An Res-Unet Method for Pulmonary Artery Segmentation of CT Images

Zhenhong Liu\textsuperscript{1} and Hongfang Yuan\textsuperscript{1,*}

\textsuperscript{1}College of Information Science and Technology Beijing University of Chemical Technology, Beijing 100029, China

\textsuperscript{*}E-mail: 2018210419@mail.buct.edu.cn

Abstract. In the recognition of pulmonary embolism, the accuracy of pulmonary artery segmentation plays a key role. Due to the irregular shape of pulmonary artery and the complex adjacent tissues, it is very challenging to segment pulmonary artery using traditional convolutional neural network. Therefore, an improved Res-Unet method for pulmonary artery segmentation is proposed in this paper. To begin with, the U-net structure is used as the basis structure to allow efficient information flow. Secondly, in order to improve the gradient circulation of the network, our model introduces residual connections based on the U-net structure, that is, adding a connection from the input to the output of the two convolutions and performing a convolution operation. Finally, to quick converge, we use a hybrid loss function, which is linearly combined by Dice loss and Cross Entropy loss. The experimental results show that the proposed framework ranks higher than U-net on recall, precision and Dice, yielding results comparable to that of manual segmentation.

1. Introduction
Pulmonary embolism (PE) refers to the pathological and clinical conditions caused by impacting substances entering the pulmonary artery and blocking the blood supply to the tissues. Nowadays, computed tomographic (CT) is the most common screening method for PE. Since PE is only present in pulmonary artery, physicians always diagnose PE by looking for the pulmonary artery in the CT image \textsuperscript{[1]}. However, manual positioning of the pulmonary artery labour-intensive and prone to misdiagnosis and missed diagnosis. Therefore, the use of computers to accurately extract pulmonary arteries is of great significance for the detection of PE.

In the field of medical image segmentation, many scholars at home and abroad have put a lot of energy into it. Some previous work is to use traditional image segmentation methods to segment target areas in CT images. Due to the complex shape of the pulmonary artery, most target segmentation methods cannot be universally applicable. At present, an increasing number of scholars use deep learning methods to solve complex problems in the field of medical imaging. Olaf et al. \textsuperscript{[2]} used very little data to complete the end-to-end neural network training, and obtained good results. Kumar et al. \textsuperscript{[3]} used the multi-U-net method to segment breast masses, and achieved good results. In the field of semantic segmentation in biomedical applications, U-net \textsuperscript{[4]} is a popular deep learning architecture, it can combine shallow information with deep information and can achieve good segmentation results. However, due to the different morphology of the pulmonary artery region, the U-net model did not achieve the expected results in lung CT image segmentation. In order to achieve more precise segmentation, this paper proposes a new automatic segmentation algorithm for...
pulmonary artery—improved Res-Unet neural network, which combines the original U-net network with the residual connection [5].

Experimental results show that compared with the popular U-net [4] and ReeUnet in [6], the performance of our model is improved. The main contributions of this article are as follows:

1. For the problem of insufficient feature extraction (missing edges), the residual unit is used to replace the feature extraction layer composed of common convolution kernel in U-net architecture.

2. A batch normalization (BN) layer [7] is added to the network for normalization, which can effectively speed up model convergence and enhance the generalization ability of the model.

3. We analysed the performance of various loss coefficients for semantic segmentation, and used a hybrid loss function [6] to improve network training performance which is combined with Cross-Entropy loss and Dice loss.

2. Methodology

2.1. The Main Framework
U-Net is an image segmentation network proposed in 2015 by Ronneberger et al. [6]. It has been widely studied and applied in the field of medical image segmentation. It replaces the convolutional layer with the fully connected layer in the network to achieve end-to-end training of images, and can combine the information of the bottom and high layers to reduce the loss during feature extraction.

ResUnet architecture in [6] uses the residual module (as shown in figure 1) to replace the ordinary module of the U-net architecture.

![Figure 1. Residual Block.](image)

For our Res-Unet network, residual connection and batch normalization (BN) layer were added to normalize the upper and lower sampling layers (as shown in figure 2(a), (b)), the residual connection is used to avoid gradient vanishing occurring in the deep network section. The batch normalization layer can effectively accelerate the speed of network convergence and improve the generalization ability of the model. The above strategies can make the network converge faster. We added network layers to improve network performance and achieve more precise segmentation. We construct ResUnet with four encoder blocks and decoder blocks instead of three in [7].

