Improved U-NET network for pulmonary nodules segmentation

Guofeng Tong, Yong Li⁎, Huairong Chen, Qingchun Zhang, Huiying Jiang

College of Information Science and Engineering, Northeastern University, Shenyang, 110819, China

ABSTRACT

Since pulmonary nodules in CT images are very small and easily confusing with other tissues, there are still many problems in the pulmonary nodule segmentation. This paper presents an improved lung nodule segmentation algorithm based on U-NET network. Firstly, CT images are transformed and normalized, and the lung parenchyma is obtained by simple and efficient morphological method. Then, the U-NET network is improved, which mainly includes the dataset rebuilding, convolutional layer, pooling layer and upsampled layer. And we introduced residual network, which has improved the network training effect. Besides, we designed batch standardization operation, which has speeded up the network training and improves the network stability. Finally, we used the new dataset to train and test the improved U-NET network. A large number of experiments show that the proposed method can effectively improve the segmentation accuracy of pulmonary nodules. It is a great work with theoretical and practical value.

1. Introduction

Pulmonary nodules are very small globular shadows in the CT images, which are approximately spherical. Pulmonary nodule is very easy to be confused with other tissues, so it is very difficult for the human eyes to identify. And it is particularly difficult to identify the pulmonary nodules with vascular adhesions. Therefore, the segmentation of pulmonary nodules is a difficult problem in medical image processing.

There are many segmentation methods for pulmonary nodules in recent years. The common segmentation methods are divided into two classes: (1) traditional unsupervised pulmonary nodule segmentation method [1], (2) pulmonary nodule segmentation based on machine learning method.

The commonly used methods based on unsupervised traditional methods include morphological method, threshold segmentation, clustering and so on. Although these methods are fast and simple, there are many kinds of problems, such as under - segmentation or over-segmentation. The morphological method [1] may remove the marginal burr of the pulmonary nodule, and the parameters involved in the operation are not easy to control. The method of threshold segmentation [2] is not ideal for the segmentation of vascular adhesions tuberculosis. Ref. [3] uses EM and mean-shift methods to extract nodules. But if the number of adhesion nodules is greater than or equal to 2, the segmentation of the method is not ideal. Other traditional approaches also have problems. For example, the method proposed by Armato et al. [4] is time-consuming, limited and inconvenient to use. Kanazawa et al. [5] lost 3D spatial feature information. Okumura et al. [6] lead into too much manual operation, and the degree of automation is low.

Pulmonary nodule segmentation based on machine learning, such as the methods [7,8] of continuous k-means classifier for localization and image segmentation. However, machine learning requires artificially design representative features, which is relatively waste of time and energy. In recent years, deep learning has made great progress in computer vision, which makes some people...
try to use deep learning to solve the problem of pulmonary nodule detection. Deep learning can automatic extract features from training data. Currently, the use of convolutional neural networks for CADe/x is still in infancy. However, previous studies have shown that compared with traditional segmentation methods, deep learning can produce less false positive judgment. In addition, DNN(Deep Neural Networks) has a great potential for application in various CADe/x tasks involving volume medical data [9,10].

Deep learning describes various computational models consisting of multiple processing layers (for example, a deep network composed of multiple layers). These layers mainly learn about the abstract representation of different levels of data. With the powerful feature extraction capabilities of deep learning, it can gradually replace features that are artificially defined based on machine learning methods [11]. However, the application of deep learning methods for medical images is still in its infancy. Earlier studies successfully using deep neural networks to medical image segmentation tasks have shown some progress [12,13].

Aiming at the problem of pulmonary nodule segmentation, the U-NET network is introduced into pulmonary nodule segmentation in this paper. We proposed a pulmonary nodule segmentation method based on deep learning. First of all, we preprocess the CT images, and morphological methods are used to segment the image to obtain the lung parenchyma, then the segmented lung parenchyma is divided into 64*64 cubes. Finally, we use the improved U-NET network to train and get the pulmonary nodule segmentation region.

