Ultrasound-assisted diagnosis of benign and malignant cervical lymph nodes in patients with lung cancer based on deep learning

Limin Xia, Shanglin Lei, Huiyao Chen, Haoyang Wang

School of Automation, Central South University, Changsha, Hunan, 410083, China.

Keywords: Deep learning; noise excitation function; medical image.

Abstract: Deep learning technology assisted ultrasound image diagnosis can improve the accuracy and efficiency of detection. In this paper we propose an improved U-Net convolution network for ultrasound image segmentation. The network replaces the noise excitation function NHReLU and NHSeLU with the noise excitation function ReLU, and adds weight parameters to the cost function. By predicting on two scales, it handles the problem of the size change of the marked area in the ultrasound image well, and improves the segmentation effect of the lymph node ultrasound image. Use networks such as VGG, ResNet, and DenseNet to predict the benign and malignant areas of lymph node lesions. Experiments show that the segmentation network has excellent performance, its Dice coefficient reaches 0.90, and the model can well prevent overfitting. In addition, the prediction of benign and malignant indicators under a small sample has been improved, which provides a new method for the application of deep learning technology in ultrasound image detection.

1. Introduction

Cervical lymph node metastasis, especially supraclavicular lymph node metastasis, is an important form of metastasis of lung cancer. Lymph node metastasis is directly related to the stage, operation and prognosis of lung cancer, so the differential diagnosis is very important. Ultrasonic imaging equipment is widely used in modern medical testing because of its low cost, portability, non-trauma and non-radiation.Traditional medical image analysis is based on the subjective evaluation of the image obtained by the doctor, which is time-consuming. The accuracy of analyzing medical images depends on the experience of the operator, and the evaluation results are often subjective.In recent years, computer-aided differentiation between metastatic lymph node lesions and benign lesions of lung cancer has an important clinical application prospect, and the application of deep learning technology in image detection has made great progress. In the fields of classification , segmentation and target detection, the accuracy is better than that of human. Therefore, the application of deep learning to ultrasonic image analysis has a strong theoretical basis. The deep learning technology is applied to the detection of medical images (CT, MRI and PET). Through a large number of labeled data, the image analysis model is trained, which greatly improves the diagnostic efficiency and accuracy of doctors. The resolution of ultrasonic image is lower than that of CT and MRI, and there are a lot of artifacts and noise in ultrasonic image. there is a large space for the application of deep learning technology in ultrasonic image analysis, which has important theoretical significance and practical application value[1].

At present, the common deep learning models used in ultrasonic image segmentation, such as FCN (FullyConvolutionalNetworks), UNet, etc., have achieved some research results, but the noise excitation function is applied to the network structure, and the improvement of network performance is worthy of in-depth research. When the amount of data is small, the traditional machine learning method is usually used, and the deep learning model needs a lot of data to train, otherwise the model will be overfitted. How to make the model keep the high accuracy of benign and malignant classification when dealing with small samples is also a difficult point in the research.

2. Method

2.1 Transfer learning

Transfer learning technology refers to pre-training examples such as VGGNet model parameters as initialization parameters on the existing image data set, and then using the labeled ultrasound image data set to greatly speed up the training speed of the network. At the same time, the over-fitting phenomenon caused by the lack of training data is effectively avoided, and the ultrasonic image recognition result of depth network recognition is improved. A more elaborate approach is to use networks pre-trained on large data sets (usually trained on ImageNet data sets), which will have learned features that are useful for most computer vision problems, and using these features will enable us to achieve better accuracy than any method that relies solely on available data. A transfer learning strategy is proposed in reference , which transfers the trained basic CNN knowledge from a large number of natural image databases to task-specific CNN. Literature established a diagnostic tool based on deep learning framework for screening patients with common treatable blinding retinal diseases. The framework uses transfer learning to train a small portion of the data of neural networks and traditional methods. The labeled data set reached 110,000[2].

This paper uses two transfer learning methods, one is to use the bottleneck feature of the pre-trained convolution neural network (bottleneckfeatures) to predict the benign and malignant lymph nodes; the other is to fine-tune the top layer of the pre-trained convolution neural network to predict the benign and malignant lymph nodes.

Using bottleneck features of pre-trained convolutional neural network to predict lymph node benign and malignant

Fine-tuning the top layer of the pre-trained convolution neural network to predict the benign and malignant lymph nodes

Fine-tuning should be done at a very slow learning speed, usually using an SGD optimizer rather than an adaptive learning rate optimizer like rms. This is to ensure that the size of the update is kept very small so that it does not break the features you have learned before.

