Lung Parenchyma Segmentation Based on Improved Unet Network

Lei Lv, Xin Sun
School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 200444, China
Corresponding authors’ e-mail: xsun@staff.shu.edu.cn

Abstract. Segmentation of lung parenchyma is an essential link in the diagnosis of lung diseases and also the premise of disease analysis. The accuracy of lung parenchyma segmentation affects the diagnosis and treatment of lung diseases. The size of input data will be reduced by using the traditional Unet network. In this paper, an improved Unet network structure is proposed to segment lung parenchyma automatically. In the process of convolution, the size of input data is kept constant by padding same and dropout layer is introduced into the network. We use cross entropy loss function to train the model for the first time. After the model converges, we use custom Dice loss function to fine tune to improve the accuracy. By calculating Jaccard coefficient and DSC coefficient, our lung parenchyma segmentation method has a very high accuracy, which is better than the earlier researchers’ segmentation algorithm. The significance of this study is to provide pretreatment for the diagnosis and treatment of lung diseases.

1. Introduction
Nowadays, air pollution and bad eating habits lead to a sharp increase in lung cancer patients, lung cancer has become the biggest threat to health, so lung cancer treatment is imminent. At present, Computed Tomography (CT) is the main method to examine lung diseases. CT image can effectively reflect the information of the lesion area. Doctors diagnose and treat patients according to the information obtained from CT image. In Computer Aided Diagnosis (CAD) [1] lung disease system, lung parenchyma should be segmented first, and the accuracy of segmentation directly affects the diagnosis of doctors. Therefore, it is of great significance to improve the accuracy of segmentation for the diagnosis and treatment of lung diseases.

Earlier, some scholars have put forward a lot of algorithms for lung parenchyma segmentation of lung CT images and got some achievements. Many researchers use threshold segmentation method [2,3], which is the segmentation algorithm mentioned by Sluimer [4] in the overview. Soliman et al [5] used the modeling method to model the lung CT image as a three-dimensional Markov Gibbs random field (MGRF) for lung parenchyma segmentation. P. Korfeitis et al [6], combined 3D threshold and 2D wavelet transform to realize the automatic segmentation of lung parenchyma. Mansoor et al [7] used fuzzy connectivity (FC) to extract lung parenchyma image, which is a region segmentation algorithm based on local feature description. Hosseiniasl et al [8] put forward an algorithm based on nonnegative matrix decomposition. By unsupervised learning, the accuracy of lung segmentation is 96.5%, and the Dice Similarity Coefficient is 0.96.

Deep learning is an important branch of machine learning [9]. Deep learning adds feature extraction and learning to the establishment of the model. It can automatically learn the relationship
between features and target samples. Therefore, it can reduce the characteristics of defects caused by human participation in design features. At present, the application of deep learning is very extensive. In some specific tasks, the effect of deep learning has exceeded the performance of existing algorithms. At present, the research field of deep learning has been extended to medical image processing. Yu et al [10]. used a fully convolutional residual network (FCRN) to segment skin melanoma lesions accurately. Mansoor et al [11]. used SAE to infer the characteristics of visual conduction pathway in MR images, and realized the automatic division of left and right optic nerve.

We propose a method of lung parenchyma segmentation based on the improved Unet [12] network structure to raise the accuracy of segmentation. At the same time, dropout layer is introduced into the network to inhibit overfitting of model, so as to improve the ability of lung parenchyma segmentation. Because of the traditional Unet network, the size of the output image is smaller than that of the input image, which cannot fully show the real segmentation effect. In this paper, the whole image is segmented by filling. The overall scheme includes preprocessing, improved Unet network model and segmentation effect evaluation. By training the data set composed of lung CT images, we get good results. Eventually, we segmented the image of the test set by the trained model, and the Dice coefficient is 0.988.

2. Research methods

2.1. Method flow
Firstly, the original CT image of lung is preprocessed to remove noise, filter and smooth. Because the number and quality of the train set images have a great influence on the training of neural network, we need to add the number of train set images after preprocessing the original lung CT images. The method to add the number of images is through translation, rotation and flipping. The transformed image is reconstructed into the training set and read into the improved Unet network. In the process of training, the size of input image is kept unchanged by padding same. We select the cross entropy loss function for the first training. After the model converges completely, we use our customized Dice loss function to fine-tune the model. Finally, we use the final model to predict the pictures of the test set and calculate their Jaccard [13] coefficients and DSC [14] coefficients. By comparing with the models proposed by predecessors, our method has proven to be more advanced and effective.