The main contributions of this paper are in the following three aspects. Firstly, in order to solve the unbalanced samples in the lung nodule segmentation net training process, a samples balanced lung nodule segmentation data set based on CT slice image with labels was rebuilt. The new data set can train a higher precision network model. Secondly, residual network is added to U-NET network. The residual network resembles an ensemble approach in a way that can improve our final predictions. The other is adding the Batch Normalization layer into U-NET network, which can speed up training, prevent over-fitting.

2. Proposed method

As shown in Fig. 1, the proposed algorithm firstly segments the lung parenchyma from the preprocessed CT images. The process includes binary segmentation, remove the border, mark region of interests(ROI), find two largest connected regions, corrosion operation, closed operation and lung lobe positioning, get the final segmentation image. Then design and improve the U-NET network. We chose a kernel size of 3*3 in U-NET, and the number of convolution kernels doubles after pooling. Next, we design the convolution layer, pooling layer and activation function. Furthermore, we introduce residual network and Batch Normalization to the neural networks.

2.1. CT image preprocessing

The CT images of the pulmonary nodules may cause some interference and equipment imaging errors due to various reasons. Even for the same focus, the information that is scanned by a different device can be different. A series of problems can be generated if the data set are trained directly in the model. Therefore, it is necessary to preprocess the image before the training.
The main formats of common CT images include the two formats of MHD and DICOM. For a patient, it usually contains a raw format file and a MHD format file. Firstly, input the CT image of the DICOM format or the MHD format, and then normalization is carried out. Image pixel values are different due to different scanning equipment or different layer thickness, while the radiation dose, HU (Hounsfield Unit), is a quantitative measure that is unaffected by those factors. The HU value and the pixel value can be converted to each other and have a linear relationship. Therefore, the HU value above 400 is taken as 400 in this paper, and the range of CT [-1000,400] is then normalized to [0,1].

2.2. Segmentation of lung parenchyma

The original CT images are very large. In order to segment the pulmonary nodules conveniently, it is necessary to remove some of the related interference. Firstly, the CT images of a slice shown in Fig. 2 are segmented based on morphological methods. Secondly, the regions that are not pulmonary nodules are removed, then the image of the lung parenchyma is obtained.

The CT value in the periphery of Fig. 2.a is -1000Hu, the areas of black, white, or light color are mainly air and blood vessels. For Fig. 2.a, firstly we use the threshold value of -400Hu for coarse segmentation[14] to get the binary image shown in Fig. 2.b. Then, each slice of the CT scan image is segmented. The pulmonary wall region is also seen as a region of interest (ROI) because some nodules may exist on the lung wall. Then the morphological operation (opened operation) is used and the value of the edge pixel is assigned to 0 to remove the object connected to the boundary, and get the image (Fig. 2.c) after the removing the boundary. Next, we calculate the area of all connected regions, and find two regions with the largest area. Then, we remove all the other small regions. Further we process it with corrosion and closed operation. Finally, in order to prevent some small bubbles in the middle of the two lobes and get a relatively complete lung region, we process it with hole filling, and get Fig. 2.d. The segmented images of the lungs are shown in Fig. 2.e.

2.3. U-NET network for pulmonary nodule segmentation

Given a CT image, output the location of the detected lung nodules, however, it is often difficult to obtain a large training set in medical problems. Olaf Ronneberger, Thomas Brox, Philipp Fischer and others proposed the U-NET network structure, which is evolved from a full convolution network and has a good effect on smaller training sets. The expansion network and the shrinking network are related to each other in U-NET. It merges the features of related contractile layers when expanding, supplementing the missing boundary information to predict edge information accurately. What's more, the training time of U-NET is relatively short, having a simple structure, less parameters and demanding less training data compared with other networks. Therefore, this paper improves the network based on U-NET.

