

Deep Convolution Neural Networks for Pulmonary Nodule Detection in CT imaging

Guohua Cheng, Weiyi Xie, Han Yang, Hongli Ji, Linyang He, Haiqi Xia, Yadong Zhou

CAD department: JianPei Technology Ltd
No.3 Street Xiasha District, Hangzhou, China, 310018

Abstract—Pulmonary Nodule detection in CT imaging is a very critical, yet very challenging medical imaging analysis task due to the wide variability of shape, textual and scale of pulmonary nodules. Deep Convolution Neural Networks are now considered the state-of-the-art in many object detection applications in computer vision. In this paper, we propose a two-stage nodule detection framework that detects nodule candidate with convolution neural networks trained separately on 2d axial slices and 3d CT volume, followed by a deep residual 3d neural network to reduce false alarms. The proposed framework shows the superior performance on LUNA16 dataset, yielding the final 0.9499 free receive operating characteristic score.

Keywords—Pulmonary Nodule Detection; Deep Convolution Neural Networks; CT Imaging;

I. INTRODUCTION

Lung cancer is the known leading cause of death to mankind. Early diagnosis of lung cancer increases the patient's chance to survive. In clinical practices, Computer Tomography (CT) is commonly-used lung cancer screening protocol that enables radiologists examining the existence of suspicious nodules. However, finding nodules in CT scans can be time consuming and error-prone due to the high variance of size, shape, texture, location of pulmonary nodules in patient studies. Many efforts have been made in developing automated systems that locate nodule in CT scans, so-called the Computer Aided Nodule Detection System [1]. The Deep Convolution Neural Networks (DCNN), with superior object detection performance in natural images, is considered the state-of-the-art in medical imaging object detection applications as well [2]. In this paper, we introduce a two-stage nodule detection framework based on DCNN, consisting of 1) to detect interest of regions with possible nodule appearances; 2) to reduce the non-nodule candidates from the first step; Since the CT imaging is three dimensional volume data and nodules are intrinsically 3d objects, DCNN filters learned directly on the overall 3d CT volume enables capturing the complete spatial context of nodules. However, due to the computational intensity and limit amount of Graphic Process Unit (GPU) memory, many researchers compromise for training object detection DCNN over 2d axial slices of CT imaging. In this research, we use both 2d and 3d DCNN filters in nodule candidate detection. Once the nodule candidate is located, we cropped multi-scale fixed size region patches from original CT volume based on the candidate center. Then feed into a 3d DCNN with dense and residual connections to eliminate further the false positive nodule candidates.

II. NODULE CANDIDATE DETECTION

A. Three-dimensional DCNN for segmentation

Inspired by the U-shaped DCNN [3], we similarly propose a 3d DCNN that exploits up-sampling to restore input resolution from the pooled feature maps for nodule object segmentation. The proposed DCNN takes image patches of $80 \times 80 \times 80$ size as input, feeding into three down-sampling blocks. Each block consists of consecutive convolution filters with size $3 \times 3 \times 3$, stride 1 and padding for reserving input resolution, followed by the $2 \times 2 \times 2$ max pooling layer to reduce spatial resolution by 2. The coarsest resolution features are convoluted with $3 \times 3 \times 3$ filters twice to learn high-level nodule semantics. Then the resulting feature maps are feed into consecutive up-sampling blocks, with each block up-sampling features by 2 to finally restore the input resolution. For finer restoration of spatial resolution, we merge the up-sampled features with down-sampled features at the same spatial resolution. Finally, a $1 \times 1 \times 1$ convolution is used to reshape the feature maps to original input shape, before the sigmoid activation is applied to produce probability segmentation maps. The structure of proposed 3d DCNN is shown in Figure 1. To reduce the training covariate drifting and the model complexity, we adopt batch normalization block wise and a dropout layer to keep 0.25 activations from the coarsest resolution. The RELU non-linearity is used for all of convolutions filters as activation function.

At training stage, each training CT image is firstly interpolated into 1.0 mm spacing for axial, coronal and sagittal directions and pixels are normalized to (0, 1) range from lung window. Furthermore, Normalized CT images are standardized to have zero mean and unit variance. Then we apply commonly-used data augmentations such as up-down and left-right flip, Gaussian blur and random shifting from the cropping center. Finally, a training step is to minimize the DICE coefficient over mini-batch of training samples using Adam optimization with initiative learning rate 10^{-4} . The choice of mini-batch size is up to GPU memory constrains.