2.2. Unet network structure
Today, deep learning is developing rapidly and is being used more and more widely. There are many excellent networks in the task segmentation. Unet is a new network based on CNN, which was first proposed by Olaf Ronneberger, Phillip Fischer and Thomas Brox [12] in 2015. It can better segment biomedical images. The name Unet comes from its structure ——The visualization of network structure is like a letter ‘U’. This paper improves the traditional Unet network structure. We add dropout layer before the last step of downsampling and after the first step of upsampling, which can improve the generalization ability of the network and inhibit the overfitting of the model in the training process. Our improved Unet network structure is shown in Figure 1.
Figure 1. Unet network structure

In the whole network structure, the propagation process of the purple arrow consists of the convolution layer with the size of $3 \times 3$ and the ReLU activation layer. The propagation process of the red arrow consists of a $2 \times 2$ pool layer. The propagation process of the yellow arrow is the dropout layer. The propagation process of the bright green arrow is composed of $2 \times 2$ deconvolution layers. The propagation of the dark green arrow consists of a convolution layer with a size of $1 \times 1$.

The size of the image entered into the Unet network is $512 \times 515$. This network is composed of left and right paths, in which the left path contracts from the top down and the right path expands from the bottom up. Among them, the contracting path is mainly used to collect the feature map of image data, while the expanding path is used to expand the size of the feature map, and the two paths are symmetrical to each other. The contracting path is the convolution layer from left to right and the size of each convolution layer is $3 \times 3$. After the convolution operation, the nonlinear function ReLU is used to overcome the gradient vanishing phenomenon. Each two convolution layers form a group, and we add dropout layer after the 4th and 5th group to suppress overfitting. A $2 \times 2$ pooling layer is introduced between each convolution layer to maximize the pooling operation. In the process of next sampling, the number of characteristic channels is increasing. In each step of the expanding path, the obtained characteristic image is sampled up, and then the number of characteristic channels is reduced by deconvolution operation with the size of $2 \times 2$. The expanding path consists of 8 convolution layers of $3 \times 3$. At the end of each convolution layer, we use the ReLU as the active function. The characteristic information of contracting path and expanding path are fused to form a U-shaped network structure. The final step of the network operation is to restore the feature map with the same size as the input image through the convolution layer of size $1 \times 1$.

2.3. Image data preprocessing

The principle of lung CT image makes CT image affected by external factors such as volume effect, offset field effect and motion. CT image has the disadvantages of blurred image boundary, poor contrast between light and dark, and artifacts such as strip, ring and motion compared with ordinary image. Before using lung CT image for Unet network training, image preprocessing is needed. Firstly,
we should enhance the contrast of CT image and transform the gray level of image. Figure 2 shows the pre-processed image of the original image and the real segmentation label image.

In Figure 2, (a) is the original CT image of the lung. (b) is the preprocessed lung CT image. (c) is the real split label image.

Increasing the number of images in the train set helps us train the neural network. Generally, the number of lung CT images we get is relatively small. In order to improve the accuracy of model training, we use image translation, rotation, flipping and other operations to increase the number of training pictures. Figure 3 shows the pre-processed image after translation, rotation, and inversion, as well as the corresponding label image after transformation.
In Figure 3, (a) is the pre-processed image, (b) is the image after translation, rotation and inversion, and (c) is the transformed label image.

3. Training

The training part includes the selection of training algorithm, the definition of loss function and model fine-tuning.

3.1. Training algorithm

The choice of optimization algorithm has a direct influence on the efficiency of model training. There are many kinds of training algorithms for convolutional neural network. The Adaptive Learning Rate Method (delta) and Adaptive Moment Estimation (Adam) are commonly used for Adaptive training. We adopt Adam algorithm with fast convergence speed as the training algorithm.

3.2. Training process

In the experiment, Intel Core 2.6GHz i7-9750H CPU and NVIDIA RTX2060 GPU with 6GB memory are used to train the Unet network model. This experiment uses Python as programming language to develop and design, and builds network model with tensorflow deep learning framework. The data set used in the experiment comes from ‘Finding and Measuring Lungs in CT Data’ [15] provided in the kaggle competition. During the training, 230 pulmonary CT images were transformed to increase the number of images to 5280. We divided 5280 images into three parts. 60% as the train set, 20% as the validation set, and the remaining 20% as the test set.

