
3D Deep Convolution Neural Network Application in Lung Nodule Detection on CT Images

Fonova
z1953@nyu.edu

Abstract

Pulmonary cancer is the leading cause of cancer-related death worldwide, and early stage of pulmonary cancer detection using low-dose computed tomography (CT) could prevent millions of patients being killed every year. However, reading millions of those CT scans is an enormous burden for radiologists. Therefore, an immediate need is to read, detect and evaluation CT scans automatically and fast.

In the paper, we propose a new implementation method which is an 3D-UNet like network that is suitable for lung-nodule detection and three 3D Convolution Neural Networks that are suitable for false positive reduction. And we ensemble the 3 networks together to get a final probability for each candidate. A 3D network performs better than a 2D network as it also captures the vertical information of nodules. Also, leaky ReLUs are used to replace the ReLUs in all activation layers. Furthermore, we don't apply any extra fully connected layers to speed-up the train process and also obtained good result.

The team interpolated the original images and crop out a cube where centered as the given coordinations of all candidates, and apply the common data augmentation methods to balance the positive and negative samples. Finally result is evaluated by a free receive operating characteristic (FROC), and we reach the average sensitivity of 0.926 at 7 different false positive rates.

1 Introduction

Pulmonary cancer is the leading cause of cancer-related death worldwide nowadays, causing more than 1.3 million deaths annually. In order to decrease the death number, scientists and researchers have been made huge effort on it. An effective way to resolve this problem is to apply early detection for the prospective patients. Screening high risk individuals for lung cancer with low-dose CT scans is a promising method. However, after scanning those patients reading millions, we have millions of CT scans awaiting for radiologists to read. Usually, a single radiologists can only read roughly 50 patients' scans, which is very time consuming. Therefore, an immediate need is to read, detect and evaluation CT scans automatically and fast.

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 LIDC/IDRI data set is publicly available, including the annotations of nodules by four radiologists. The LUNA16 challenge provides a list of candidate nodules (750,000 nodules) from many nodule detectors with specific coordinates. Inside this list, only 1166 nodules are true positives and remaining are all false positives. Our task is to design a powerful classifier to distinguish the minor differences between true and false positives.

2 Nodule Detection

Nowadays many objection systems use networks pre-trained on ImageNet as initial weights and fine-tune them. However medical images are quite different from natural images. It is difficult to adopt a pre-trained network based on ImageNet to medical object detection task. We design a nodule detector network motivated by the work of the 1st place team of Kaggle Data Science Bowl competition. We developed a 3-D UNet-like network where takes cropped $128*128*128$ cubes as input and produces an $5*1$ vector, representing the location of x, y, z, radius and the probability of the candidate nodule. This would produces more than millions of candidate nodules, but many of them are overlapped with each other. We regard two candidate nodules that are centered within 3mm as a single nodule, and use non maximum suppression (NMS) to calculate the final probability of that nodule for second step use. Furthermore, we discard all nodules that have probability less than 0.1. The probability derived from this part is marked as p_{step1} .

The detector model is trained on two Nvidia M40, which take around 1.5 days. A 10-fold cross validation is performed on 888 samples in order to evaluate the performance of our detector. We achieve a recall of 0.991 and a FP/TP ratio of 16.

3 Data Preprocessing

A key issue is that the original CT images have different voxel length. For example, spacing on z-axis from different CT scans varies from 0.625mm to 2.5mm, same issue exists on the xy plane as well. Thus, before feeding data to our network, we interpolated the original images to same voxel spacing (x: 0.5556mm, y:0.5556mm, z:1mm). Afterwards, using the given coordinate of the center of sphere of all candidate nodules, we crop out a $40*40*24$ voxels cuboid from the interpolated images, which is centered at the given coordinates. Finally, we randomly crop out a $36*36*24$ cube from the $40*40*24$ cube. In this way, we finally get a $20mm*20mm*20mm$ cube for training afterwards.

Hounsfield value is the unit to represent the relative density of tissues and organs on CT images. In most of the lung CT images, the Hounsfield value varies from -1000 to 1000. For those values greater than 1000, we simply convert them to 1000, the maximum value we set. Same method is applied for values less than -1000.

More importantly, the training dataset has relatively high false positive to true positive ratio (roughly 16). In order to deal with the imbalance labels, we applied common data augmentations for the true nodule positions, such as rotate 90, 180, 270 degrees, shift within 2 pixels on xy plane, zoom out, zoom in, flip respect to x-axis and y-axis. Totally, we increased the true positive sample size 7 times greater.

4 False Positive Reduction

Three different models, 3D-DCNN, 3D-WRN-16-2, and cascaded 3D-DCNN models are utilized for false positive reduction. Final result ensembles the probabilities from 3 different models respectively and the detailed ideas are introduced below.