At testing stage, since the proposed network is fully-convolutional, it can be used for any size of input data. We interpolate test CT images into 1.0 mm spacing for axial, coronal and sagittal axes, and then slide over the resulting image with the predefined window size and 25% overlapping rate. Each sliding window is further convolved with the proposed 3d DCNN straightly. This approach is very efficient

with GPU implementation and due to the memory limitation of our experimental hardware; the size of sliding window is up to 128. Finally, we simply average the probability outputs to smooth the aggregated responses over overlapping region caused by sliding window.

B. Two-dimensional volume DCNN for segmentation

We adopt the same DCNN structure as the proposed 3d detection network, with exceptions of using 2d filters and input images of size 128×128 for training. Moreover, at training and testing stage, the input axial slices are interpolated into 0.6 mm spacing for all three directions, to allow convolution filters extract finer details from local regions comparing to the proposed 3D DCNN. For data augmentation, the 2d DCNN in addition uses random cropping from the nodule center, together with common approaches used in the 3d network. We train the 2d DCNN using DICE coefficient loss and Adam optimizer as the as used in training the 3D counterpart. Since the 2d DCNN requires much less parameters to compute and less memory occupation, we apply the 2d DCNN straightly over the input image in convolution manner for detection at testing stage, instead of sliding window approach used in proposed 3D

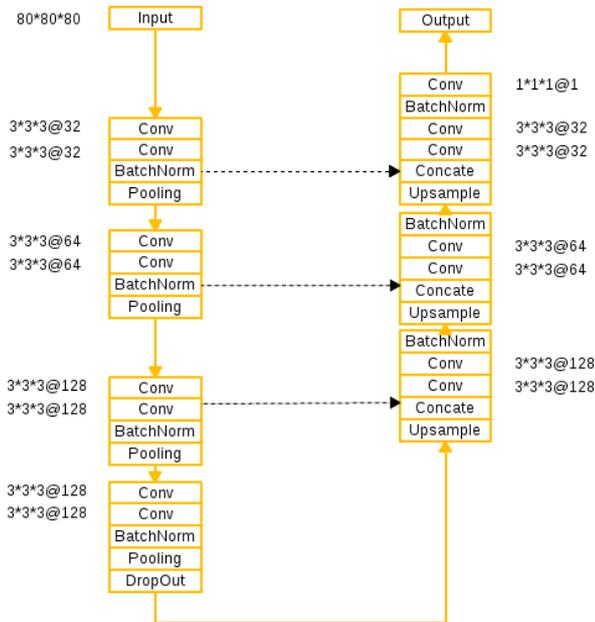


Figure 1: 3D DCNN for segmentation, Dash line denotes passing finer resolution feature maps to coarser level using concatenation

DCNN.

C. Combination Of Nodule Candidates

After thresholding of output probability maps from 2d and 3d detection DCNNs, the nodule regional masks are obtained. We hereby denote M2 for the mask generated by 2d DCNN and M3 for 3d DCNN. we merge the two masks for each image in following manner 1) for M2 nodule regions, reserve ones that do not intersect with any M3 regions, and keep nodule regions in M3 which are not overlapping with any M2 regions; 2) for overlapping regions, we use the intersection region from M3 and M2 regions.

III. FALSE POSITIVE REDUCTION

Nodule candidates from detection step may contain many false alarms, including blood vessels, thickening of lung fissures and pleura, and scars etc. The major reason is that classification of nodule and non-nodule objects from only visual clues is considered difficult even for clinical experts, especially with the objects of small size and focal opacity. The goal of false positive reduction step is to eliminate non-nodule candidates to the utmost extent. In this paper, we design a residual structure of 3d DCNN binary classifier for the false positive reduction, shown in Figure 2.