This paper presents a network model for pulmonary nodules segmentation in CT images based on U-NET structure. The network...
The improvements of the U-NET network are mainly those two aspects. Firstly, residual network is introduced in the improved network. Residual network is to solve the problem of various errors and too much training time in deep network training. The most important one in the residual network is the residual module, which uses a method of fitting the residual map. It does not output the result of the convolutional layer directly, but the method of choosing the residual mapping, which we call “shortcuts”. Suppose that we have a hidden layer named $F(x)$, and the hidden layer or a few of hidden layers satisfy $F(x) = H(x) - x$. Then combining such non-linear modules, we can get a fairly complicated network. To form a residual module, the residual network adds up the output of multiple convolutional layers and the initial input, which extracting the features of the image and reducing the training parameters in a way. And the residual network structure is relatively simple, it can also speed up training. To some extent, that solves the problem of network degradation in deep conditions. This residual module is a network implemented by a combination of neural networks and shortcut, where shortcut is equivalent to having a simple identity map with no extra parameters.

(1) Residual network

In order to reduce the cost of computer, we improved the residual module. As shown in Fig. 4, the size of the convolution layer is 3 * 3. Two 3 * 3 convolution layers were replaced with three convolution layers, where the sizes of the three convolution layers are 1*1, 3*3, and 1*1. By replacing the residual module, the number of channels are reduced, and so does the amount of computation. On the other hand, the number of channels is reduced by the last 1 * 1 convolutional layer. To some extent, this ensures both accuracy and a reduction in computation.

(2) Introducing of Batch Normalization (BN)

Batch Normalization (BN) is introduced to normalize the data. When using SGD for training, the activation function in the mini-batch is calculated for normalization. After normalization, the mean of each dimension is 0 and the variance is 1. The normalizing
expression is as shown in Eq. (1):
\[
\chi^{(k)} = \frac{x^{(k)} - E[x^{(k)}]}{\sqrt{\text{Var}[x^{(k)}]}}
\]
(1)

In order to solve the problem of expression ability of linear model, two parameters are added in the BN layer to make the model learn these two parameters adaptively. So it can restore the normalized data itself. Specific batch normalization process is as follows,
Input: Values of \(x\) over a mini-batch, \(B = \{x_i\}_{i=1}^m\): parameters to be learned,
Output: \(y_i = BN_{\gamma, \beta(x_i)}\)

\[
\mu_B \leftarrow \frac{1}{m} \sum_{i=1}^m x_i
\]
(2)
\[
\sigma_B^2 \leftarrow \frac{1}{m} \sum_{i=1}^m (x_i - \mu_B)^2
\]
(3)
\[
\hat{x}_i \leftarrow \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}}
\]
(4)
\[
y_i \leftarrow \hat{x}_i + \beta \equiv BN_{\gamma, \beta(x_i)}
\]
(5)

The mean and variance of the testing phase are for the entire dataset, not for a batch. Therefore, the variance and mean of each batch are wrote down when training, and then the unbiased estimates of variances and means are calculated as the testing variance and mean. Those are given by:

\[
E[x] \leftarrow E_B[\mu_B]
\]
(6)
\[
\text{Var}[x] \leftarrow \frac{m}{m-1}E_B[\sigma_B^2]
\]
(7)

Batch Normalization scales the scale of each dimension in each level to the same. So it can use a relatively large learning rate when training, which will speed up the network training. When there is no BN layer, a smaller learning rate is needed when the promotion of the accuracy is not obvious. However, the initial learning rate can be larger after BN being introduced. Because the convergence of the algorithm is relatively fast, the attenuation coefficient of the learning rate can also be larger. On the other hand, it solves the problem of gradient disappearance or gradient explosion in the process of back-propagation during training to some extent.

3. Results and discussions

3.1. Dataset

The data set used in this paper is the dataset of LUNA2016 contest. All the data are the images of lung CT in the format of mhd. The patient’s sample number is 1000, all containing pulmonary nodules. Nodules 5–10 mm and nodules 10–30 mm each account for 50%. Each image of a pulmonary nodule contains multiple slices, each 512 × 512 in size. However, the exact number of slices of per CT image will vary depending on the scan machine, the layer thickness, and the patient. For example, a CT image with a slice thickness of 1.25 mm generally contains more than two hundred slices, and the size of each image is 512 × 512 pixels.

