Lung Tumor Segmentation Algorithm Based on Deep Convolution Neural Network

CHEN Shu[1,2], JI Jianbing [2], YANG Yuanyuan[3*]

1.College of Physics and Information Engineering, Fuzhou University, Fuzhou 350108, China

2.College of Information Engineering, Fujian Business University, Fuzhou 350012, China

3.Department of General Surgery, Fujian Medical University Union Hospital, Fuzhou 350001, China

*Corresponding author E-mail: spring0086@163.com

Keywords: Cnn, Lung tumor, Segmentation

Abstract: In order to segment lung tumor in CT image automatically, a segmentation method based on improved 3DUnet is proposed. The convolution kernel size is increased, PReLU is used as the activation function, the training set is expanded and optimized by nonlinear transformation, the network is trained by DICE loss function and SGDM gradient descent algorithm, post-processing is added to optimize the segmentation results. The experiment is carried out by using medical segmentation decathlon open data set and 5 indicators are used for evaluation. Results show that the algorithm achieves high accuracy and is better than 3DUnet network.

1. Introduction

In this study, the CT scan image sequence was used as the source of image data and the three-dimensional visualization of lung tumors was displayed through image segmentation, with the purpose to assist doctors in the diagnosis and analysis of tumors. The accuracy of segmentation directly affects the decision-making of the subsequent surgical plan, and thus image segmentation of lung tumors will bring huge social benefits to the formulation of preoperative plans and to reduce the difficulty of doctor-patient communication.

Researchers have proposed methods for the automatic segmentation of lung tumors, respectively based on K-means clustering algorithm [1], on random walk algorithm [2], and on machine learning. The method of 3DUnet convolutional neural network (CNN) segmentation proposed by Uday Kamal n 2018 presented a high accuracy rate [3]. Different from other methods, the method in this research adopts volume segmentation, which can not only effectively solve the problem of importing the image sequence slice by slice into model training, but also greatly improve the training efficiency. And the segmentation result is a three-dimensional matrix, which is convenient for three-dimensional visualization of the data.

This study uses improved 3DUnet network to segment lung tumors in CT images. First of all, the researcher replaced the 3x3x3 convolution kernel in 3DUnet with a convolution kernel sized 5x5x5, replaced ReLU with PReLU, and replaced the weighted cross-entropy loss with a DICE loss function. Second, random nonlinear transformation was added to expand and optimize the training data, and post-processing was used to optimize the segmentation results. The public data set was used to verify the performance of the algorithm.

  DOI: 10.25236/mmmce.2021.026

2. Network Structure

3DUnet is a three-dimensional fully convolutional network proposed on the basis of a fully convolutional neural network (CNN) [4], which presents a left-right symmetrical structure. The left half of it is down-sampling, that is, the feature extraction process is composed of convolutional layer units. The right half is up-sampling, that is, the feature reconstruction process is composed of deconvolution units. The convolution unit is used for feature extraction, and the deconvolution unit is used to restore the image size. The network structure is shown in Figure 1.

