Accurate Pulmonary Nodule Detection in Computed Tomography Images Using Deep Convolutional Neural Networks

Jia Ding, Aoxue Li, Zhiqiang Hu, and Liwei Wang[✉]

The Key Laboratory of Machine Perception (MOE), School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China
{dingjia,lax,huzq}@pku.edu.cn, wanglw@cis.pku.edu.cn

Abstract. Early detection of pulmonary cancer is the most promising way to enhance a patient's chance for survival. Accurate pulmonary nodule detection in computed tomography (CT) images is a crucial step in diagnosing pulmonary cancer. In this paper, inspired by the successful use of deep convolutional neural networks (DCNNs) in natural image recognition, we propose a novel pulmonary nodule detection approach based on DCNNs. We first introduce a deconvolutional structure to Faster Region-based Convolutional Neural Network (Faster R-CNN) for candidate detection on axial slices. Then, a three-dimensional DCNN is presented for the subsequent false positive reduction. Experimental results of the LUng Nodule Analysis 2016 (LUNA16) Challenge demonstrate the superior detection performance of the proposed approach on nodule detection (average FROC-score of 0.893, ranking the **1st place** over all submitted results), which outperforms the best result on the leaderboard of the LUNA16 Challenge (average FROC-score of 0.864).

1 Introduction

Pulmonary cancer, causing 1.3 million deaths annually, is a leading cause of cancer death worldwide [8]. Detection and treatment at an early stage are required to effectively overcome this burden. Computed tomography (CT) was recently adopted as a mass-screening tool for pulmonary cancer diagnosis, enabling rapid improvement in the ability to detect tumors early. Due to the development of CT scanning technologies and rapidly increasing demand, radiologists are overwhelmed with the amount of data they are required to analyze.

Computer-Aided Detection (CAD) systems have been developed to assist radiologists in the reading process and thereby potentially making pulmonary cancer screening more effective. The architecture of a CAD system for pulmonary nodule detection typically consists of two stages: nodule candidate detection and false positive reduction. Many CAD systems have been proposed for nodule detection [7,10]. Torres et al. detect candidates with a dedicated dot-enhancement filter and then a feed-forward neural network based on a small set of hand-craft features is used to reduce false positives [10]. Although conventional CAD systems have yielded promising results, they still have two distinct drawbacks as follows.
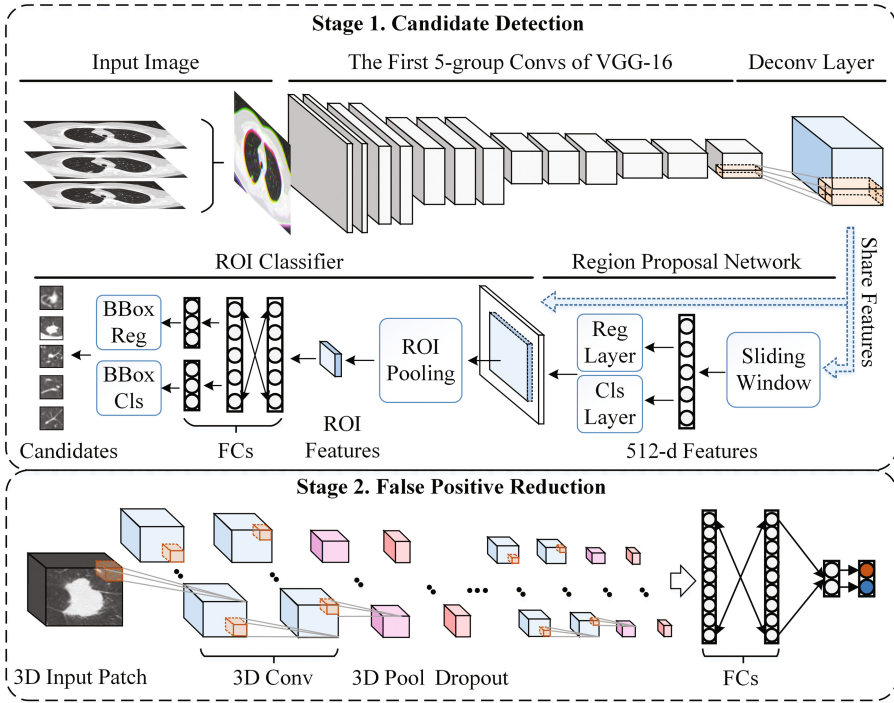
Fig. 1. The framework of the proposed CAD system.

