



Automated Pulmonary Nodule Detection: High Sensitivity with Few Candidates

Bin Wang^{1,3}, Guojun Qi², Sheng Tang^{1(✉)}, Liheng Zhang², Lixi Deng^{1,3},
and Yongdong Zhang¹

¹ Key Lab of Intelligent Information Processing, Institute of Computing Technology,
Chinese Academy of Sciences, Beijing, China

ts@ict.ac.cn

² University of Central Florida, Orlando, FL, USA

³ University of the Chinese Academy of Sciences, Beijing, China

Abstract. Automated pulmonary nodule detection plays an important role in lung cancer diagnosis. In this paper, we propose a pulmonary detection framework that can achieve high sensitivity with few candidates. First, the Feature Pyramid Network (FPN), which leverages multi-level features, is applied to detect nodule candidates that cover almost all true positives. Then redundant candidates are removed by a simple but effective Conditional 3-Dimensional Non-Maximum Suppression (Conditional 3D-NMS). Moreover, a novel Attention 3D CNN (Attention 3D-CNN) which efficiently utilizes contextual information is proposed to further remove the overwhelming majority of false positives. The proposed method yields a sensitivity of 95.8% at 2 false positives per scan on the LUNG Nodule Analysis 2016 (LUNA16) dataset, which is competitive compared to the current published state-of-the-art methods.

1 Introduction

Lung cancer is one of the most lethal diseases worldwide [1]. It can be early diagnosed among high-risk individuals through screening with low-dose computed tomography (LDCT). Compared with traditional chest radiography screening, LDCT has reduced the mortality of lung cancer by 20% in seven years with early diagnosis [3]. The prevalence of CT technology has generated enormous CT data. However, it is challenging for radiologists to accurately localize every pulmonary nodule appearing on all CT slices.

Over the past two decades, researchers have developed many Computer-Aided Diagnosis (CAD) systems for automatic detection of lung nodules [2, 3]. The performances of these CAD systems have been improved significantly over previous systems. However, these techniques still have a long way to enable practical applications. These CAD systems are often designed by using hand-craft features that

This work was supported by the National Natural Science Foundation of China under Grant (61525206, 61572472, U1703261, 61571424).

are based on some low-level features such as size, shape, texture and intensity. They cannot deal with the large variations of nodules, thus may fail to distinguish nodules from those ambiguous regions. Fortunately, the development of deep learning enables us to engineer representative features to recognize various appearances of pulmonary nodules, which shows promising detection accuracy with significantly improved performances. Over the past two years, many deep learning based systems have been proposed and delivered exciting results [3–6]. However, it is still a challenging task to detect small pulmonary nodules from CT slices. Detecting small or even tiny nodules plays an important role in early diagnosis and treatment, which can effectively lower the risk of lung cancers before they develop into worse stages. This inspires us to develop a high sensitive CAD system.

Conventional pulmonary nodule detection systems usually consist of two stages: proposing nodule candidates and removing false positives. In this paper, we carefully design Convolutional Neural Network (CNN) structures to address the challenge of detecting small pulmonary nodules at each stage: First of all, in order to ensure detecting all the nodules without missing those small nodules, we design the *candidate detection network* (see Fig. 1-Stage 1) by exploring the property of Feature Pyramid Networks (FPN) [7]. This detection network could cover almost all nodules with only very few nodules missed ($16/1186 \approx 1.3\%$). Then, we propose an effective Conditional 3D-NMS to remove the redundant candidates. Moreover, we propose a novel Attention 3D-CNN model (see Fig. 1-Stage 2) allowing the model to focus on the most relevant regions for reducing false positives. We will demonstrate that the proposed pulmonary nodule detection architecture achieves as high as 95.8% sensitivities with only two false positives per scan on LUNA16 dataset, which is very competitive compared to the current state-of-the-art methods.

2 Methods

As shown in Fig. 1, the proposed framework consists of two main parts: (1) the detection of nodule candidates by using FPN and (2) false positive reduction with the Attention 3D-CNN.

