IEEE Access
Multidisciplinary ⁞ Rapid Review ⁞ Open Access Journal

Received March 8, 2019, accepted March 25, 2019, date of publication March 29, 2019, date of current version April 17, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2908195

Pulmonary Nodule Detection in Volumetric Chest CT Scans Using CNNs-Based Nodule-Size-Adaptive Detection and Classification

JUN WANG[1], JIAWEI WANG[2], YAOFENG WEN[3], HONGBING LU[4], TIANYE NIU[1], JIANGFENG PAN[5], AND DAHONG QIAN[6], (Senior Member, IEEE)

[1]Institute of Translational Medicine, Zhejiang University, Hangzhou 310058, China
[2]Department of Radiology, The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou 310009, China
[3]Shanghai Industrial Technology Institute, Shanghai 201203, China
[4]College of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China
[5]Medical Imaging Department, Jinhua Municipal Central Hospital, Jinhua 321000, China
[6]Deepwise Healthcare Joint Research Laboratory, School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

Corresponding authors: Jiangfeng Pan (panjiangfeng967@163.com) and Dahong Qian (dahong.qian@sjtu.edu.cn)

This work was supported in part by the talent project of Shanghai Industrial Technology Institute: Application of artificial intelligence and molecular detection technology on the construction of models for tumor assistant diagnosis under Grant CYY2017002, in part by the National Key Research and Development Program of China under Grant 2017YFC0113400, and in part by the Major Special Projects Of Jinhua Science and Technology Bureau: Action Plan for Lung Nodule Screening under Grant 2017-3-004.

ABSTRACT In computed tomography, automated detection of pulmonary nodules with a broad spectrum of appearance is still a challenge, especially, in the detection of small nodules. An automated detection system usually contains two major steps: candidate detection and false positive (FP) reduction. We propose a novel strategy for fast candidate detection from volumetric chest CT scans, which can minimize false negatives (FNs) and false positives (FPs). The core of the strategy is a nodule-size-adaptive deep model that can detect nodules of various types, locations, and sizes from 3D images. After candidate detection, each result is located with a bounding cube, which can provide rough size information of the detected objects. Furthermore, we propose a simple yet effective CNNs-based classifier for FP reduction, which benefits from the candidate detection. The performance of the proposed nodule detection was evaluated on both independent and publicly available datasets. Our detection could reach high sensitivity with few FPs and it was comparable with the state-of-the-art systems and manual screenings. The study demonstrated that excellent candidate detection plays an important role in the nodule detection and can simplify the design of the FP reduction. The proposed candidate detection is an independent module, so it can be incorporated with any other FP reduction methods. Besides, it can be used as a potential solution for other similar clinical applications.

INDEX TERMS Computed tomography, pulmonary nodule, object detection, deep-learning, convolutional neural networks.

I. INTRODUCTION

Spiral computed tomography (CT) is one of the most widely used diagnostic tools for detecting pulmonary lesions [1]. With respect to the lesions, nodule screening is an important task in hospitals, because the nodule may indicate lung cancer, which is a cause of high mortality in human beings [2]. Nodule screenings entail a high workload for radiologists: they must find nodules of various sizes, shapes and locations across a large number of CT images generated from thin-sliced reconstructions. It is relatively easy for radiologists to locate big nodules, but some small ones, especially the ground-glasses and the solids surrounded by other tissues, are difficult to find even by experienced radiologists. The heavy

The associate editor coordinating the review of this manuscript and approving it for publication was Yudong Zhang.

workload and human subjectivity can lead to omission of nodules [3].

Reliable computer-aided detection (CAD) systems present a solution to alleviate the workload of radiologists and in helping them reduce the number of omitted nodules. However, it is an enormous challenge to design clinically applicable systems, because nodules have various features and different position distributions in the lung regions as illustrated in figure 1. In the past decades, many systems have been developed, which can be roughly classified into two categories: traditional systems [4]–[9] and deep-learning based systems [10]–[20]. Traditional systems were usually designed for specific scenarios based on handcrafted descriptors and image processing techniques. They are unable to cope with a broad spectrum of cases and thus their clinical applications are limited.