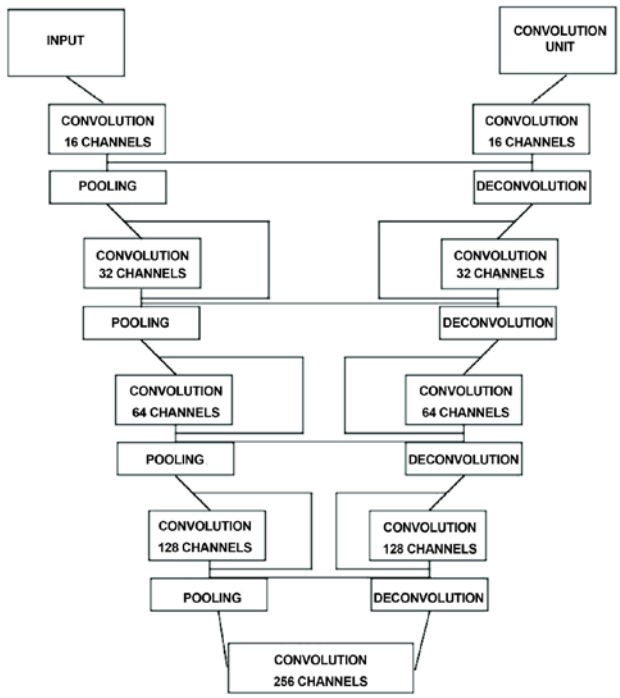


Fig.1 Network Structure

This research was carried out on the basis of the basic structure of 3DUnet. In view of the large differences between individual lung tumors, it is necessary to retain as much effective information as possible during the convolution process. First, the size of the convolution kernel that was originally 3x3x3 in the convolution unit was increased to 5x5x5. The change of this item extended the visible range of the convolution larger, which not only was conducive to more information after the convolution, but also could increase the parameters of the network. Second, ReLU was replaced with PReLU as the activation function. This change retained more input details that were less than 0 after convolution.

The structure of each part of the network is explained as follows:

(1) Convolution unit

The convolution unit consists of $Conv-BN-P\mathrm{Re}LU-Pool$. The size of the input data is set to be $A=[m,n,d]$, the size of the convolution kernel is f, the step size is s, the padding is p, and the size of the pooling convolution kernel is k.

The output size relationship of the data after three-dimensional convolution is:

$$\mathrm{Conv}(i)=floor(\frac{A(i)-f+2*p}{s}+1) \text{ (Equation 1)}$$

All the convolution configurations are filled with the SAME method, that is, when the step size is 1, the size of the output feature map and that of the input feature map remain

unchanged. Two sets of convolution kernel standards $\{f=5, s=1\}$ and $\{f=2, s=2\}$ are used in this research. When the data passes through the convolution kernel $\{f=2, s=2\}$, its resolution will be reduced by half (through Formula 1), and a deeper feature with the mapping size halved will be generated.

BN (Batch Normalization) converts the data to obey the standard normal distribution, with the purpose to speed up the convergence speed. The main calculation process includes calculating the sample mean, calculating the sample variance, standardizing the sample data, and performing translation and zoom processing.

$P\mathrm{Re}LU$ (Parametric Rectified Linear Unit) is the activation function, and the output relationship is:

$$P\mathrm{Re}LU(x_i)=\begin{cases} x_i, x_i>0 \\ a_i x_i, x_i \le 0 \end{cases} \quad \text{(Equation 2)}$$

When the input x is greater than 0, PReLU degenerates to ReLU. In Formula 2, α is a small constant that is automatically calculated when the network feeds back. Different from ReLU, this algorithm achieves the purpose of activating functions while retaining part of the information less than 0.

Pool (Convolutional pooling) is used to reduce the dimensionality of features, and the output size is:

$$Pool(i)=ceil(\frac{X(i)-k+2*p}{s}+1) \text{ (Equation 3)}$$

(2) Deconvolution unit

The deconvolution unit is composed of the deconvolution + convolution unit. When the data is deconvolved with the convolution kernel $\{f=2, s=2\}$, its resolution will be doubled to restore the size. Then there is one to three convolution units, and the size of the convolution layer kernel of the last layer is 1x1x1.

(3) Regression model

The network prediction consists of two voxels which have the same resolution as the original input. The Softmax regression model outputs the probability that each voxel is divided into a background area and a target area. When inputting $z=(z_1, z_2, ...z_n)$, the Softmax function maps it to $\sigma=(\sigma_1, \sigma_2, ...\sigma_n)$, and the mapping relationship is:

$$\sigma_i(z)=\frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}}, i=1,2,...n \quad \text{(Equation 4)}$$

In this equation, $\sum \sigma_i = 1$

3. Lung Tumor Segmentation

3.1 Algorithm Flow

The researcher divided the data set into training set and test set after preprocessing, and then imported it into the network for training, and finally used the network to predict the output of the test set. After post-processing on the prediction results, the entire segmentation model is output, and the algorithm flow is shown in Figure 2.