Then, the data set was preprocessed and input to the Unet network for training. The model parameters are iteratively updated by calculating the loss function between the predicted data and the real label data. After several epochs, the value of the loss function gradually decreases. When the loss function decreases gradually and reaches stability, the model tends to converge. At this time, we save the model. In the training process, we use the cross entropy loss function as the first training, and the expression of the function is as follows:
\[ \text{Ent_loss} = -\sum_{i=0}^{m} [v^i \log \hat{v}^i + (1 - \hat{v}^i) \log (1 - \hat{v}^i)] \]  

(1)

In this formula: \(v^i\) is each pixel in the label image of the training image. \(\hat{v}^i\) on behalf of the predicted value with each pixel in the training image; \(m\) is the total number of samples. The reason for choosing the loss function is that when the subtraction between the pixel value of the two images is larger, the greater the value of the Ent_loss will be. This means that the more 'penalty' the parameters in the current model are, and the degree of 'penalty' is close to the exponential nonlinear growth. This is determined by the characteristics of the log function itself. The advantage of doing so is that the predicted results will be more inclined to the real labeled data.

We use the deformation form of dice function, the formula is as follows:

\[ D(\alpha, \beta) = \frac{|\alpha \cap \beta|}{|\alpha| + |\beta|} \]  

(2)

In the formula, \(\alpha\) is the predicted value with each pixel in the training image and \(\beta\) is each pixel in the label image of the training image.

When the model reaches a certain degree of convergence, we take \(1 - D(\alpha, \beta)\) as the new loss function to fine-tune the model. In the task of binary segmentation of lung parenchyma in CT image segmentation, we divide the entire image into two areas, with the lung parenchyma in the foreground and the other part is the background. There are two Dice coefficients and Dice loss functions corresponding to foreground and background respectively. In the experiment, we assigned different loss weight ratios to the foreground region and the background region, as follows:

\[ D_{\text{loss}} = 0.9 \times \text{Dice_loss}_1 + 0.1 \times \text{Dice_loss}_2 \]  

(3)

In this formula: \(\text{Dice_loss}_1\) represents Dice loss function of foreground area; \(\text{Dice_loss}_2\) represents Dice loss function of background area. Figure 4 shows the flow of the operation.

4. Experimental results

We input the training pictures into the Unet network and train 25 epochs. After 15 epochs, the model tends to converge, the error decreases and remains at a very low level, the accuracy increases to 99.3% and remains stable. At this time, we can see that the model has completely converged. Figure 5 shows the accuracy curve and error curve in the training and validating process.
After 25 epochs, we used the trained model to predict the image of the test set. Figure 6 visualizes the experimental results.

In Figure 6, (a) is the original image, (b) is the actual label image and (c) is the predicted image.

5. Algorithm evaluation
Today, deep learning is developing rapidly. The application of deep learning to image segmentation is
expanding. The methods to evaluate the accuracy of segmentation usually include Dice coefficient and Jaccard coefficient.

Dice similarity coefficient (DSC), or overlapping comparison method, is used to evaluate the similarity between two contours. Generally speaking, DSC > 0.7 means that the repeatability of automatic segmentation and manual segmentation is high and the segmentation effect is good.

The application scope of Jaccard similarity coefficient is to compare the degree of similarity between two samples. Generally speaking, if the similarity between the two samples is higher, the value of Jaccard coefficient will be larger.

5.1. Jaccard coefficient
The Jaccard coefficient was first proposed by Jaccard in 1912 and applied to the floristic distribution analysis of alpine areas. Tan et al [16] introduced this method into neural network. The expression of the Jaccard coefficient is shown below:

\[ J(\alpha, \beta) = \frac{|\alpha \cap \beta|}{|\alpha \cup \beta|} \quad (4) \]

We can see that the value range of \( J(\alpha, \beta) \) is 0 to 1. In particular, when \( \alpha \) and \( \beta \) are empty sets, \( J(\alpha, \beta) \) is 1. The value of \( J(\alpha, \beta) \) determines the similarity between the two samples. The bigger the calculated value of \( J(\alpha, \beta) \), the higher the similarity between the two samples will be.

Jaccard distance is used to evaluate the difference between two samples, which complements the Jaccard coefficient. The Jaccard distance is expressed as:

\[ d_j(\alpha, \beta) = 1 - J(\alpha, \beta) = \frac{|\alpha \cup \beta| - |\alpha \cap \beta|}{|\alpha \cup \beta|} \quad (5) \]

Jaccard coefficient in the segmentation image evaluation algorithm is as follows:

Given two binary images \( \alpha \) and \( \beta \). \( P_{00} \) is defined as the number of pixels of image \( \alpha \) and image \( \beta \) with value of 0. \( P_{01} \) is the number of image \( \alpha \) with pixels value of 0 and image \( \beta \) with pixels value of 1. \( P_{10} \) is that the number of pixels value of image \( \alpha \) is 1 and the number of pixels value of image \( \beta \) is 0. \( P_{11} \) is the number of pixels value of image \( \alpha \) and image \( \beta \) that is 1 at the same time.