2.1 High Sensitivity Candidate Detection with FPN

Candidate detection is a crucial stage for pulmonary nodule detection. The purpose of this step is to recall all possible nodules subject to a reasonable number of candidates. In principle, sliding windows of various scales can cover all possible nodules. However, it is practically impossible because of a too large number of candidates which bring great challenges for subsequent operations. On the contrary, Region Proposal Network (RPN) [8] makes a better balance between the computational cost and the amount of recalled candidates. By using the RPN, Ding *et al.* [5] showed a higher sensitivity with even fewer candidates compared with traditional CAD systems. However, small nodules are still hard to detect

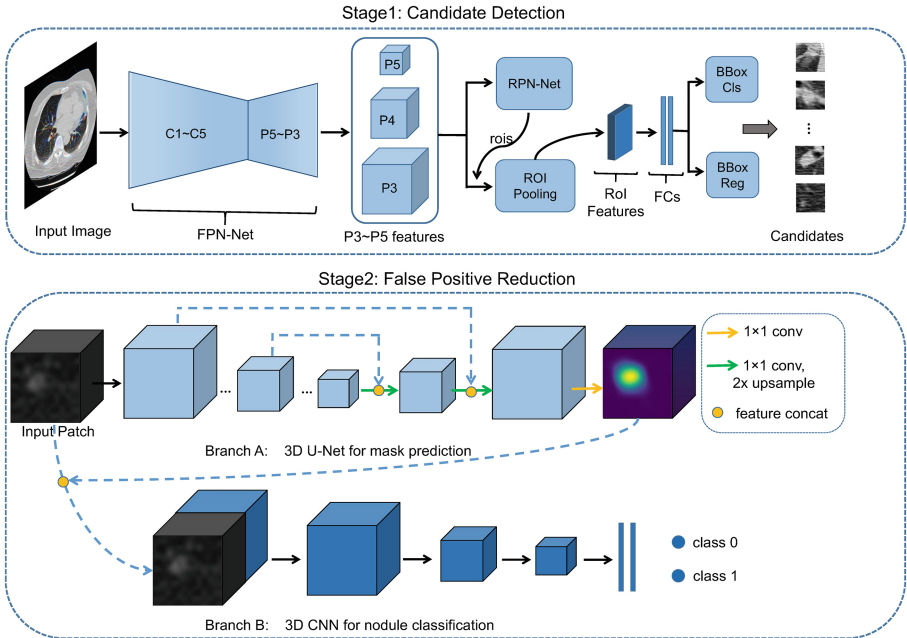


Fig. 1. The framework of proposed pulmonary nodule detection system.

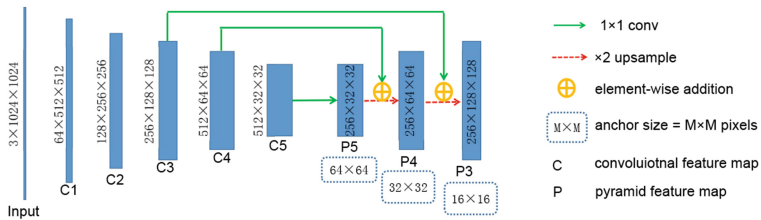


Fig. 2. Feature pyramid networks in our candidate detection architecture.

with the original RPN. To this end, we propose the FPN-based detection architecture in detail below.

Like [5], we take three consecutive layers of CT slices as input and resize them to $3 \times 1024 \times 1024$ before feeding into the network. To detect candidates, we design the FPN-based architecture as illustrated in Fig. 2. The C1~C5 layers in the proposed network correspond to the VGG16 network [9], just adding a new 2×2 pooling layer with stride 2 after Conv5.3 of the original VGG16 network. In most cases, the C5 feature map is already effective for object detection in general natural image. However, since nodules in CT are usually very small ($3 \sim 30$ mm in LUNA 2016 [3]), after passing through several pooling layers, the features of small nodules would become too weak or even disappear. We then take P3~P5 as the final features instead of C5. A 1×1 convolutional layer is attached on C5 to

produce the coarsest resolution map P5 of 256 channels. Since P5 has a lower resolution feature map, we attempt to find larger nodule proposals of size 64×64 on it. Then, we upsample the P5 by a factor of two and add it to channel-reduced C4 to obtain a middle-resolution P4 layer. In a similar way, we can obtain the P3 layer at the highest resolution. Both P3 and P4 contain higher-resolution information (which comes from C3 and C4) than P5 layer for detecting smaller nodules. Thus, 32×32 and 16×16 nodule proposals are found at P4 and P3 layers respectively. With these feature maps at different resolutions, we use RPN-Net to get nodule proposals and then classify them into nodule candidates or not.