- Traditional CAD systems detect candidates based on some simple assumptions (eg. nodules look like a sphere) and propose some low-level descriptors [10]. Due to the high variability of nodule shape, size, and texture, low-level descriptors fail to capture discriminative features, resulting in inferior detection results.
- Since CT images are 3D inherently, 3D contexts play an important role in recognizing nodules. However, several 2D/2.5D deep neural networks have achieved promising performance in false positive reduction [6,11] while rare works focus on introducing 3D contexts for nodule detection directly.

In this paper, to address the aforementioned two issues, we propose a novel CAD system based on DCNNs for accurate pulmonary nodule detection. In the proposed CAD system, we first introduce a deconvolutional structure to Faster Region-based Convolutional Neural Network (Faster R-CNN), the state-of-the-art general object detection model, for candidate detection on axial slices. Then, a three-dimensional DCNN (3D DCNN) is presented for false positive reduction. The framework of our CAD system is illustrated in Fig. 1.

To evaluate the effectiveness of our CAD system, we test it on the LUng Nodule Analysis 2016 (LUNA16) challenge [7], and yield the 1st place of Nodule Detection Track (NDET) with an average FROC-score of 0.893, which outperforms the best

result on the leaderboard (average FROC-score of 0.864). Our system achieves high detection sensitivities of 92.2% and 94.4% at 1 and 4 false positives per scan, respectively.

2   The Proposed CAD System

In this section, we propose a CAD system based on DCNNs for accurate pulmonary nodule detection, where two main stages are incorporated: (1) candidate detection by introducing the deconvolutional structure into Faster R-CNN and (2) false positive reduction by using a three-dimensional DCNN.

2.1   Candidate Detection Using Improved Faster R-CNN

Candidate detection, as a crucial step in the CAD systems, aims to restrict the total number of nodule candidates while remaining high sensitivity. Inspired by the successful use of DCNNs in object recognition [5], we propose a DCNN model for detecting nodule candidates from CT images, where the deconvolutional structure is introduced into the state-of-the-art general object detection model, Faster R-CNN, for fitting the size of nodules. The details of our candidate detection model are given as follows.

We first describe the details of generating inputs for our candidate detection network. Since using 3D volume of original CT scan as the DCNN input gives rise to high computation cost, we use axial slices as inputs instead. For each axial slice in CT images, we concatenate its two neighboring slices in axial direction, and then rescale it into $600 \times 600 \times 3$ pixels (as shown in Fig. 1).

In the following, we describe the details of the architecture of the proposed candidate detection network. The network is composed of two modules: a region proposal network (RPN) aims to propose potential regions of nodules (also called Region-of-Interest (ROI)); a ROI classifier then recognizes whether ROIs are nodules or not. In order to save computation cost of training DCNNs, these two DCNNs share the same feature extraction layers.

Region Proposal Network. The region proposal network takes an image as input and outputs a set of rectangular object proposals (i.e. ROIs), each with an objectness score [5]. The structure of the network is given as follows.

Owing to the much smaller size of pulmonary nodules compared with common objects in natural images, original Faster R-CNN, which utilizes five-group convolutional layers of VGG-16Net [9] for feature extraction, cannot explicitly depict the features of nodules and results in a limited performance in detecting ROIs of nodules. To address this problem, we add a deconvolutional layer, whose kernel size, stride size, padding size and kernel number are 4, 4, 2 and 512 respectively, after the last layer of the original feature extractor. Note that, the added deconvolutional layer recovers more fine-grained features compared with original feature maps, our model thus yields much better detection results than the original Faster R-CNN. To generate ROIs, we slide a small network over the feature map of the deconvolutional layer. This small network takes a 3×3

spatial window of deconvolutional feature map as input and map each sliding window to a 512-dimensional feature. The feature is finally fed into two sibling fully-connected layers for regressing the boundingbox of regions (i.e. Reg Layer in Fig. 1) and predicting objectness score (i.e. Cls Layer in Fig. 1), respectively.