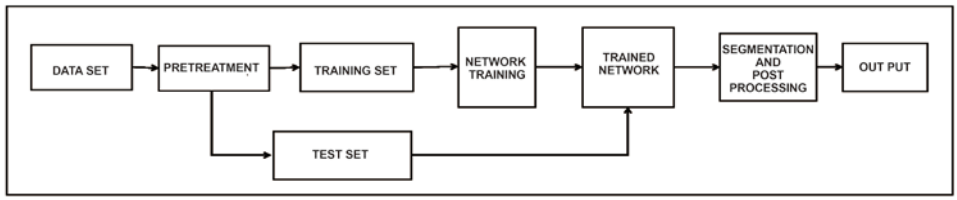


Fig.2 Algorithm Flow

3.2 Pretreatment

The data set is composed of CT image sequence and label sequence, and the preprocessing mainly includes resolution standardization and gray-scale standardization. First of all, the CT image sequence and the label sequence were synthesized into a three-dimensional matrix, and then the three directions (x, y, z) were expanded by 48 pixels according to the size of the label body. Each dimension was cut to a multiple of 8, and the cutting method was “remainder removal”. Then, it was scaled by the window size, and the scaling method was linear interpolation. Due to the different parameters of CT scan, the gray value distribution of each slice in the image matrix was inconsistent. In this study, the gray value in the image slice was subtracted from the mean gray level and then divided by the standard deviation of the gray level to perform normalization.

3.3 Training Network

3.3.1 Expansion and Optimization of the Training Set

Regard to limited training sets, the researcher expanded the training data by random non-linear transformation. The method group of nonlinear transformation mainly included rotating 90 degrees counterclockwise, mirroring left and right, mirroring up and down, and rotating 90 degrees counterclockwise after mirroring left and right. Each set of training data randomly selected a method transformation in the non-linear transformation method groups. In addition to expanding the training data set, the transformation result also enhanced the robustness of the network's predictive ability.

3.3.2 Loss Function

The loss function was used to measure the gap between the predicted value and the true data value obtained after the forward pass. In order to avoid the unstable training process caused by the extremely unbalanced data, Dice loss was used to replace the weighted cross-entropy loss function in this research. The DICE loss algorithm described the degree of difference between the two sets through the Dice coefficient. The Dice coefficient between the algorithm segmentation result *p* and the real mark can be expressed as:

$$D = \frac{2\sum_i^N p_i g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2} \qquad \text{(Equation 5)}$$

When the N voxels of the predicted segmentation result exceed the sum of the actual labeled N voxels, the gradient generated by the Dice coefficient corresponding to the predicted j^{th} voxel is:

$$\frac{\partial D}{\partial p_j} = 2\left[\frac{g_j(\sum_i^N p_i^2 + \sum_i^N g_i^2) - 2p_j(\sum_i^N p_i g_i)}{\sum_i^N p_i^2 + \sum_i^N g_i^2}\right]$$

(Equation 6)

3.3.3 Gradient Descent Algorithm

Gradient descent is a method to find the minimum value of the loss function. In the process of network training, gradient descent is often used to adjust the parameters to find the small value of the loss function. In this research, Sgdm (stochastic gradient descent algorithm with momentum) was used to realize the network training. This method was designed to accelerate learning, adding inertia to the gradient descent process, and determining the direction of descent through momentum accumulation.

Suppose the learning rate is ε, the initial parameter is θ, the momentum parameter is α, and the initial speed is v. And the rule of algorithm update is as follows:

while (stop condition is not satisfied)

A mini-batch of m samples $\{x^{(1)},...x^{(m)}\}$ is sampled from the training set, and $x^{(i)}$ corresponds to $y^{(i)}$;

Gradient estimation $g = \frac{1}{m}\nabla_{\theta}\sum_{i} L(f(x^{(i)},\theta), y^{(i)})$;

Speed update $v = \alpha v - \varepsilon g$;

Parameter update $\theta = \theta + v$;

end while

3.4 Post-Processing

The researcher input the test sets into the trained network and performed network segmentation on the input, and then a three-dimensional binary matrix of the same size was output. In order to optimize the segmentation results and eliminate the salt-and-pepper noise generated by the probability segmentation, three-dimensional median filtering was adopted for post-processing in this research.