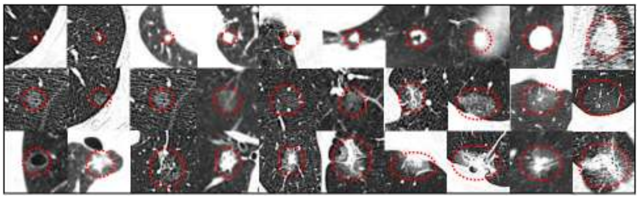


FIGURE 1. Examples of pulmonary nodules (see the red dashed ellipses) with various sizes, locations and types.

Deep-learning based systems benefit from the development of deep learning techniques for computer vision and pattern recognition [21]–[32]. A basic technique behind these applications is the so-called convolutional neural network (CNN), which can automatically learn high-level features from many training images. A nodule detection system usually contains two major stages: candidate detection and false positive (FP) reduction. Some existing studies focused on FP reduction using 2D CNNs [13] or 3D CNNs [12]. There is no doubt that candidate detection also plays an important role in a complete system, because it determines the maximum sensitivity of the nodule detection. Predecessors in this field have developed 3D semantic segmentation [15], [18] and 3D objection detection [16]–[17] [20] for the candidate detection and achieved promising results. However, it is still a challenge to detect small nodules, *e.g.*, nodules of size smaller than 5 mm. A potential solution for small nodule detection is voxel-wise detection which has been used for detection of cerebral microbleed (CMB) in brain MRI images [46] and segmentation of organs-at-risks in head and neck CT images [47]. Unfortunately, voxel-by-voxel detection may demand more computing power, which may not be suitable for our target clinical applications where processing speed is also important.

In this paper, we propose a new CAD system for nodule detection in chest CT images. In our system, we designed a novel CNN-based nodule-size-adaptive model for fast and accurate candidate detection. The proposed model is constructed from three unified neural networks which are sharing the convolutional features to reduce the detection time, namely, a feature extraction network (FEN), a region proposal network (RPN) and a region classification network (RCN). We use FEN to extract a convolutional feature pyramid from a 3D image that is fed into the model. In the RPN, we assign a series of boxes on the 3D image based on the shared feature pyramid and we predict the suspicious lesions (ROIs) from the boxes. To deal with nodules of different sizes and shapes, we set boxes with different scales and aspect ratios. By applying the third network RCN, we reduce most FPs from the ROIs through a binary classifier which is also based on the shared feature pyramid and we keep the remaining ROIs as candidates. Finally, we apply the well-known inception-v4 neural network [31] to the candidates for further suppression of FPs.

Our major contributions can be summarized as: (1) we designed and trained a deep CNN-based model which can be used for fast detection of pulmonary nodules. The detection can cover nodules of various appearances, locations and sizes ranging from 3 mm to 70 mm (in clinical practice, nodule-like lesions of size larger than 30 mm are usually called masses). Evaluations demonstrated that our detection can achieve high sensitivity with a low FP ratio. (2) We proposed a simple yet effective strategy to minimize the number of false positives (FPs) and false negatives (FNs) in the candidate detection. The detection was performed slice by slice, from the first slice to the last in a scan and for the slice that was being detected, more contextual information was incorporated through the concatenation of its adjacent slices. (3) The effectiveness of our candidate detection greatly reduces the design complexity of FP reduction. From the candidates, we trained two inception-v4 models of different receptive fields for FP reduction.

II. MATERIALS

A. TIANCHI AI DATASET

Pulmonary nodule detection is one of TIANCHI AI challenges. In the challenge, a total of 1000 scans [33] were provided. The maximum slice thickness of all scans was limited to 2 mm. The nodule size distribution was as follows: 5–10 mm nodules comprised 50% and 10–30 mm nodules comprised the other 50%. More details of the dataset can be found at the website of the challenge [33].