So \( J(\alpha, \beta) \) and \( d_j(\alpha, \beta) \) can be deformed in the following form:

\[ J(\alpha, \beta) = \frac{P_{11}}{P_{11} + P_{10} + P_{01}} \quad (6) \]
\[ d_j(\alpha, \beta) = \frac{P_{01} + P_{10}}{P_{11} + P_{10} + P_{01}} \quad (7) \]

5.2. DSC coefficient
The DSC coefficient is determined by the area or volume of the overlapping part of the contour of two images as a percentage of their total area. The DSC coefficient is calculated as follows:

\[ \text{DSC} \ (\alpha, \beta) = \frac{2(\alpha \cap \beta)}{\alpha + \beta} \quad (8) \]

In the formula above, \( \alpha \cap \beta \) indicates the area of overlap between two images. \( \alpha + \beta \) is the sum of the areas of the two images. It can be known from the calculation formula that DSC can take values from 0 to 1. When the two calculation areas coincide with each other, the maximum value of DSC coefficient is 1, and when the two calculation areas do not coincide completely, the minimum value of DSC coefficient is 0. Although DSC is sensitive to the difference of location and size, the difference of location relative to size can reflect the difference more in the process of DSC measuring image similarity. For example, if an image is moved so that only half of the regions are overlapped, the DSC coefficient is only 0.5.

In order to expand the range of DSC coefficient, the DSC calculation formula can be transformed as follows:
\[
\text{Logit}(\text{DSC}) = \ln\left(\frac{\text{DSC}}{1-\text{DSC}}\right)
\]

The value range of Logit (DSC) is the whole real number field and has monotonicity. When DSC value is 0.5, the corresponding Logit (DSC) value is 0. By adjusting the value range of DSC, we can clearly see the differences between the similar data. Although the calculation process of DSC is simple, it cannot describe all cases.

5.3. Comparison with other algorithms
We randomly selected 8 lung CT images for testing. The results of Jaccard coefficient and DSC coefficient are shown in the table:

| Algorithm | Image | I   | II  | III | IV  | V   | VI  | VII |
|-----------|-------|-----|-----|-----|-----|-----|-----|-----|
| Jaccard   |       | 0.9753 | 0.9734 | 0.9723 | 0.9676 | 0.9780 | 0.9813 | 0.9877 | 0.9804 |
| DSC       |       | 0.9875 | 0.9865 | 0.9860 | 0.9835 | 0.9889 | 0.9896 | 0.9914 | 0.9879 |

The maximum number of Jaccard coefficients is 0.9877, the minimum number is 0.9676, and the average number is 0.977. The maximum number of DSC coefficient is 0.9914, the minimum number is 0.9835, and the average number is 0.9877. From the above data, we can see that the data are very close to 1, and the fluctuation range of the data is very small. So, we can know that the segmentation accuracy of lung CT image is very high.

In order to prove the ideal effect of our segmentation, we compared our results with several methods proposed by previous researchers, such as intensity (I) [17] model, intensity and spatial (IS) [17] model, iterative threshold (IT) [2] model, nonnegative matrix factorization (NMF) [18] model, multiple resolution segmentation (MRS) [19] model and gradient vector flow (GVF) [20] model. From the data in Table, we can see that our segmentation method is more accurate than other methods.

| DSC | Segmentation Algorithms |
|-----|-------------------------|
| Our | I | IS | IT | NMF | MRS | GVF |
| Min | 0.984 | 0.578 | 0.634 | 0.613 | 0.945 | 0.539 | 0.692 |
| Max | 0.991 | 0.742 | 0.883 | 0.891 | 0.988 | 0.750 | 0.941 |
| Mean | 0.988 | 0.632 | 0.783 | 0.816 | 0.966 | 0.613 | 0.848 |

6. Conclusion
An improved Unet network is proposed to segment lung parenchyma. Firstly, the data of lung CT images are preprocessed and transferred to the improved Unet network, and then we use the cross entropy function to update the parameters to get the optimal network model. After the first training convergence of the model, we fine-tune the model using the custom Dice loss function to optimize the model again. Using our model to segment lung parenchyma can get more accurate segmentation results. In order to reasonably evaluate the model obtained by our algorithm, we compare our algorithm with those proposed by previous scholars, and the calculated Jaccard coefficient and DSC coefficient are better than the results of other models. The comparison result indicates that our segmentation algorithm has better performance than previous segmentation algorithms. At present, our work only stays in two-dimensional image segmentation, and then our work will deepen the segmentation algorithm into three-dimensional space.