![Figure 2. (a).Encoder Block,(b).Decoder Block](image)
Our Res-Unet in this article is shown in figure 3. It can be divided into two parts, namely, the encoding path and the decoding path. The left half of the network is the encoding path, while the right one is the decoding path. The first residual unit performs two convolutions on input image and an identification mapping. The later portion of encoding path consists of four encoder blocks. In each block, a convolutional layer with the stride of 2 is used to reduce the spatial dimension of the feature maps by half, rather than the pooling layer [8]. Figure 2(a) shows the structure of the encoder block. The structure of the decoder block (as shown in figure 2(b)) is similar to the encoder block, but the stride of all convolutional layers in it is 1. In each unit, there is an up-sampling from lower level and a concatenation with the feature maps from the corresponding encoding path. At the end of the network, a 1x1 convolution and a sigmoid activation layer are used to convert the channel map into a feature map with two categories.

![Residual Block Diagram](image)

**Figure 3.** Residual Block.

### 2.2. Loss function

The model was trained by using different loss functions. For example, Cross-Entropy loss, MSE loss, Dice loss, the combination of two losses. By analysing their performance and conducting comparative experiments, we observed that the model achieved a higher performance with the combination of Cross-Entropy loss and Dice loss.

F Milletari et al. [9] firstly used Dice loss in their neural network to segment medical image. Dice coefficient is a similarity measurement, it is defined as follows:

$$\text{Dice} = \frac{2\sum_i^n p_i g_i}{\sum_i^n p_i^2 + \sum_i^n g_i^2}$$  \hspace{1cm} (1)

$$\text{Loss}_{\text{Dice}} = 1 - \frac{2\sum_i^n p_i g_i}{\sum_i^n p_i^2 + \sum_i^n g_i^2}$$  \hspace{1cm} (2)

Where $N$ is the sum of pixels, $p_i$ is the predicted label and $g_i$ is the ground truth label, $p_i, g_i \in [0,1]$.

Dice coefficient is a similarity measurement, and is quite effective in dealing with imbalanced categories. However, Dice loss is unstable during the training process, which is not conducive to network convergence. Cross-Entropy loss [10] is used extensively in network training by calculating the coincidence degree between prediction and label, it maintains a smooth gradient for all pixels and helps the convergence of loss function, so we add cross entropy to stabilize training. Its formula is as follows:
\[
Loss_{CE} = -\frac{1}{N} \sum_{i}^{N} (g_i \log p_i + (1-g_i) \log(1-p_i))
\]  (3)

Where \( p_i \) is the predicted label and \( g_i \) is the ground truth, \( p_i, g_i \in [0,1] \). \( Loss_{CE} \) is the average of \( N \) samples. In order to take full advantage of the two losses, we combine \( Loss_{Dice} \) and \( Loss_{CE} \) as \( Loss_{DC} \), which is defined as follow:

\[
Loss_{DC} = \alpha Loss_{Dice} + (1-\alpha) Loss_{CE}
\]  (4)

The fix factor \( \alpha \) was defined to control the balance between \( Loss_{Dice} \) and \( Loss_{CE} \). When the factor are fixed to \( \alpha = 0.3 \) through experiments, the network training loss is the smallest.

3. Experiment

3.1. Dataset

Our own data set is the lung CT image sequence provided by the China-Japan Friendship Hospital. There are 70 patients in total. Each patient has 30 CT images. The pulmonary artery area of each CT image is manually marked by professional physicians. We also conducted experiments on the public data set CHAOS. CHAOS contains CT images of 40 patients and MRI images of 120 patients. This article only uses the CT image part, which contains 6315 CT images of 40 patients.

The image used in the experiment in this article is a two-dimensional CT image. First, the images in the DICOM format in the data set are uniformly converted into a common picture format (this article uses the bmp format), and then the data set is normalized.

In order to expand the lung CT image data provided by China-Japan Hospital, We expanded the data set to 6 times the original through data enhancement techniques. The training data are randomly assigned in an 8:1:1 ratio as training set, test set and verification set.

3.2. Implementation details

The experimental environment of this research is Lunix, and the experiment is completed on a hardware platform of 2 * Gold 6148, 256G, 4 * Nvidia V100 GPU. All models are implemented using Keras framework with Tensorflow as backend. The optimization is solved by Adam (adaptive moment estimation), the initial learning rate of neural network is: 1e-4. The batch size is 2.

**Figure 4.** The segmentation result of CHAOS dataset. (a) image; (b) mask; (c) U-net; (d) ResUnet; (e) Proposed.
3.3. Results and comparisons
To prove the validity of the ResUnet architectures proposed in this paper, we conducted experiments on the lung CT image sequence and the public data set CHAOS. U-net and ResUnet architectures in [6] was trained as the comparative experiments to evaluate our network, because they are common structures for semantic segmentation tasks. Figure 4 is the result of segmentation of the pulmonary artery CT sequence. (a) represents the input image, (b) is the mask of the target area generated from the labeled file, representing the label value. (c) represents the segmentation result of U-net. (d) represents the segmentation results of ResUnet in [6], the rightmost figure is the result of the prediction of the architectures proposed in this paper.