In our study, all 1000 scans were used as the training datasets. To create the database for training the candidate detection model, we extracted 3D images from the annotations. Each image had three channels which were built by concatenating three adjacent axial slices. Compared to a single slice, three slices have richer features for distinguishing between nodules and other tissues. Figure 2 shows a special and representative case. In the single slice, it is difficult to discriminate the small nodule (see the red rectangle) from the vessel (see the yellow rectangle). But in the 3D image, the differences become more obvious, *e.g.*, the nodule has less change in shape than the vessel. To focus the detection on the lung regions, the pixel values of each image were converted

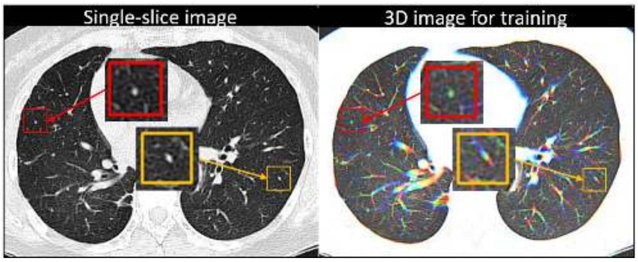


FIGURE 2. An example of 3D image for training. In the single-slice image, it is difficult to distinguish between small nodules (the red rectangle encloses a small nodule) and vessels (the yellow rectangle shows an example of a vessel). In the 3D image, the nodule and the vessel become more distinguishable.

from the CT numbers (Hounsfield Units, HU) using the L/W setting −700 HU/1000 HU. This L/W setting can highlight the lung regions while removing most of the anatomical information of the body regions, see the 3D image as shown in Figure 2. Furthermore, we cropped the image based on a coarse lung segmentation to reduce most redundant and useless regions, *e.g.*, the air regions of the CT image. Details of the segmentation are introduced in part D of Sec. III. To augment the total number of training samples, we extracted multiple 3D images from a certain nodule. For example, if a nodule covered five slices, we extracted five 3D images (one image per slice). A ground-truth bounding box (gt-bbox) is required for each nodule during training. Therefore, we generated the gt-bbox of a certain nodule according to the location and the size of the nodule: using the location as the center of the gt-bbox and using the diameter (in voxel-coordinates) as the length and width of the gt-bbox.

Normally, more slices should be used to build a 3D image, but our experiments demonstrated that the use of three slices was the right choice because: (1) Theoretically, three slices have enough information to distinguish most nodules from other tissues. A small nodule covers at most three or four slices in the volume data, and a large nodule should be discriminated from even just one slice. (2) Our model was developed from an existing model that was used for object detection on natural RGB images. Thus, the use of three slices made it possible to train our model from the pre-trained weights. This is the so-called transfer-learning [34]–[37]. (3) If more slices were used, it requires more samples to train a good model because more redundant information are included.

B. LUNA-16 DATASET

LUNA-16 [38] is another well-known open challenge in pulmonary detection. There are two tracks in the challenge: nodule detection (NDET) and false positive reduction (FPRED). Using raw CT scans, the goal of NDET is to identify locations of possible nodules and assign a probability for being a nodule to each location. The pipeline typically consists of candidate detection and FP reduction. Given a set of candidate locations, the goal of FPRED is to assign a probability for being a nodule to each candidate location.

The organizers of the challenge provided a total of 888 scans which were divided into ten subsets. All scans were selected from the Lung Image Database Consortium (LIDC-IDRI) [39]. The selection criteria were: (1) the slice thickness is less than or equal to 2.5 mm and (2) the nodule size is larger than or equal to 3 mm and accepted by at least three out of four radiologists. Based on the criteria, a total of 1186 nodules were selected. Non-nodules and the remaining nodules were referred to as irrelevant findings and were ignored during the evaluation.

In our study, all ten subsets were only used for evaluation. More details of the dataset can be found at the website of the challenge [38] and the online LIDC-IDRI database [39].