The data format of this paper is shown in Table 1. The CT image data is in mhd format, and the csv file marked the size and location of pulmonary nodules. Where seriesuid is the patient’s label, CoordX, coordY and coordZ represent the corresponding lung nodules X–Z coordinate information, and diameter_mm is the size of the pulmonary nodules radius, in units of mm. Because the image contains a large amount of information, the goal of segmentation is based on the small target of the pixel. The input image of training set is divided into 64 × 64 image blocks. As shown in Fig. 5, First, the lung CT image is processed to obtain the pulmonary parenchyma region. Then, a 64×64 binary mask is obtained centering on the pulmonary nodule in the label. Finally, according to the mask, the corresponding pulmonary parenchyma image block and label block are obtained from the pulmonary parenchyma image and label image respectively. The training set consists of all these 64×64 image blocks containing details of pulmonary nodule and corresponding label images. In Fig. 6(1) and (2), (a) shows the pulmonary parenchyma region obtained by morphological processing, and (b) is a 64×64 binary image calculated from the given nodular center coordinate, whose center is the pulmonary nodule. (c) represents the input image of the net obtained by (a) and (b). (d) is a label image corresponding to the input image of the net.

For a total of 1000 patient cases, 1245 lung nodules were labeled. The whole dataset is often randomly split into training,
validation and test sets with the ratio of 8:1. Here, we take 975 lung nodules as training set and 130 lung nodules as validation set and 140 lung nodules as testing set. In addition, the 140 samples in the test Set were divided and compared by Level Set and Graph Cut respectively, and then the statistical comparison was conducted according to the evaluation indicators.

Table 2
Computer configurations.

<table>
<thead>
<tr>
<th>Item</th>
<th>Configuration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processor</td>
<td>Intel Core i5-3570 <a href="mailto:CPU@3.4Hz">CPU@3.4Hz</a>*4</td>
</tr>
<tr>
<td>Graphics card</td>
<td>GTX660/PCIe/SSE2</td>
</tr>
<tr>
<td>RAM</td>
<td>8G</td>
</tr>
<tr>
<td>Hard disk</td>
<td>500G</td>
</tr>
</tbody>
</table>

Table 3
Test results of two different structures (Bold indicates the best results).

<table>
<thead>
<tr>
<th>Model</th>
<th>Dice coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>U-NET</td>
<td>0.713</td>
</tr>
<tr>
<td>Our method</td>
<td>0.736</td>
</tr>
<tr>
<td>Level Set</td>
<td>0.601</td>
</tr>
<tr>
<td>Graph Cut</td>
<td>0.695</td>
</tr>
</tbody>
</table>

Fig. 5. The flowchart of the dataset production processing.

Fig. 6. Input image and label.
3.2. Evaluating indicator

The loss function is mainly used to evaluate the inconsistency between the predicted value and the real value. If the loss function is relatively small, the corresponding model is better, and the robustness is also stronger. The loss function in this paper is:

Table 4  
Training parameters for two different networks.

<table>
<thead>
<tr>
<th>Network model</th>
<th>Training parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>U-NET</td>
<td>7,848,130</td>
</tr>
<tr>
<td>Our method</td>
<td>8,301,439</td>
</tr>
</tbody>
</table>

3.2. Evaluating indicator

The loss function is mainly used to evaluate the inconsistency between the predicted value and the real value. If the loss function is relatively small, the corresponding model is better, and the robustness is also stronger. The loss function in this paper is:
\[ L(X, Y) = \frac{1}{N} \sum_{i=1}^{N} S(X, Y) \tag{8} \]

Where X is the predictive value of the model and Y is the true value. \( S(X, Y) \) is as follows,

\[ S(X, Y) = 1 - \frac{2|X \cap Y| + k}{|X| + |Y| + k} \tag{9} \]

In order to prevent the case that there is no pulmonary nodules in CT images, smoothing coefficient K is added to the loss function. When the loss function is relatively good, the robustness of the model will be better.