Then the proposed architecture was applied to the lung CT sequence. It can be seen that the segmentation results of our architecture (figure 5(e)) are superior to the U-net and original ResUnet in [6], and are highly similar to those labelled by doctors (Fig. 5(b)).

Figure 5. The segmentation result of the pulmonary artery. (a) image; (b) mask; (c) U-net; (d) ResUnet; (e) Proposed

After establishing the model, it is very important to evaluate its performance. According to [11], recall and precision are useful and appropriate metrics when dealing with unbalanced classes. In addition, we also use the Dice coefficient.

When experimenting on CHAOS dataset, Table 1 shows the performance of different models. To evaluate the performance of hybrid loss function, we compare the segmentation results of Res-Unet with Cross-Entropy loss and the segmentation results of Res-Unet with the hybrid loss.

Then the proposed architecture was applied to the pulmonary artery segmentation of lung CT sequence. Table 2 shows the performance of different architectures.

Table 1. The performance comparison of different algorithm on CHAOS dataset.

| Architecture    | Dice  | Precision | Recall |
|-----------------|-------|-----------|--------|
| U-net           | 0.923 | 0.931     | 0.926  |
| ResUnet in[6]   | 0.939 | 0.948     | 0.944  |
| Our             | 0.963 | 0.966     | 0.961  |
| Our+LossDC      | **0.965** | **0.969** | **0.961** |
Table 2. The performance comparison of different algorithm on our dataset.

| Architecture      | Dice   | Precision | Recall  |
|-------------------|--------|-----------|---------|
| U-net             | 0.957  | 0.961     | 0.965   |
| ResUnet in[6]     | 0.971  | 0.969     | 0.973   |
| Our               | 0.980  | 0.976     | 0.983   |
| Our+LossDC        | 0.981  | 0.979     | 0.984   |

It can be seen from Table 1 and Table 2 that our improved Res-Unet architecture shows better performance in Dice, Recall and Precision, which shows the effectiveness of proposed architecture. For the improved Res-Unet with a single Dice loss, the indicators are shown in the third row of the table. For the improved Res-Unet with hybrid loss, Dice coefficient is 0.98, Precision is 97.6%, and Recall is 0.983. Experimental results show that the proposed Res-Unet can improve the performance of pulmonary artery segmentation.

4. Conclusion
In order to segment the pulmonary artery region accurately, this paper proposes an improved Res-Unet. First, to solve the problem of insufficient feature extraction of U-net, residual connection was added to the feature extraction layer composed of the ordinary convolution layer. Then, batch normalization was added, which effectively accelerates the network convergence speed and prevents the disappearance of the network gradient. Finally, a hybrid loss function is used to improve training performance. The experimental results show that the model proposed in this paper can accurately segment the pulmonary arteries of the lung CT images, which facilitates the subsequent three-dimensional reconstruction of the pulmonary arteries and the calculation of the pulmonary embolism volume, which is convenient for experts to evaluate the severity of pulmonary embolism.

References
[1] F. G. Meinel, N. J. Jr, U. J. Schoepf, et al., “Predictive Value of Computed Tomography in Acute Pulmonary Embolism: Systematic Review and Meta-analysis,” Am. J. Med., vol. 128, pp. 747–759, 2015.
[2] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation.". in MICCAI, 2015.
[3] Kumar V, Webb JM, Gregory A et al. Automated and real-time segmentation of suspicious breast masses using convolutional neural network[J]. PloS one, 2018, 13(5): e0195816-e0195816.
[4] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in Proceedings of International Conference on Medical image computing and computer-assisted intervention. Springer, 2015, pp. 234–241.
[5] Kaiming H, Xiangyu Z, Shaoqing R, et al. Deep Residual Learning for Image Recognition. arXiv preprint arXiv:1512.03385 [cs.CV] (2015).
[6] Z. Zhang, Q. Liu, and Y. Wang, “Road extraction by deep residual u-net,” IEEE Geoscience and Remote Sensing Letters, vol. 15, no. 5, pp.749–753, 2018.
[7] Ioffe, Sergey, and Christian Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift.", in arXiv preprint arXiv, 2015.
[8] Xu W, Liu H, Wang X, et al. Liver Segmentation in CT based on ResUNet with 3D Probabilistic and Geometric Post Process[C]// IEEE International Conference on Signal and Image Processing. IEEE, 2019.
[9] Milletari, Fausto, Nassir Navab, and Seyed-Ahmad Ahmadi. "V-net: Fully convolutional neural networks for volumetric medical image segmentation.", in 3DV, 2016.
[10] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation.", in MICCAI, 2015.
[11] A. Tharwat, “Classification assessment methods,” in Applied Computing and Informatics. Riyadh, Saudi Arabia: King Saud Univ, 2018.