4. Results

4.1 Experimental Data

The Medical Segmentation Decathlon public data set was used in the experiment, and the data set contained a total of 96 labeled chest CT images with lung tumors. 90 cases were randomly selected as the training set, and the rest 6 cases were used for algorithm testing.

4.2 Evaluation Methods of the Experiment

The following five evaluation indicators were adopted in this research. P represents algorithmic segmented volume data, and G represents real labeled volume data. Evaluation indicator are defined as follows:

(1) DICE: the ratio of the volume where two objects intersect to the total volume.

$$DICE = \frac{2*(P \cap G)}{P+G} \qquad \text{(Equation 7)}$$

(2) VOE: volume overlap error.

$$VOE = 1 - \frac{P \cap G}{P \cup G} \qquad \text{(Equation 8)}$$

(3) RVD: relative volume difference.

$$RVD = \left(\frac{P}{G} - 1\right) * 100\% \qquad \text{(Equation 9)}$$

(4) ASD: Average Symmetric Surface Distance. In this research, $S(P)$ and $S(G)$ respectively denoted the collection of surface voxels of P and G; the distance from any point v on $S(P)$ to $S(G)$ was defined as the minimum Euclidean distance from point v to all points on $S(G)$, expressed as $d(v,S(P))=\min_{S_p\in S(P)}(\|v-S_p\|)$. Then the average symmetrical surface distance was defined as:

$$ASD(P,G)=\frac{1}{|S(P)|+|S(G)|}\times\left(\sum_{S_P\in S(P)}d(S_P,S(G))+\sum_{S_G\in S(G)}d(S_G,S(P))\right)\text{(Equation10)}$$

(5) MSD: Maximum Symmetric Distance.

$$MSD(P,G)=\max\left\{\max_{S_P\in S(P)}d(S_P,S(G)),\max_{S_G\in S(G)}d(S_G,S(P))\right\}\text{(Equation10)}$$

Among the five evaluation indicators, the closer DICE is to 1, the better the segmentation, and the smaller the other four indicators, the better the segmentation effect.

4.3 Experimental Design

The initial learning rate was set to 0.001, and the number of training iterations was set to 200; the training was stopped when the loss function output was obviously stable.

4.4 Experimental Results

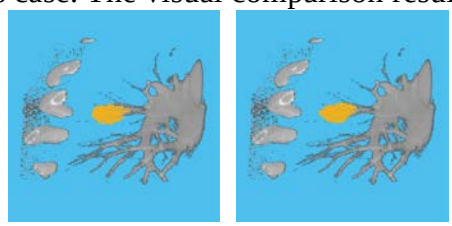
According to the Medical Segmentation Decathlon data experiment, the average test results of the experiment are shown in Table 1.

Table 1 Experimental Results

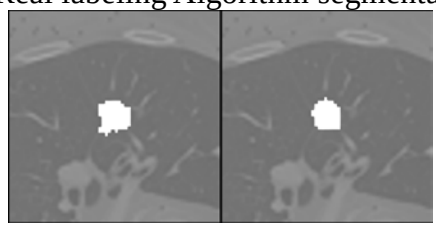
Experiment	DICE	VOE	RVD	ASD	MSD
Lung tumor	0.716	0.429	-0.339	1.032	3.534

It can be seen from Table 1 that the algorithm in this research can segment lung tumors very accurately, and the other four indicators show that the algorithm is stable and reliable.

The researcher randomly selected a case of CT data from the test set and randomly sliced the segmentation results of this case. The visual comparison results are shown in Figure 2.



(a) Real labeling Algorithm segmentation



(b) Comparison of real labeling and algorithmic segmentation of a random slice

Fig.2 Visual Comparison

5. Discussion

The experimental results show that the improved 3DUnet network can segment lung tumors accurately, and it can be seen from Table 1 that the average accuracy of the experiment is 71.56. In order to compare the segmentation performance of the algorithm, 3DUnet was used for verification on the same data set. The verification results show that the method in this paper is superior to the 3DUnet network in various indicators, and the comparison results of the algorithms are shown in Table 2.