The evaluation criterion adopted in this paper is the Dice coefficient, which is used to evaluate the quality of the model segmentation results. The Dice coefficient mainly evaluates the similarity between two templates, and it is an evaluation function of the similarity between the sets. If the two sets are similar, the dice coefficient will be larger. Therefore, the larger the Dice coefficient, the more similar two images are, and the corresponding segmentation effect is more accurate.

### 3.3. Analysis of results

The experiment of the segmentation algorithm in this paper is carried out under the environment of 64-bit Ubuntu16.04 operating system. Table 2 shows the configuration of the computer.

Due to the limitations of computer computing ability, the batch size used in this article is 1. We use Adam to optimize and segment processing learning rate when training. Specifically, a fixed value learning rate is set when training the first 100 epochs. Then during the 100–500 epochs, the learning rate is 0.1 times of the previous one, namely 0.1 \( \eta \). And so on, you can solve the problem of choice of learning rate. By analogy, the problem of choosing the learning rate can be solved. In this paper, the learning rate of the network is set to 0.001 firstly, and the loss value is displayed when training. When the loss value is no longer reduced for many times, the learning rate is reduced to 0.1 times that of the previous one. In this way, the learning rate is segmented and it will better train the network.

(1) Analysis of test results

The result of the prediction is compared with the real nodule image, and the final evaluation results are shown in Table 3.

It can be seen from the table that our improved U-NET network has a better segmentation effect compared to the original U-NET network. And our algorithm is better than the other lung nodules segmenting method such as Level Set [15] and Graph Cut [16].

As shown in Fig. 7, there are there complete CT slice images segmentation results using our net to segment. The first row is the complete CT slice image, and the second row is the segmentation result. The segmentation results of our improved U-NET compared with U-NET, Level set, Graph cut and Ground Truth for the CT slice block images are shown in Fig. 8.

(2) Evaluation of training parameters

The relationship between the training parameters of the U-NET network and the improved U-NET network can be seen from Table 4. Since the improved U-NET network has introduced the residual network, the depth of the network is deepened to some extent. Therefore, the parameters of the improved U-NET network are increased. The improved U-NET network has introduced the batch Normalization processing method. As shown in Fig. 9, the loss of our method and U-Net is tend to stabilize at a similar rate. While, the Loss optimization rate of our method is faster than the U-Net. Thus, the increased parameters have little influence on the model training process.
3.4. Interface implementation

The system interface is built to show the experimental results more intuitively. The structure of the visualization system is mainly based on the Tkinter of Python. The GUI built with Tkinter is shown in Fig. 10. It can display the original image, the segmentation of pulmonary parenchyma, the image of 64*64 in size, the pulmonary nodule and the final predicted image.

4. Conclusion

In this paper, we mainly propose a pulmonary nodule segmentation algorithm based on the improved U-NET network. The algorithm mainly includes image preprocessing, lung segmentation to obtain lung parenchyma and using improved U-NET network to get a model. By designing the U-NET network and introducing the residual network, the depth of the network is deepened to a certain extent, so that the training effect of the network is better. In addition, the introduction of a batch of standardized operations speeds up the network training and a greater learning rate can be used. This will facilitate the network training and alleviate the problem of gradient disappearance or gradient explosion. Experiments show that our method has a better effect for pulmonary nodules segmentation. This work still needs to be improved:
(1) The experimental data used in this paper is from the published data set on LUNA2016, but the sample size is relatively insufficient. Consider increasing the sample size of pulmonary nodules in the future, and then use the improved U-NET network for training to further improve the training effect of the network.

(2) This article mainly focuses on the segmentation of pulmonary nodules, pulmonary nodules can be further studied whether it is malignancy or benign also. This is a better way to detect lung cancer and help more patients find lung cancer as soon as possible.

(3) This article uses a two-dimensional image based on the U-NET network to train, which may cause some loss to the original three-dimensional CT image data. As computing power increases, follow-up studies can be performed using three-dimensional convolutional networks and using the three-dimensional data of lung nodules.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (No. 61005034), the Natural Science Foundation of Hebei Province (No. F2012203185).

References