At each sliding-window location, we simultaneously predict multiple ROIs. The multiple ROIs are parameterized relative to the corresponding reference boxes, which we call anchors. To fit the size of nodules, we design six anchors with different size for each sliding window: 4×4, 6×6, 10×10, 16×16, 22×22, and 32×32 (See Fig. 2). The detailed description of RPN is given in [5].
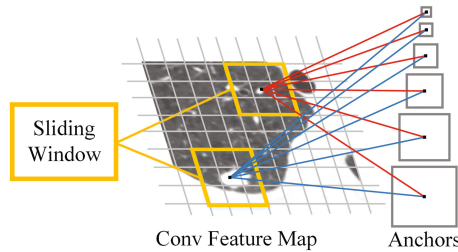

Fig. 2. Illustration of anchors in the improved Faster R-CNN

ROI Classification Using Deep Convolutional Neural Network. With the ROIs extracted by RPN, a DCNN is developed to decide whether each ROI is nodule or not. A ROI Pooling layer is firstly exploited to map each ROI to a small feature map with a fixed spatial extent $W \times H$ (7×7 in this paper). The ROI pooling works by dividing the ROI into an $W \times H$ grid of sub-windows and then max-pooling the values in each sub-window into the corresponding output grid cell. Pooling is applied independently to each feature map channel as in standard max pooling. After ROI pooling layer, a fully-connected network, which is composed of two 4096-way fully-connected layers, then map the fixed-size feature map into a feature vector. A regressor and a classifier based on the feature vector (i.e. BBox Reg and BBox Cls in Fig. 1) then respectively regress boundingboxes of candidates and predict candidate confidence scores.

In the training process, by merging the RPN and ROI classifier into one network, we define the loss function for an image as follows.

$$\mathcal{L}_t = \frac{1}{N_c} \sum_i \mathcal{L}_c(\hat{p}_i, p_i^*) + \frac{1}{N_r} \sum_i \mathcal{L}_r(\hat{t}_i, t_i^*) + \frac{1}{N_{c'}} \sum_j \mathcal{L}_c(\tilde{p}_j, p_j^*) + \frac{1}{N_{r'}} \sum_j \mathcal{L}_r(\tilde{t}_j, t_j^*)$$

(1)

where N_c, N_r, $N_{c'}$ and $N_{r'}$ denote the total number of inputs in Cls Layer, Reg Layer, BBox Cls and BBox Reg, respectively. The $\hat{p}_i$ and p_i^* respectively denote the predicted and true probability of anchor i being a nodule. $\hat{t}_i$ is a vector representing the 4 parameterized coordinates of the predicted bounding box of

RPN, and t_i^* is that of the ground-truth box associated with a positive anchor. In the same way, $\tilde{p}_j$, p_j^*, $\hat{t}_j$ and t_j^* denote the corresponding concepts in the ROI classifier. The detailed definitions of classification loss $\mathcal{L}_c$ and regression loss $\mathcal{L}_r$ are the same as the corresponding definitions in the literature [5].

2.2 False Positive Reduction Using 3D DCNN

In the consideration of time and space cost, we propose a two-dimensional (2D) DCNN (i.e. Improved Faster R-CNN) to detect nodule candidates (See Sect. 2.1). With the extracted nodule candidates, a 3D DCNN, which captures the full range of contexts of candidates and generates more discriminative features compared with 2D DCNNs, is utilized for false positive reduction. This network contains six 3D convolutional layers which are followed by Rectified Linear Unit (ReLU) activation layers, three 3D max-pooling layers, three fully connected layers, and a final 2-way softmax activation 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 overfitting. We initialize the parameters of the proposed 3D DCNN by the same strategy using in the literature [3]. The detailed architecture of the proposed 3D DCNN is illustrated in Fig. 3.
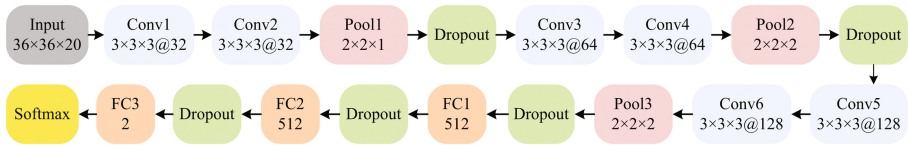

Fig. 3. The architecture of the proposed three-dimensional deep convolutional neural network. In this figure, 'Conv', 'Pool', and 'FC' denote the convolutional layer, pooling layer and fully-connected layer, respectively.