2.2 Conditional 3D-NMS for Redundant Candidate Removal

Most pulmonary nodule candidates generated from the FPN may be repetitively detected because they often exist in multiple consecutive slices. To reduce the unnecessary computational burden, we propose a simple yet effective Conditional 3D-NMS method to remove redundant candidates. The basic idea is to divide candidates of the same CT scan into different sets of highly overlapped candidates based on their positions and radius. Then we choose the candidate with the highest mean pixel value as the final candidate of the current set. The reason we use pixel mean value as the condition is that the mean value corresponds to the Hounsfield unit (HU) value which reflects the characteristics (pulmonary nodule or lung parenchyma) of the current region. The overview of the algorithm is summarized in Algorithm 1.

Algorithm 1. Conditional 3D Non-Maximum Suppression.

Input:

- The set of candidates which belong to the same CT scan: C
 $C = \{C_1, \dots, C_N\}$; $C_i = (C_i^w, C_i^v, r_i^w, r_i^v, s_i)$, in which
 $C^w = (x^w, y^w, z^w)$ represents world coordinates,
 $C^v = (x^v, y^v, z^v)$ represents voxel coordinates,
 r and s represent the radius and the confidence of the candidate respectively.
- The image array of the CT scan: I

Output:

- The set of remaining candidates after Conditional 3D-NMS: L ;
- 1: Initialize $L = \{\}$
 Note $V(z, y, x, r) = I[z, y - \frac{r}{2} : y + \frac{r}{2}, x - \frac{r}{2} : x + \frac{r}{2}]$; which means the mean pixel value of the square area with (y, x) as center and $r/2$ as length in the z -th slice.
 - 2: **repeat**
 - 3: Initialize temporary set $T = \{\}$;
 - 4: Sort all the candidates in C by confidence s in descending order, mark the first candidate as C_h ;
 - 5: For any $C_i \in C$: if $\|C_i^w - C_h^w\|_2 \leq r_h^w$, Do $T = T \cup \{C_i^v\}$, $C = C - \{C_i\}$;
 - 6: Conditional Selection: Find the minimum z_{min} and the maximum z_{max} of coordinate z in T ,
 $z_{choose} = \arg \max_{z_i} V(z_i^v, y_h^v, x_h^v, r_h^v), i \in [z_{min}, z_{min}+1, \dots, z_{max}]$;
 - 7: $L = L \cup \{(x_h^v, y_h^v, z_{choose})\}$
 - 8: **until** $C = \{\}$
-

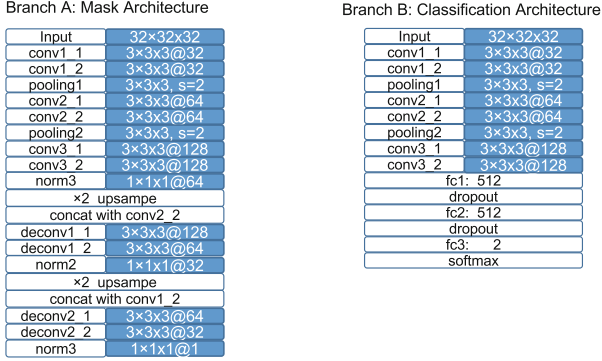


Fig. 3. The network structure of attention 3D-CNN.

2.3 Attention 3D-CNN for False Positive Reduction

The proposed candidate detection method can recall almost all of the nodules. Nevertheless, a large number of false positives exist among those candidates since it is difficult to distinguish true nodules from highly similar false positives without using three-dimensional spatial contextual information. Some researchers take advantage of 3D CNN for false positive reduction [5, 6, 10]. By contrast, to use spatial information more effectively, we propose a novel Attention 3D-CNN architecture for false positive reduction.

As shown in Fig. 1 (Stage2), the Attention 3D-CNN has two components: Branch A is the attention subnet (U-net structure [11]), which produces a 3D mask that is supposed to have a high response near the nodule. We apply the resultant 3D mask to the source patch before it is fed into the 3D CNN classification network. This allows the network to focus on the lesion area while ignoring the noisy irrelevant background. The detailed architecture of this Attention 3D-CNN network is presented in Fig. 3, where the convolutional layers are followed by batch normalization and ReLU activation.

