3DCNN for Lung Nodule Detection And False Positive Reduction

Ping An Technology (Shenzhen) Co.,Ltd, China

WUTIANBO484@pingan.com.cn

Abstract

The diagnosis of pulmonary nodules can be roughly divided into three steps: lung area extraction, lung nodule detection, false positive reduction. This article focuses on the nodule detection and false positive reduction. For nodule detection, we design a 3D nodule detector network motivated by the work of Feature Pyramid Networks and the 1st place team of Kaggle Data Science Bowl competition. For the large number of candidates detected by the detection model, we use two 3DCNN classifiers to determine which candidates are false-positive. A 3D network performs better than a 2D network as it also captures the vertical information of nodules. Also, focal loss [2] function is used to replace the cross-entropy loss function, which has been proved that it can deal with the problem of unbalanced classification. Furthermore, we use Batch Normalization [3] and weight initialization of convolution kernel by Xavier [4] and obtained good result.

1 Introduction

Lung cell proliferation or foreign body will lead to the production of pulmonary nodules, in an increasingly deteriorating environment, more and more human lungs have produced pulmonary nodules. Nowadays, pulmonary nodules are already a very common symptom nowadays. Many young people go to the hospital to remove their lung nodules this morning. Pulmonary cancer survival rates are highly correlated with the stage of disease at the time of first diagnosis. As no obvious symptoms of early lung cancer, resulting in clinical diagnosis of lung cancer has often reached late, high cost of treatment but ineffective. Therefore, the early detection of lung cancer and early diagnosis is particularly important. At present, the application of cutting-edge technologies such as big data and artificial intelligence has become a trend in the medical field. The application of big data-driven artificial intelligence in the early diagnosis of lung cancer can not only save the lives of countless patients, but also help alleviate medical resources and patients Contradictions are also of great significance.

The LUNA16 challenge [1] also realized this problem and held a competition that focuses on a large-scale evaluation of automatic nodule detection algorithms on the LIDC/IDRI data set. The LUNA16 challenge provides a list of candidate nodules (750,000 nodules) from many nodule detectors with specific coordinates. In 'false positive reduction' challenge track, our task is to design a powerful classifier to distinguish the minor differences between true and false positives.

2 Data Preprocessing

At first, we get the lung area by using traditional methods, and then preprocessing is performed on the lung area. Use -600HU as a threshold to get a 0-1 3D map, based on the size of the 3D map block and the average distance traveled by the tile to the center, and make up for all small cave depths. The final 0-1 three-dimensional map is the lung area. As the CT picture up and down there will be some slices connected with the outside world, should be removed. The final image pixel values are clip to [-1200,600], then zoom to [0,255]. Pixels for non-lung regions are set to 170. Pretreatment can eliminate the noise, such as the bright spots of the bones, the metal lines of the CT bed. And finally, we get 128*128*128 cube.

For false positive reduction track, we use a multi-scale strategy, and prepare two sizes little cube: 36*48*48 and 20*36*36. We crop little cube from the whole lung area to feed into the two classification networks. Obtain different fields of vision size of nodules to predict the overall results.

More importantly, the training dataset has extremely high false positive to true positive ratio (735418:1557). To solve the problem of category imbalance, in addition to the focal loss function, we used oversampling to increase the number of positive samples. Specific methods are sliding window crop, flip respect to x-axis, y-axis and z-axis, rotate 90, 180, 270 degrees, multi-scale transform. Finally, we expanded the positive sample more than 300 times.

3 Nodule Detection

Compared to U-net [8], a top-down architecture with skip connections, where predictions are made on the finest level, FPN [7] has a similar structure but leverages it as a feature pyramid, with predictions made independently at all levels. FPN shows significant improvements over all existing state-of-the-art methods and can achieve higher accuracy. It suggests that it is effective to address multiscale problems using pyramid representations. We develop a 3-D FPN detector network motivated by the work of the 1st place team of Kaggle Data Science Bowl competition, the detail architecture of network is illustrated in Fig. 1