As for inputs of the proposed 3D DCNN, we firstly normalize each CT scan with a mean of -600 and a standard deviation of -300. After that, for each candidate, we use the center of candidate as the centroid and then crop a $40 \times 40 \times 24$ patch. The strategy for data augmentation is given as follows.

- **Crop.** For each $40 \times 40 \times 24$ patch, we crop smaller patches in the size of $36 \times 36 \times 20$ from it, thus augmenting 125 times for each candidate.
- **Flip.** For each cropped $36 \times 36 \times 20$ patch, we flip it from three orthogonal dimensions (coronal, sagittal and axial position), thus finally augmenting $8 \times 125 = 1000$ times for each candidate.
- **Duplicate.** In training process, whether a candidate is positive or negative is decided by whether the geometric center of the candidate locates in a nodule or not. To balance the number of positive and negative patches in the training set, we duplicate positive patches by 8 times.

Note that 3D context of candidates plays an important role in recognizing nodules due to the inherently 3D structure of CT images. Our 3D convolutional filters, which integrate 3D local units of previous feature maps, can 'see' 3D context of candidates, whereas traditional 2D convolutional filters only capture 2D local features of candidates. Hence, the proposed 3D DCNN outperforms traditional 2D DCNNs.

3   Experimental Results and Discussions

In this section, we evaluate the performance of our CAD system on the LUNA16 Challenge [7]. Its dataset was collected from the largest publicly available reference database for pulmonary nodules: the LIDC-IDRI [2], which contains a total of 1018 CT scans. For the sake of pulmonary nodules detection, CT scans with slice thickness greater than 3 mm, inconsistent slice spacing or missing slices were excluded, leading to the final list of 888 scans. The goal of this challenge is to automatically detect nodules in these volumetric CT images.

In the LUNA16 challenge, performance of CAD systems are evaluated using the Free-Response Receiver Operating Characteristic (FROC) analysis [7]. 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.

3.1   Candidate Detection Results

The candidate detection results of our CAD system, together with other candidate detection methods submitted to LUNA16 [7], are shown in Table 1. From this table, we can observe that our CAD system has achieved the highest sensitivity (94.6%) with the fewest candidates per scan (15.0) among these CAD systems, which verifies the superiority of the improved Faster R-CNN in the

Table 1. The comparison of CAD systems in the task of candidate detection.

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
Baseline (w/o deconv)	0.817	22.7
Baseline (4 anchors)	0.895	25.8
Ours	**0.946**	**15.0**

task of candidate detection. We also make comparison to two baseline methods of the improved Faster R-CNN (See Table 1). 'Baseline (w/o deconv)' is a baseline method where the deconvolutional layer is omitted in feature extraction and Baseline(4 anchors)' is a baseline method where only four anchors (i.e. 4×4, 10×10, 16×16, and 32×32) are used in improved Faster R-CNN. According to Table 1, the comparison between 'Ours' vs. 'Baseline (w/o deconv)' verifies the effectiveness of the deconvolutional layer in the improved Faster R-CNN, while the comparison between 'Ours' vs. 'Baseline (4 anchors)' indicates that the proposed 6 anchors are more suitable for candidate detection.