2.2 Data enhancement

Data enhancement is a technology that is used to fine-tune the parameters of the pre-trained network model in order to prevent over-fitting. When collecting data to fine-tune the deep learning model, we often encounter a serious shortage of classified data. When a model is trained on a training set with too few examples, the learned model can not be generalized to new data, that is, when the model begins to use irrelevant characteristics for prediction, overfitting will occur. Therefore, data enhancement is to transform the image of region of interest (ROI) horizontally and vertically from top to bottom, left and right, to enlarge the size of the data set, so that our model will not see exactly the same image, which helps to prevent over-fitting and contribute to the generalization of the model.

Data enhancement is a good way to combat over-fitting, but it is not enough because our enhanced samples are still highly correlated. The main concern for over-fitting is the entropy capacity of the model-how much information the model is allowed to store. A model that can store a lot of information can predict more accurately by using more features, but it is also more likely to start storing extraneous features.At the same time, a model that can only store some features will have to focus on the most important features in the data, which are more likely to be truly relevant and can be better generalized.

There are different ways to adjust entropy capacity. The most important thing is to select the number of parameters in the model, that is, the number of layers and the size of each layer. Another method is to use weight regularization, such as L1 or L2 regularization, which enables the model weight to take a smaller value. In the process of building a convolution network structure, inserting the dropout layer also helps to reduce overfitting, preventing the same pattern from repeating layer by layer, thus operating in a manner similar to data enhancement (both Dropout and Data Augmentation tend to destroy random correlations that occur in the data).

2.3 UNET

The UNet network makes use of the inherent multi-scale characteristics of the neural network, the shallow output preserves the spatial detail information, and the deep output preserves the relatively abstract semantic information. The underlying information is used to supplement the high-level information, which is suitable for medical image segmentation and natural image generation. It has better segmentation accuracy than RPN and FCN network in medical image segmentation. On the left side of the UNet model is a downsampling process, which is divided into four groups of convolution operations; the right upsampling process uses four groups of deconvolution, each of which expands the image to twice the original value, then clips and copies the image (feature map) of the corresponding layer, and then concat to the result of the upper convolution[3].
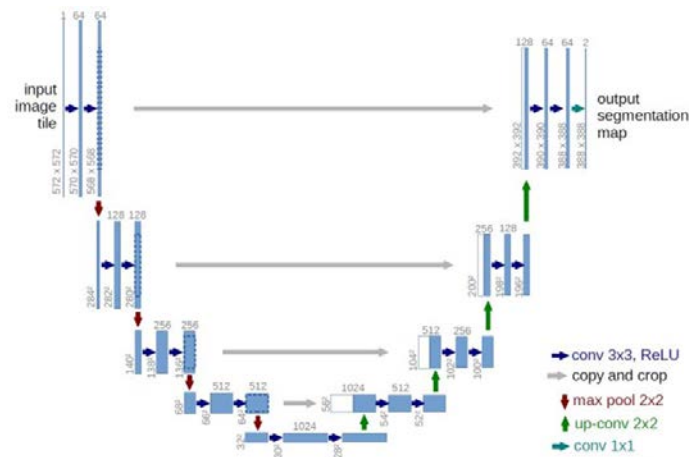

Fig 1 Unet

The UNet network makes use of the inherent multi-scale characteristics of the neural network, the shallow output preserves the spatial detail information, and the deep output preserves the relatively abstract semantic information. The underlying information is used to supplement the high-level information, which is suitable for medical image segmentation and natural image generation. It has better segmentation accuracy than RPN and FCN network in medical image segmentation. On the left side of the UNet model is a downsampling process, which is divided into four groups of convolution operations; the right upsampling process uses four groups of deconvolution, each of which expands the image to twice the original value, then clips and copies the image (feature map) of the corresponding layer, and then concat to the result of the upper convolution.

3. Empirical analysis

In this study, 420 lymph node ultrasound images were collected from 360 patients with lung cancer in the Ultrasound Department of Shanghai Thoracic Hospital, including 190 males, 230 lymph nodes, 170 females and 190 lymph nodes. All lymph nodes were examined by fine needle aspiration cytology and fine needle aspiration biopsy, and the results of ultrasonic diagnosis were compared with pathological results. The collected ultrasound image data of cervical lymph nodes are binary images. The lymph node ultrasound image data set is divided into training set, verification set and test set. Each image is scanned by special ultrasound doctors and marked with the long diameter, short diameter, benign and malignant of the diseased cells.