Acknowledgments
This work is supported by the National Key Research and Development Program of China (No. 2019YFB1405500).
References

[1] Singadkar, G., Talbar, S., Sanghavi, P., et al. Automatic Lung Field Segmentation Based on Non Negative Matrix Factorization and Fuzzy Clustering. *Smart Trends in Systems, Security and Sustainability*. 2018, 18: 57-66.

[2] Nakagomi, K., Shimizu, A., Kobatake, H. et al. Multishape graph cuts with neighbor prior constraints and its application to lung segmentation from a chest CT volume. *Medical Image Analysis*. 2013, 17(1): 62-77.

[3] S. Hu, E. A. Hoffman, J. M. Reinhardt. Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. *IEEE Trans. Med. Imaging*. 2001, 20(6): 490-498.

[4] I. Sluimer, A. Schilham, M. Prokop, et al. Computer analysis of computed tomography scans of the lung: a survey. *IEEE Trans. Med. Imaging*. 2006, 25(4): 385-405.

[5] Soliman A, Khalifa F, Elnakib A, et al. Accurate Lungs Segmentation on CT Chest Images by Adaptive Appearance-Guided Shape Modeling. *IEEE Trans. Med. Imaging*. 2017, 36(1): 263-276.

[6] P. Korfiatis, S. Skiadopoulos, P. Sakellaropoulos, et al. Combining 2D wavelet edge highlighting and 3D thresholding for lung segmentation in thin-slice CT. *The British Journal of Radiology*. 2008, 80(96): 996-1004.

[7] Mansoor, Awais, U. Bagci, et al. Near-optimal Keypoint Sampling for Pathological Lung Segmentation. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE Engineering in Medicine and Biology Society. 2014, 6032-6035.

[8] Hosseini asl, Ehsan, Zurada Jacek, et al. 3D Lung Segmentation Based on Incremental Constrained Nonnegative Matrix Factorization. *IEEE T BIO-MED ENG*. 2015, 63(5):952-963.

[9] G. E. Hinton, S. Osindero, Y. The. A Fast Learning Algorithm for Deep Belief Nets. *Neural Computation*. 2006, 18(7):1527-1554.

[10] Yu L, Chen H, Dou Q, et al. Automated Melanoma Recognition in Dermoscopy Images via Very Deep Residual Networks. *IEEE Trans Med Imaging*. 2017, 36(4): 994-1004.

[11] A. Mansoor, J. Cerrolaza, R. Idrees, et al. Deep Learning Guided Partitioned Shape Model for Anterior Visual Pathway Segmentation. *IEEE Trans. Med. Imaging*. 2016, 35(8):1856-1865.

[12] O. Ronneberger, P. Fischer, T. Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Medical Image Computing and Computer-Assisted Intervention–MICCAI. 2015, 234-241.

[13] Jaccard P. The Distribution of Flora in the Alpine Zone. *New Phytologist*. 1912, 11(2): 37-50.

[14] Dice L R. Measures of the Amount of Ecologic Association Between Species. *Ecology*. 1945, 26(3): 297-302.

[15] “Finding and Measuring Lungs in CT Data”, Kaggle Competition, https://www.kaggle.com/kmader/finding-lungs-in-ct-data

[16] Tan P N, Steinbach M, Kumar V, et al. Introduction to Data Mining: Pearson New International Edition. *Pearson Schweiz Ag*. 2013, 14(2):279-288.

[17] Abdollahi B. Soliman A. Civelek A C, et al. A novel Gaussian Scale Space-based joint MGRF framework for precise lung segmentation. *IEEE International Conference on Image Processing*. IEEE, 2013.

[18] Asl, Ehsan Hosseini, J. M. Zurada, et al. Lung Segmentation Based on Nonnegative Matrix Factorization. *Image Processing (ICIP)*. 2014 IEEE International Conference on IEEE. 2014.

[19] Bouman C, B. Liu. Multiple resolution segmentation of textured images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1991, 13(2): 99-113.

[20] Xu Chenyang, J. L. Prince. Snakes, shapes, and gradient vector flow. *IEEE Transactions on Image Processing*. 1998, 7(3): 359-369.