C. INDEPENDENT DATASET

In addition to the publicly available data, we collected 2470 chest scans from the cooperative hospital to augment the training samples (2440 scans) and for further validation (30 scans). The devices included SOMATOM Perspective, Sensation 16 from the Siemens company and Optima CT540 from the General Electric company. Scan protocols contained (120kV, 140 mAs) and (130 kV, 100 mAs). Images with size of 512 × 512 were reconstructed using filtered back projection (FBP) of lung kernel. The slice thickness included 1.25 mm and 1.5 mm. The average age of the included male and female patients was 49.6 years old.

For the training scans, annotating all 2440 scans is labor-intensive and time consuming. To reduce the total labeling time, we first used the TIANCHI AI datasets to train the model. Then, we applied the pre-trained model on the 2440 scans with a low scoring threshold value $T_1 = 0.1$ that was used to generate candidates (the definition of T_1 is shown in Fig. 3(c)). Finally, two subspecialized radiologists (radiologists $\boldsymbol{A}$ and $\boldsymbol{B}$, both have over ten years of experiences) of chest image were invited to make annotations based on the candidates. For a certain lesion (nodule or mass) of diameter less than 70 mm in the candidates, each slice of the nodule was annotated with a 2D bounding box and each box was assigned with a number that indicates the type of the nodule: number 1 for the ground-glass, number 2 for the part-solid and number 3 for the solid. The ImageJ [40] was used as the annotation tool. For the annotated nodules, we invited a senior radiologist of chest image to make a double check. Finally, a total of 9577 nodules were obtained: about 80% of the cases were small nodules of size smaller than 6 mm. About 16% of the cases had size ranging from 6mm to 30mm and 4% of the cases were larger than 30 mm. The nodule type distribution was as follows: ground-glass, part-solid and solid nodules comprised about 23%, 5.5% and 71.5%, respectively.

For the 30 validation scans, radiologists $\boldsymbol{A}$ and $\boldsymbol{B}$ were invited to find nodules from these scans in a blind fashion. Once all findings were obtained, radiologist $\boldsymbol{A}$ found nodules again from all scans without accessing to the first set of findings. So, in total there were three sets of independent findings to the validation scans: (1) the first set of findings from radiologist $\boldsymbol{A}$, which were double-checked by a