Using open source image annotation tools Labelme and VGG Image Annotator to re-label and extract the mask of each lymph node for training. Because the malignant is in the majority, that is, there is the problem of data imbalance, so the data enhancement method is used to expand the training data for each category. There are two groups of enhanced data, which are trained and verified respectively. The first group of data (a total of 3000 pieces) randomly sampled 2000 pieces as the training set, and the rest as the verification set, during which several groups of training samples and verification samples were regenerated to test the network performance. The second

group of data (5000 pieces) randomly sampled 3000 pieces as the training set and the rest as the verification set.

The methods of data enhancement are rotation 90 degrees (can be less than 90 degrees), random horizontal movement 0.2 (compared to the original picture width), random vertical movement 0.2, shear strength 0.2 (counterclockwise shear), random scaling 0.2. Random channel offset 0.2, random horizontal flip, random vertical flip. Borderline-SMOTE2 (Synthetic Minority Oversampling Technique), is used to synthesize minority class oversampling technology, analyze minority class samples and artificially synthesize new samples according to minority class samples to add to the data set. The noise added to the noise excitation function is 0.05 and the noise standard deviation is 0.5.

3.1 Lymph node focus segmentation

The indexes of image, UNet downsampling and upsampling convolution layer used for training are clear, and the Deiss coefficient (DiceCoefficient) is used to evaluate the similarity between the image segmentation results and the original label data. The Dice coefficient of ReLU is 0.85, and the Dice coefficient of NHReLU and SeLU is 0.86. The Dice coefficient and loss function of NHReLU are shown in figure
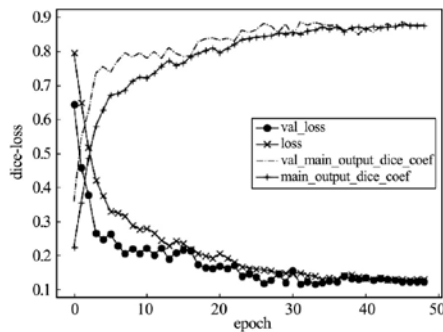

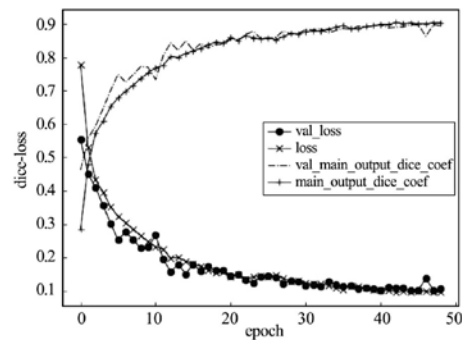Fig 2 NHReLU Dice coefficient and loss function

Fig 3 NHSeLUDice coefficient and loss function

Through experiments, we can conclude that the model uses NHSeLU noise excitation function better than ReLU, SeLU and NHReLU segmentation effect, the Dice coefficient reaches 0.9. The Dice coefficient and loss function using NHSeLU are shown in Figure 3.

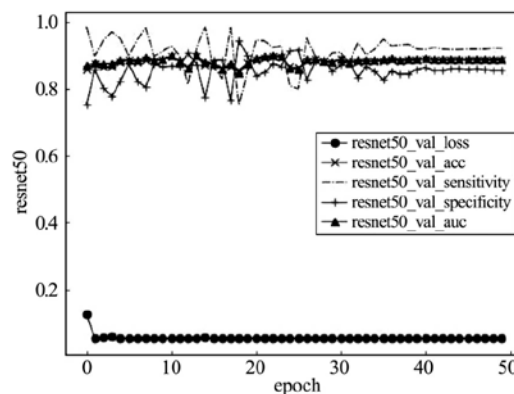### 3.2 Lymph node focus segmentation
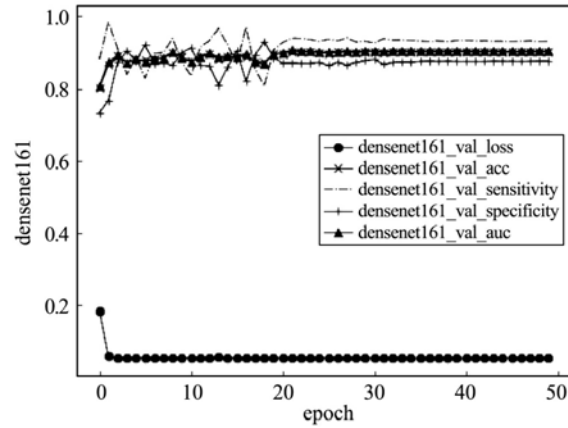