While Branch A produces a mask of the nodule’s Gaussian distribution, its ground truth used to train the network is calculated below:

$$V = \frac{K}{\sqrt[3]{2\pi}\sigma^3} \cdot \exp\left(-\frac{(x - \bar{x})^2 + (y - \bar{y})^2 + (z - \bar{z})^2}{2\sigma^2}\right) \quad (3\sigma = 1.5r) \quad (1)$$

where $(\bar{x}, \bar{y}, \bar{z})$ represents the nodule centroid, (x, y, z) represents the voxel point of the mask in the CT scan, V is the mask value, and K is a constant term. We adopt $3\sigma = 1.5r$ instead of r , because we aim to consider context information in the model to better recognize nodules. Finally, Branch B outputs a classification probability to decide whether the current voxel is a nodule or not. We use a multi-task loss L to jointly train the network:

$$L = \lambda L_{mask} + L_{cls} \quad (2)$$

where the mask loss L_{mask} is mean squared error (MSE) between the ground truth mask and the prediction mask, the classification loss L_{cls} is focal loss [7] that is more effective than cross entropy loss to classify hard examples on an imbalanced set, and λ is a hyperparameter balancing between the two losses.

Table 1. Detection performance comparison (Sensitivity vs Candidates/scan)

System	Sensitivity	Candidates/scan
ISICAD	0.856	335.9
SubsolidCAD	0.361	290.6
LargeCAD	0.318	47.6
M5L	0.768	22.2
ETROCAD	0.929	333.0
FUSION	0.983	850.2
FRCNN	0.817	22.7
FRCNN+3DCNN [5]	0.946	15.0
FCN+OSF [6]	0.971	219.1
Cand-Det	0.987	179.6
3D-NMS	0.967	103.3
Conditional 3D-NMS	0.981	91.5
3D-CNN	0.946	8.5
Attention 3D-CNN	0.958	6.6

3 Experiments

We evaluate the proposed framework on the LUNA16 [3] dataset. It contains 888 CT scans whose pulmonary nodules have been well annotated by four experienced radiologists. The LUNA16 dataset is divided into ten subsets for ten-fold cross validation. The performances of detection algorithms are evaluated by sensitivity and average number of false positives per scan (FPs/scan). The overall score (CPM score) is defined as the average of sensitivity at seven predefined false positive rates – 1/8, 1/4, 1/2, 1, 2, 4 and 8 FPs per scan.

3.1 Implementation Details

On the candidate detection stage, we normalize the values of CT scans (Hounsfield Unit between -1000 and 400) to the range $(0, 1)$. In the stage of false positive reduction, we crop $36 \times 36 \times 36$ voxels from the detected candidates that are preprocessed by Conditional 3D-NMS. Then, data augmentation is used for training the Attention 3D-CNN network: we flip the voxels from coronal, sagittal and axial dimensions and crop $32 \times 32 \times 32$ patches as the input into the network. The constant term K in generating Gaussian masks is set to 1000 and the hyper-parameter λ is set based on the cross-validation result.

3.2 Ablation Study and Results

To check the contribution of each step in the proposed framework, we perform an ablation study. As shown in Table 1, our candidate detection network (Cand-Det) can achieve a sensitivity of 98.7% with an average of 179.6 detected candidate nodules per scan. Compared with the ‘FUSION’ result that combines five traditional CAD systems (ISICAD~ETROCAD), the proposed candidate detection method can achieve higher sensitivity with fewer candidates.

After candidate detection, the Conditional 3D-NMS (Cand-Det-CNMS) method is adopted to remove redundant candidates. From the table, we can see that, compared with the normal 3D-NMS, the Conditional 3D-NMS is more effective for removing redundancy. Finally, we use the Attention 3D-CNN to remove massive false positives and reach a sensitivity of 95.8% with only 6.6 detected nodules per scan. By comparison, if we remove the branch A from the Attention 3D-CNN network, it becomes a typical 3D CNN, and the result would reduce to 94.6% sensitivity with 8.5 detections per scan. This shows that adding attention branch can help improve the sensitivity as well as reduce false detections. The FROC curve for each step is plotted in Fig. 4. We can see that the performance of the proposed framework is improved by combining these steps.