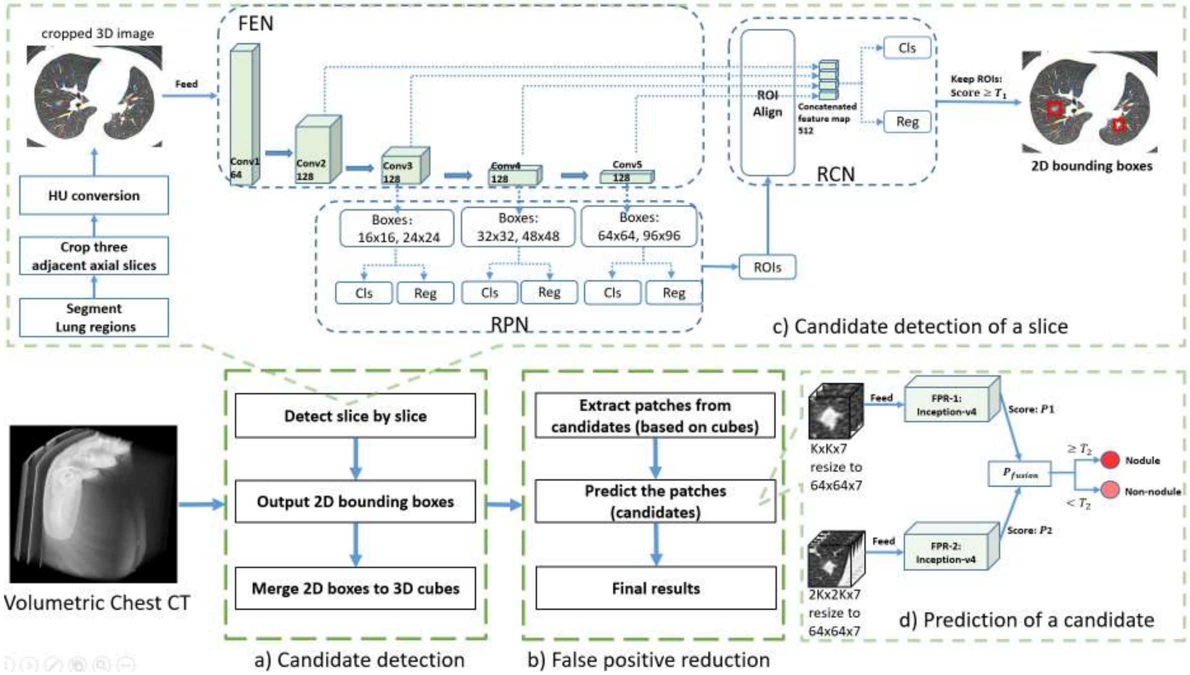


FIGURE 3. The architecture of the proposed CAD system which contains two major stages: (a) candidate detection and (b) FP reduction. For the candidate detection, a deep object detection model as shown in (c) was developed to located each candidate with bounding boxes. For the FP reduction, a CNN-based classifier built with two inception-v4 models as shown in (d) was designed to predict a candidate whether it is a nodule or a non-nodule.

senior radiologist. We used these findings as the ground-truth for comparison purpose. A total number of 90 nodules were found which included 23 ground-glasses, 11 part-solids and 56 solids, respectively. About 90% of the nodules were smaller than 6 mm. The other 10% cases had size ranging from 6mm to 30 mm. (2) The secondary set of findings from radiologist ***A***, which was used for intraobserver validation. (3) The set of findings from radiologist ***B***, which was used for interobserver validation.

III. METHODS

Figure 3 shows the architecture of the proposed CAD system. Like most existing systems, two major stages are incorporated: candidate detection as shown in Fig. 3(a) and FP reduction as shown in Fig. 3(b). Details of the proposed system are described in the next sections.

A. FASTER-RCNN

To achieve the goal of candidate detection, we designed a deep object detection model. We were inspired by the Faster-RCNN [24] which is an object detection model for natural images. The authors of Faster-RCNN first designed three networks, the FEN, RPN and RCN, and merged them together by sharing the convolutional features that extracted by the FEN from the input image.

In the RPN, the authors slide a 3 × 3 spatial window over the convolutional feature map output by the last shared convolutional layer and map each sliding window to a lower-dimensional feature (512-d for the VGG16 [32]). The feature is then fed into two sibling fully connected layers, a box-classification layer and a box-regression layer. The classification layer simultaneously predicts multiple boxes that are centered at the sliding-window location and associated with different scales and aspect ratios. For each box, the classification layer outputs two scores that estimate the probability of object (ROI) or not object. Meanwhile, the regression layer predicts four coefficients of each ROI which are used to move and scale the ROI to enclose the object more accurately.

After all ROIs are obtained and adjusted, their richer features are extracted from the last shared convolutional layer of the feature pyramid according to their locations and sizes using ROI-pooling [23]. Then, the extracted features are fed into RCN for another classification and regression, respectively. The function of the regression is similar to the counterpart in the RPN. The major difference lies in the classification. In RCN, the classification layer is a multi-category classifier which predicts multiple scores of each ROI. For example, it was implemented as a 21-category (20 classes plus the background) classifier on the PASCAL VOC dataset [45].

Faster-RCNN processes an image very quickly by sharing features between the RPN and RCN and it achieves promising results by taking full advantage of the deep features' strong semantic information. However, it has a drawback which is poor performance in detecting small objects. The root causes of this disadvantage can be described as: (1) the stride between boxes is too big. In the feature pyramid, the spatial resolution of features is decreased layer by layer as the depth increases. Two neighboring pixels in the feature map output by the last shared convolutional layer in VGG16 have sixteen pixels' stride after being mapped back to the original image. Therefore, the small objects located at the gaps between boxes may be neglected. (2) There is a lack of semantic information for classification of small objects. Much of the semantic information of the small objects has been lost in the last layer after the sub-sampling operations inside the feature pyramid. (3) A so-called "mis-alignment" issue as discussed in [25]. In the ROI-pooling layer, the location and size of an ROI is rescaled to the spatial space of the last layer and the results (floating numbers) are directly quantized to integers to extract features. The quantization will reduce precision, which is amplified when mapped back to the spatial space of the original image. For example, if the width of an ROI is 88 pixels, the corresponding width is 5 pixels in the last layer of VGG16 ($88 \div 16 = 5.5$, which is quantized to 5). The missed precision of 0.5 signifies 8 ($0.5 \times 16 = 8$) pixels' mis-alignment in the original image.