Table 2 Algorithm Comparison

Experiment	DICE	VOE	RVD	ASD	MSD
Method in this research	0.716	0.429	-0.339	1.032	3.534
3DUnet	0.695	0.446	-0.387	1.103	8.794

Due to the large differences in the individual shape of lung tumors, the network still presents a certain degree of uncertainty in predicting segmentation, especially in terms of the segmentation results of tumor spread. Therefore, the improvement of segmentation accuracy needs to be supported by larger training data, and the marking of the training set needs to be completed under the guidance of professional doctors. In addition, the number of samples of public data sets with real labels is relatively limited, and thus building a larger amount of training data and optimizing the network based on it will be the focus of research in the next stage.

6. Conclusion

Based on the improved 3DUnet CNN, a lung tumor segmentation method was proposed in this research through increasing the size of the convolution kernel, using PReLU as the activation function, and adding random nonlinear transformation to the training set. In addition, the researcher trained the network with the DICE loss function and the Sgdm gradient descent algorithm, post-processed the segmentation results with three-dimensional median filtering, and verified the accuracy of the segmentation algorithm with public data sets. Subsequent improvement goals and related strategies were also proposed based on the discussion of the experimental results in this research.

Acknowledgment

Fund Project: supported by Fujian Provincial Young and Middle-aged Teacher Education Research Project of the Education Department of Fujian Province: "*Research on the Segmentation Method of Lung Tumor Based on Convolutional Neural Network*" (No. JAT190017).

References

[1] LI Ling, YU Houqiang. Retrieval of Lung Tumor CT Images Based on K-means Clustering Algorithm [J]. Journal of Gansu Sciences, 2015, 27(1):58-61.

[2] ZHANG Xin, WANG Jie, KONG Hang, et al. Large lung tumor segmentation in CT images based on random walk algorithm [J]. Journal of Hebei University (Natural Science Edition), 2019, 39(3):311-315.

[3] Kamal, U. et al. Lung cancer tumor region segmentation using recurrent 3D-DenseUNet.Preprint at https://arXiv.org/abs/1812.01951 (2018).

[4] Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation[J].2016.

[5] Ronneberger O, Fischer P, Brox TU-Net: convolutional networks for biomedical image segmentation[J]. Lecture Notes in Computer Science,2015,9351:234-241.

[6] SUN Mingjian, XU Jun, MA Wei, et al. A new fully convolutional network for 3D liver region segmentation on CT images[J].Chinese Journal of Biomedical ngineeering,2018,37(4):4-12.

[7] TIAN Xuan, WANG Liang, DING Qi. Review of image semantic segmentation based on deep learning[J].Journal of Software,2019,30(2):440−468

[8] TANG Si-yuan, XING Jun-feng, YANG Min. New method for medical image segmentation based on BP neural network. COMPUTER SCIENCE,2017,44(6):250-253.

[9] LIU Fei, ZHANG Junran, YANG Hao. Research progress of medical images recongnition based on deep learning [J]. Chinese Journal of Biomedical Engineering, 2018,37(1).

[10] ZHAO Ziqi, PEI Yun, CHANG Zhendong, et al. Lung nodules based on deep learning. Journal of Jilin University (Information Science Edition),2019,37(05):572-581.

[11] ZHOU Yan, ZENG Fanzhi, WU Chen, et al. 3D shape feature extraction method based on deep learning. COMPUTER SCIENCE,2019,46(9):47-58.

[12] Currie Geoff, Hawk K Elizabeth, Rohren Eric, et al. Machine Learning and Deep Learning in Medical Imaging: Intelligent Imaging.[J].Journal of medical imaging and radiation sciences,2019.

[13] A. Fourcade, R. H. Khonsari. Deep learning in medical image analysis: A third eye for doctors[J]. Journal of Stomatology oral and Maxillofacial Surgery,2019,120(4).

[14] Daoud Bilel, Morooka Ken'ichi,Kurazume Ryo,et al.3D segmentation of nasopharyngeal carcinoma from CT images using cascade deep learning.[J].Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society,2019,77.

[15] LI Xiaomeng, CHEN Hao, QI Xiaojuan, et al. H-Dense UNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation From CT Volumes.[J]. IEEE transactions on medical imaging,2018,37(12).