(1). The 3D-DCNN network contains three small stages. In each of the stage, there are two convolutional layers, followed by a batch normalization layer and a leaky Rectified Linear Unit (leaky-ReLU) activation layer respectively. And a max-pooling layer is connected to the second activation layer. The channels of convolutional layers for stage 1,2,3 are 32,64,128 respectively. Finally, a 2-way softmax activation layer is connected to the last max-pooling layer to classify the candidates from nodules to none-nodules. Moreover, dropout layers are added after max-pooling

layers and fully-connected layers to avoid over-fitting, dropout ratio is set to be 0.5. The detailed architecture of the proposed 3D DCNN is illustrated in 1.

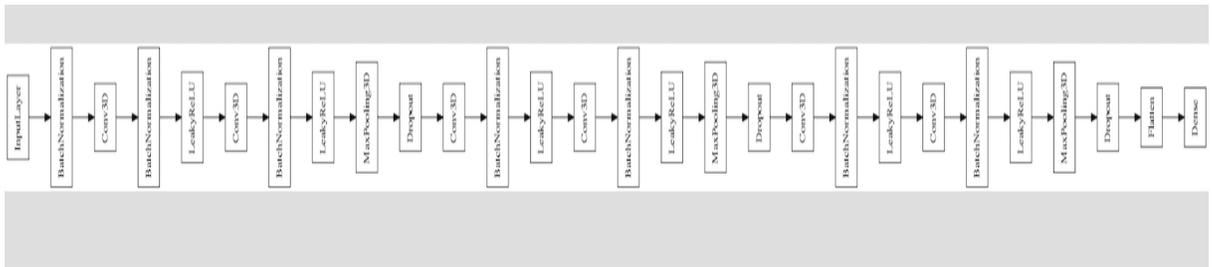


Figure 1: Structure of 3D-DCNN Model

(2). The 3D-WRN-18-2 network is the 3D version of the WRN-18-2 [6], where 18 represents the total convolutional layer numbers and 2 represents the widen factor. We recommend to use a 3D model here as it also captures the vertical information of a patient CT images. There are 3 stages in total, each stage contains 6 convolutional layers. Finally, a global max-pooling layer is applied before the final 2-way softmax activation layer. Moreover, dropout layers are added after max-pooling layers and fully-connected layers to avoid over-fitting, dropout ratio is set to be 0.1. The detailed architecture of the proposed 3D-WRN-18-2 is illustrated in 2. Or if you are suffering from reading the image details, please feel free to contact us via email above.

in (1). After each stage, the training set will be downsized to only those hard to be classified ($0.001 < \text{probability} < 0.999$). The reason we came up with this idea is that we found single 3D-DCNN model can classifier most of the nodule candidates perfectly, only very small size of the candidates are hard to be classified (namely have $0.001 < \text{prob} < 0.999$). In this case, different models at different stages are focused on different features. Generally, after each stage, only 10%-50% of the candidates are left for next stage and hence can speed up the training process as well. For those easy classified, we assign $p = 1$ for those with probability ≥ 0.999 and $p = 0$ for those with probability ≤ 0.001 .

After we get 3 probabilities for all candidates from 3 different models, we take the average of these 3 probabilities for each candidate as p_{step_2} , and the final ensemble probability is calculated as $p_{final} = 0.4 * (p_{step_1} + 0.6 * p_{step_2})$ as the result we submit. The weights 0.4, 0.6 are derived from multiple experiments and may not have any instructional meaning.

5 Experiments

In this section, we evaluate the performance of our DCNN network system on the LUNA16 Challenge. 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.

In the cross-validation process. The leaky-ReLU was used in the network and the softmax function was used for logistic regression to yield the prediction probabilities. The weights of the 3D CNNs were initialized from the normal distribution [2] and trained by minimizing the cross-entropy loss with Adam. The learning rate was initialized to 0.0001. Dropout and momentum were used during the training procedure. The method was implemented with Python based on the deep learning library of Keras which is a simplified interface of TensorFlow [3].

6 Evaluation

The evaluation of the result is based on 10-fold cross validation using the provided dataset, containing 888 patients.

7 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] K. He, X. Zhang, S. Ren, and J Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *IEEE international conference on computer vision*, page 1026-1034, 2015
- [3] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Rafal Jozefowicz, Yangqing Jia, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Mike Schuster, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent

Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[4] Tsung-Yi Lin, Piotr Dollar, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature Pyramid Networks for Object Detection. In *arXiv:1612.03144v2*, 2017

[6] Sergey Zagoruyko, and Nikos Komodakis. Wide Residual Networks, British Machine Vision Conference (BMVC), 2016. In *arXiv:1605.07146*