Fig 4 the value of each index of ResNet50

Fig 5 the value of each index of DenseNet161

The results of ResNet50 experiment (enhanced 3000 samples): the sensitivity, specificity and accuracy were 86.7%, 82.8% and 84%, respectively. The AUC was 0.848. The results of DenseNet161 experiment (enhanced 3000 samples): the sensitivity, specificity and accuracy were 92%, 80% and 86.5%, respectively. The AUC was 0.861.

The results of ResNet50 experiment (enhanced 5000 samples): the sensitivity, specificity and accuracy were 92%, 86% and 89%, respectively, and the AUC was 0.89, as shown in figure 4. The results of DenseNet161 experiment (enhanced 5000 samples): the sensitivity, specificity and accuracy were 93%, 88% and 90%, respectively, and the AUC was 0.90, as shown in figure 5.

The accuracy of ResNet50 is compared with that of DenseNet161 (enhanced 5000 samples) and AUC, as shown in figure 6. It can be seen that DenseNet161 is better than ResNet50 network. The experimental results show that the training accuracy and Deiss coefficient are higher than those on the verification set, mainly because the amount of data is not enough. In the future, the sample size should be further increased to avoid over-fitting of the model.
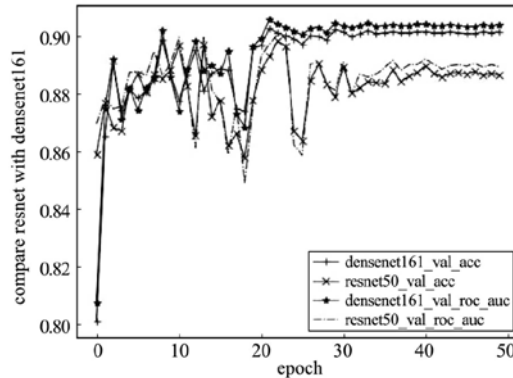

Fig 6 Comparison of accuracy and AUC between ResNet50 and DenseNet161

The experimental results show that the network can achieve rapid and accurate segmentation of cervical lymph node ultrasound images of lung cancer patients. The sensitivity, specificity and accuracy of the segmentation results are 93%, 88% and 90%, respectively, and AUC is 0.90. These quantitative indicators will greatly help differentiate metastatic lymph nodes from benign lymph nodes in patients with lung cancer, and also verify the effectiveness of the model. Some scholars use the mean-max pooling layer to improve the accuracy, but this method does not improve the accuracy of the experiment in this paper, which is also worth further study.

4. Empirical analysis

In this paper, ultrasonic images are segmented based on UNet network, and a variety of noise excitation functions are introduced to improve the performance of the network and improve the Deiss coefficient; VGG, ResNet and DenseNet networks are used to predict benign and malignant lymph node lesions; transfer learning, data enhancement and loss function are used to prevent network overfitting and improve the accuracy, sensitivity and specificity of predicting benign and malignant lymph nodes under small samples. There are two directions worthy of further research in the future: (1) in order to improve the generalization ability of the model and increase the diversity of time, the generating countermeasure network (GAN), should be used to generate new positive samples from random noise and learn to generate new forms of samples; (2) the deep learning model in this paper uses a single network to predict the benign and malignant classification, and does not use the model fusion technology. Therefore, the detection uses a variety of model fusion to predict the benign and malignant probability of the lesion site respectively. The method of obtaining the final probability according to the weight ratio between models is worthy of in-depth study.

References

[1] U-net:Convolutional networks for biomedical image segmentation. Ronneberger O,Fischer P,Brox T. International Conference on Medical image computing and computer-assisted intervention . 2015

[2] Noisy activation functions. Gulcehre C,Moczulski M,Denil M,et al. 33rd International Conference on Machine Learning(ICML 2016) . 2016

[3] Aggregated Residual Transformations for Deep Neural Networks. Xie S,Girshick R,Dollar P,et al. 2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR 2017) . 2017