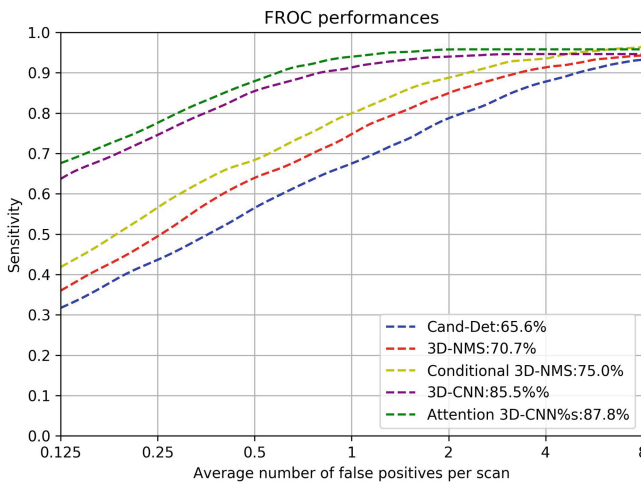


Fig. 4. FROC performances in each stage of the proposed method

To further analyze the performance of the proposed framework, we compared with state-of-the-art methods [5, 6, 10] by using CPM score. In Table 2, we can see that the CPM score of the proposed system is 87.8%, which is better than the methods proposed in [6, 10], and only a little lower than the FRCN+3DCNN. However, in clinical practice, radiologists are more concerned with the sensitivities when the FPs per scan rates vary from 1 to 4 [5, 6], where the proposed architecture achieves the best performance.

Table 2. Comparison with state-of-art methods on LUNA16 dataset [3] (CPM score). We achieve best results when FPs per scan vary from 1 to 4 which is more worth noting in clinical practice.

Sensitivity \ FPs/scan	0.125	0.25	0.5	1	2	4	8	CPM score
MultiLevel[10]	0.677	0.737	0.815	0.848	0.879	0.907	0.922	0.827
FCN+OSF [6]	0.659	0.745	0.819	0.865	0.906	0.933	0.946	0.839
FRCN+3DCNN[5]	0.748	0.853	0.887	0.922	0.938	0.944	0.946	0.891
Ours	0.676	0.776	0.879	0.949	0.958	0.958	0.958	0.878

4 Conclusion

In this work, we propose an architecture for the detection of pulmonary nodules. The architecture can first produce nodule candidates with high sensitivity using a FPN-based detection network. Then, a simple and effective Conditional 3D-NMS method removes the redundant candidates. Finally, a novel Attention 3D-CNN network is proposed to reduce the abundance of false positives. Experiments show that our architecture can achieve a high sensitive result with few candidates. The architecture can also be extended to other similar object detection tasks in CT scans.

References

1. Siegel, R.L., Miller, K.D., Jemal, A.: Cancer statistics, 2018. *CA Cancer J. Clin.* **68**(1), 7–30 (2018)
2. Murphy, K., van Ginneken, B., Schilham, A.M.R., et al.: A large-scale evaluation of automatic pulmonary nodule detection in chest CT using local image features and k-nearest-neighbour classification. *Med. Image Anal.* **13**(5), 757–770 (2009)
3. Setio, A.A.A., Traverso, A., De Bel, T., et al.: Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge. *Med. Image Anal.* **42**, 1–13 (2017)
4. Setio, A.A.A., Ciompi, F., Litjens, G., et al.: Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks. *IEEE Trans. Med. Imaging* **35**(5), 1160–1169 (2016)
5. Ding, J., Li, A., Hu, Z., Wang, L.: Accurate pulmonary nodule detection in computed tomography images using deep convolutional neural networks. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) *MICCAI 2017*. LNCS, vol. 10435, pp. 559–567. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66179-7_64
6. Dou, Q., Chen, H., Jin, Y., Lin, H., Qin, J., Heng, P.-A.: Automated pulmonary nodule detection via 3d convnets with online sample filtering and hybrid-loss residual learning. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) *MICCAI 2017*. LNCS, vol. 10435, pp. 630–638. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66179-7_72

7. Lin, T.Y., Dollár, P., Girshick, R., et al.: Feature pyramid networks for object detection. In: CVPR, vol. 1, no. 2, p. 4 (2017)
8. Ren, S., He, K., Girshick, R., et al.: Faster R-CNN: towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems, pp. 91–99 (2015)
9. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556) (2014)
10. Dou, Q., Chen, H., Yu, L., et al.: Multilevel contextual 3-D CNNs for false positive reduction in pulmonary nodule detection. *IEEE Trans. Biomed. Eng.* **64**(7), 1558–1567 (2017)
11. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28