A. Training

Firstly, CT volumes are interpolated into 0.8 mm spacing for axial, coronal and sagittal directions and pixel intensity is normalized to (0, 1) from lung window. Then we crop patches from normalized CT volume using bounding boxes of nodule candidates and patches are further resized to a fixed size, $40 \times 40 \times 40$ in our experiment. The crop size is 1.5 times larger than the original bounding box size. In addition, patches are standardized to have zero mean and unit variance. Due to the imbalance of positive and negative candidates, we augment positive patches to even the size of the positive and negative set. Data augmentation approaches are 1) horizontal and vertical flips. 2) Random variations in patch intensity within (-0.15, 0.15). 3) Gaussian blur; 4) rotation at random angles within (-60, 60) degree range. 5) Random cropping, with crop size N times larger than the bounding box size, where N is uniformly distributed random variable within (1.25, 1.75). Stochastic gradient descent with the momentum of 0.9 and initiative learning rate 0.001 is used as the optimizer to minimize the cross entropy loss for training. We heavily used batch normalization in residual blocks to accelerate training. All convolution layers' activation function is RELU non-linearity.

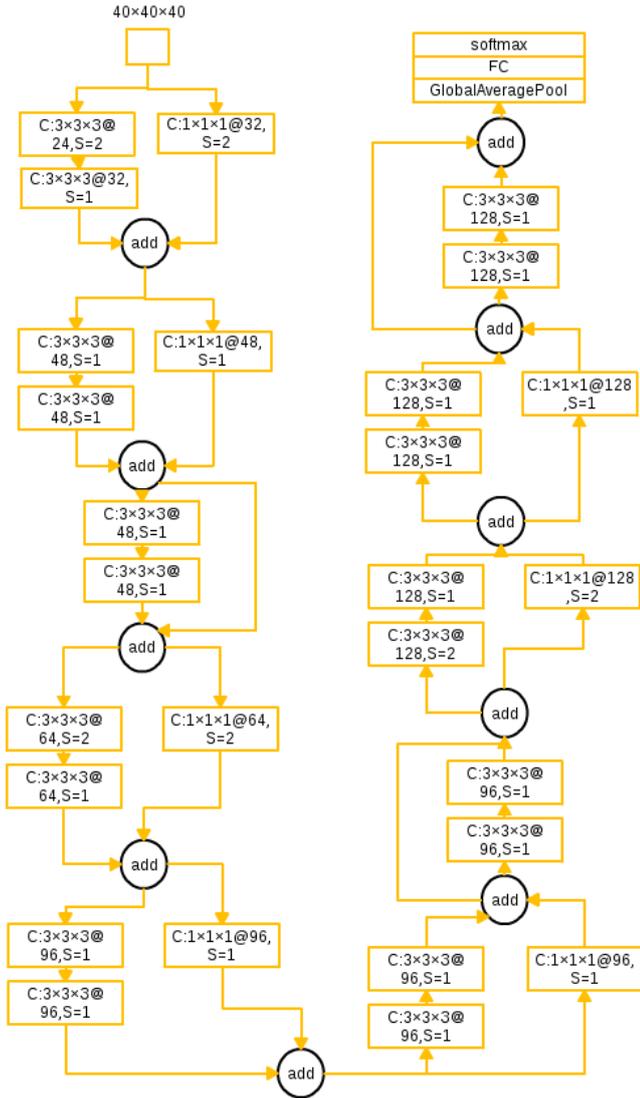


Figure 2: 3D DCNN for false positive reduction, Convolution filters are denoted with 'C' in short

B. Testing

At testing stage, CT images are preprocessed and patches are cropped the same way as we do in training stage. Also, we flip and rotate candidate patches using same approaches at training stage, and ensemble probability outputs of patches of the same candidate by averaging.

EXPERIMENT RESULTS

In this section, we evaluate the performance of the proposed nodule detection DCNN framework using Free-Response Receiver Operating Characteristic (FROC) analysis [4] on the LUNA16 Challenge. The challenge focuses on sensitivities at 0.125, 0.25, 0.5, 1, 2, 4, 8 false positives per scan of given nodule detection system. The challenge ranks submissions by evaluating average sensitivity of 7 false positive rates above. The proposed nodule detection deep convolution neural network yields 0.9499 the FROC score. The hardware settings for the experiments are GTX Titan-X \times 2 GPU, Intel E5-1650 v3 CPU and 64GB memory on Ubuntu 64bits Linux desktop. And Tensorflow is used as our deep learning framework.

REFERENCES

- [1] Tan, Maxine, et al. "A novel computer - aided lung nodule detection system for CT images." *Medical physics* 38.10 (2011): 5630-5645.
- [2] Litjens, Geert, et al. "A survey on deep learning in medical image analysis." *arXiv preprint arXiv: 1702.05747* (2017).
- [3] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, Cham, 2015.
- [4] 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.