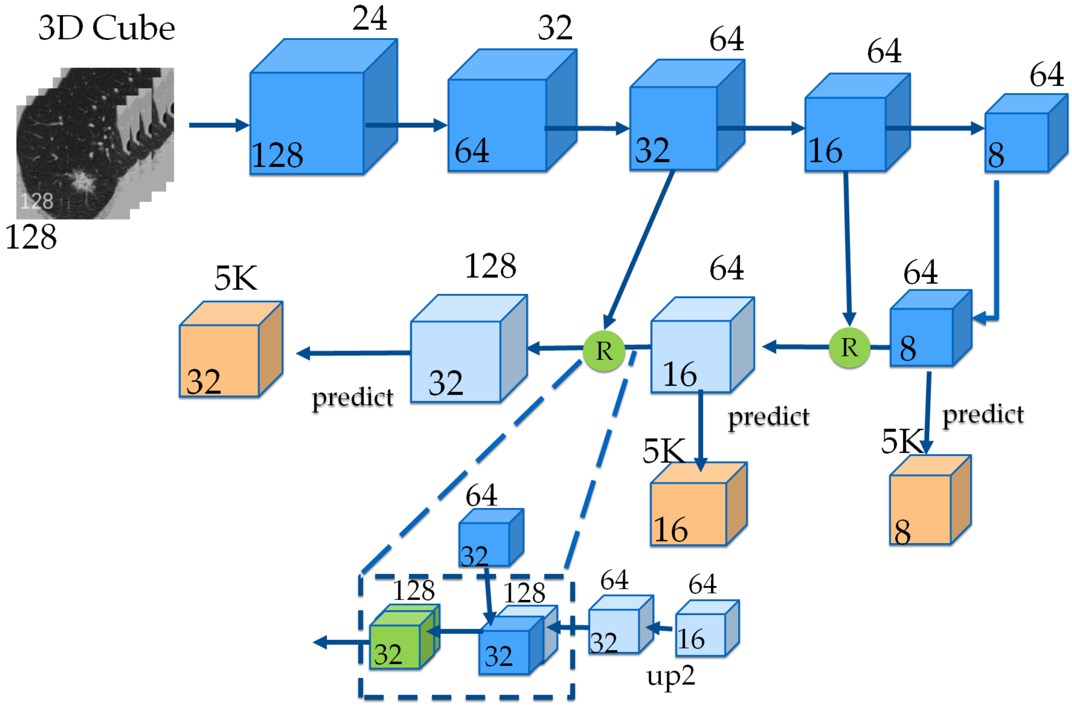

Fig.1 Structure of nodule detect Model

The input is a 128 × 128 × 128 cube cropped from the lung scans. Negative samples are much more than positive samples, to deal with the large class imbalance, we tried two ways. One is Focal loss, we use the focal loss as the loss on the output of the classification subnet. The other is adding hard negative example gradually, here is how to choose the hard negative to be included in the computation of loss. First, N negative samples are randomly selected as a candidate pool. Second, the negative samples in this pool are sorted in descending order based on their classification confidence scores, and the top n samples are selected as the hard negatives. We increase N gradually during training. By our experiment we found that the second way is more effective than focal loss, maybe for Focal loss, we need to try more parameter to get a better result.

4 False Positive Reduction

We use 3D Deep Convolution Neural Network for false positive reduction, for which captures the full range of contexts of candidates and generates more discriminative. This network contains four stages, in the first three stage, each stage contains two 3D convolutional layers, followed by Batch Normalization, Rectified Linear Unit (ReLU) activation layers, and 3D max-pooling layers. At the end of each stage, add dropout layers to avoid over-fitting. In the last stage, it contains three fully-connected layers followed by dropout layers. The detail architecture of used 3DCNN is illustrated in Fig. 1.

And another classifier is coming from 3D U-net [5,6]. At the beginning, we train 3D U-net for detecting nodules from lung, which can get good result. We found it is powerful, so we change the network to classifier, shown in Fig. 1, fine-tuning on the already trained detection model. And then the result proves it is a good idea.

Aiming at the problem of unbalanced classification, Kaiming He proposed a new loss function: focal loss. This loss function is modified based on the standard cross-entropy loss. This function can make the model more focused on hard-to-class samples during training by reducing the weight of easy-to-class samples. And focal loss has been proved effectiveness by a dense detector: RetinaNet. In this track, the problem of unbalanced categories has a great impact on the results, so we use focal loss in the two models above, the final result proves, focal loss really effective.

The final result comes from the fusion of two models.