3.2   False Positive Reduction Results

To evaluate the performance of our 3D DCNN in the task of false positive reduction, we conduct a baseline method using the NIN [4], a state-of-the-art 2D DCNN model for general image recognition. To fit the NIN into the task of false positive reduction, we modify its input size from $32 \times 32 \times 3$ into $36 \times 36 \times 7$ and the number of final softmax outputs is changed into 2. For fair comparison, we use the same candidates and data augmentation strategy with the proposed 3D DCNN. The comparison of the two DCNNs in the task of false positive reduction are provided in Fig. 4. Experimental results demonstrate that our 3D DCNN significantly outperforms 2D NIN, which verifies the superiority of the proposed 3D DCNN to 2D DCNN in false positive reduction. We further present the comparison among top results on the leaderboard of LUNA16 Challenge[1], which is shown in Fig. 4. From this figure, we can observe that our model has attained the best performance among the CAD systems submitted in the task of nodule detection. Although Ethan20161221 and resnet yield comparable performance when the number of FPs/scan is more than 2, however, they perform a significant drop with less than 2 FPs/scan, which limits their practicability in
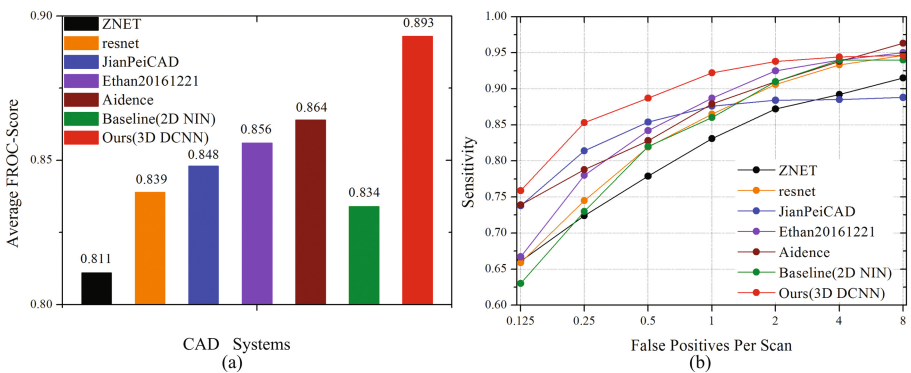

Fig. 4. Comparison of performance among our CAD system and other submitted approaches on the LUNA16 Challenge. (a) Average FROC-scores (b) FROC curves

---

[1] Until the submission of this paper, https://luna16.grand-challenge.org/results/.

nodule detection. Moreover, Aidence trained its model using the labeled data on the NLST dataset [1], therefore, its result is actually incomparable to ours. Noted that, since most CAD systems used in clinical diagnosis have their internal threshold set to operate somewhere between 1 to 4 false positives per scan on average, our system satisfies clinical usage perfectly.

4   Conclusion

In this study, we propose a novel pulmonary nodule detection CAD system based on deep convolution networks, where a deconvolutional improved Faster R-CNN is developed to detect nodule candidates from axial slices and a 3D DCNN is then exploited for false positive reduction. Experimental results on the LUNA16 Nodule Detection Challenge demonstrate that the proposed CAD system ranks the 1st place of Nodule Detection Track (NDET) with an average FROC-score of 0.893. We believe that our CAD system would be a very powerful tool for clinical diagnosis of pulmonary cancer.

References

1. Alberts, D.S.: The national lung screening trial research team reduced lung-cancer mortality with low-dose computed tomographic screening. New Engl. J. Med. **365**(5), 395–409 (2011)
2. Armato, S., McLennan, G., Bidaut, L., et al.: The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans. Med. Phys. **38**(2), 915–931 (2011)
3. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: surpassing human-level performance on imagenet classification. In: IEEE International Conference on Computer Vision, pp. 1026–1034 (2015)
4. Lin, M., Chen, Q., Yan, S.: Network in network (2013). arXiv:1312.4400
5. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems, pp. 91–99 (2015)
6. 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)
7. Setio, A.A.A., Traverso, A., Bel, T., et al.: Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the luna16 challenge (2016). arXiv:1612.08012
8. Siegel, R.L., Miller, K.D., Jemal, A.: Cancer statistics, 2015. CA Cancer J. Clin. **65**(1), 5–29 (2015)

9. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: International Conference on Learning Representations, pp. 1–9 (2015)
10. Torres, E.L., Fiorina, E., Pennazio, F., et al.: Large scale validation of the M5L lung CAD on heterogeneous CT datasets. Med. Phys. **42**(4), 1477–1489 (2015)
11. Zagoruyko, S., Komodakis, N.: Wide residual networks (2016). arXiv:1605.07146