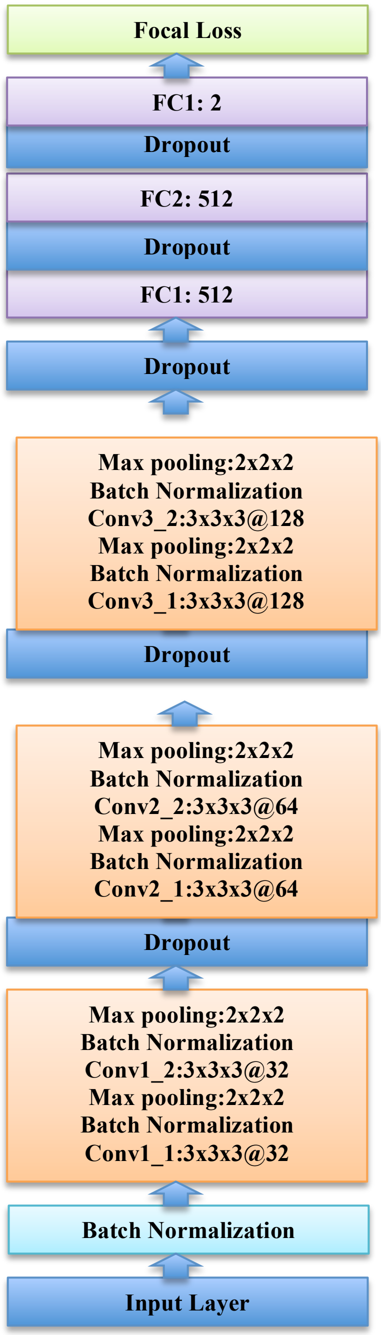

Fig.2 Structure of 3D-DCNN Model

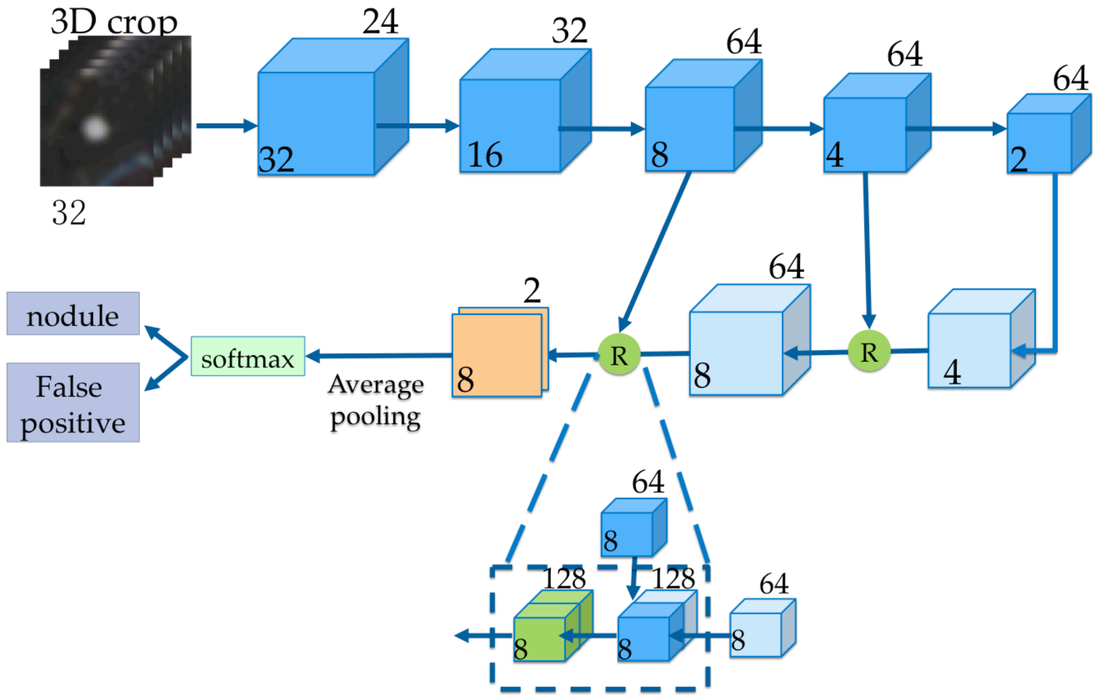

Fig.3 Structure of U-net Classifier Model

4 Experiments

In the LUNA16 challenge, performance is evaluated using the Free-Response Receiver Operating Characteristic (FROC) analysis [1]. The sensitivity is defined as the fraction of detected true positives divided by the number of nodules. In the FROC curve, sensitivity is plotted as a function of the average number of false positives per scan (FPs/scan). The average FROC-score is defined as the average of the sensitivity at seven false positive rates: 1/8, 1/4, 1/2, 1, 2, 4, and 8 FPs per scan.

For the 3D_CNN network: We did a lot of attempts on multi-scale strategies, including probabilistic weighting and finetune. Through experiments, we found that using 36*48*48 data to train the network to get the source network, and then use the 20*36*36 data on the source network to finetune, there will be better results. The method was implemented with Python based on the deep learning library of tensorflow1.0.

For the U-net classifier network: We use crop with size of 32^3. This classifier was implemented with Python based on the deep learning library of pytorch1.12.

The ReLU was used in these networks and the focal loss function was used for logistic regression to yield the prediction probabilities. The weights of the networks were initialized from Xavier and trained by minimizing the loss function with Stochastic Gradient Descent (SGD). The learning rate is step changed, (0.001,0.0001,0.00001). We use 4 GPU for training on Tesla K80, and the OS is centos 7.2.

5 References

[1] A.A.A. Setio, A. Traverso, T. Bel and et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the luna16 challenge. In *arXiv:1612.08012*, 2016.

[2] Lin T Y, Goyal P, Girshick R, et al. Focal Loss for Dense Object Detection[J]. 2017.

[3] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]//International Conference on Machine Learning. 2015: 448-456.

[4] Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks[C]//Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. 2010: 249-256.

[5] Ronneberger O. Invited Talk: U-Net Convolutional Networks for Biomedical Image Segmentation[J]. 2017, 9351:234-241.

[6] "The Winning Submission for the Kaggle Data Science Bowl 2017" by grt123 Team. 2017, 5.

[7] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie. Feature pyramid networks for object detection. In *CVPR*, 2017.

[8] O. Ronneberger, P. Fischer, and T. Brox. U-Net: Convolutional networks for biomedical image segmentation. In *MICCAI*, 2015