The above discussion demonstrates that it is a poor solution to directly apply the Faster-RCNN for nodule detection, because nodules are very small compared to most objects in natural images. A potential solution is the feature pyramid network (FPN) [27] which has better performance than Faster-RCNN in the detection of small objects by addressing the first issue. However, the secondary and the third issue remain in the FPN. Besides, the FPN is too complex and time-consuming for nodule detection, since it uses the ResNets [30] as the backbone of the convolutional network and it constructs a top-down architecture with lateral connections. Therefore, we designed a new model that should effectively avoid the three issues mentioned above. The proposed model is simple to implement and train based on the Faster-RCNN framework. Additionally, only a small computational overhead is introduced, which makes the model fast enough for clinical applications.

B. MODEL FOR CANDIDATE DETECTION

The architecture of the proposed model is schematized in Fig. 3(c). Similar to the Faster-RCNN, the model consists of three sub-networks as well: the FEN, RPN and RCN.

In the FEN, we still use the VGG16 to extract a feature pyramid from the 3D image which is fed into the network. To accelerate the detection, the channels of the output feature of the three convolutional layers, *Conv3*, *Conv4* and *Conv5* are reduced to 128.

In the RPN, we assign boxes based on *Conv3*, *Conv4* and *Conv5*, respectively. Each layer has two settings of the base size: 16×16 and 24×24 on *Conv3* for small nodules, 32×32 and 48×48 on *Conv4* for medium nodules, and 64×64 and 96×96 on *Conv5* for large lesions such as cancer and cysts. Aside from the base size, three aspect ratios, 0.6, 1.0 and 1.65, are set for each base size for the consideration of lesion shapes. Our experiments show that these settings should cover lesions ranging in size from 3 mm to 70 mm.

Because nodule sizes vary widely, it is reasonable to assign boxes on multiple layers rather than on only a single layer. A single layer may present a lack of semantic information for nodules of certain sizes. Using multiple layers should avoid this phenomenon.

Finally, in the RCN, a high-level feature map with 512 channels is constructed for each ROI through the concatenation of four feature maps, which are extracted from *Conv2*, *Conv3*, *Conv4* and *Conv5*, respectively, using "ROI-align" method [25]. The feature map of an ROI is then fed into the *Cls* layer to classify the ROI as a candidate or a non-candidate.

The high-level feature map for each ROI, constructed by concatenating features from shallower and deeper layers, has richer semantic information for classification of both small and large nodules than that provided from only a single layer. The effectiveness of using multiple layers for object detection in natural images has been demonstrated in some variants of the Faster-RCNN, *e.g.*, the FPN [27] and the HyperNet [29]. However, the high-level information in the FPN and the HyperNet was obtained through the fusion and concatenation of different layers in the feature pyramid, respectively, which involves up-sampling and sub-sampling operations to make the shallower and the deeper layers have the same spatial resolutions for the fusion and the concatenation. Sampling operations can lead to missing information and an aliasing effect. Thus, additional convolutional operations were applied to extract more semantic features and reduce the aliasing effect, which requires greater computational time. Moreover, the "mis-alignment" issue becomes severe in the nodule detection. For example, if the pixel size of a CT image is 0.68 mm, an 8-pixel "mis-alignment" means 5.44 mm of "mis-alignment". Obviously, this will degrade the detection of nodules, especially for the small nodules. Thus, we use the "ROI-align" method, rather than the ROI-pooling, to extract the features for each ROI to reduce the "mis-alignment" issue. The effectiveness of this method was demonstrated in the Mask-RCNN [25] which was developed for instance segmentation.

C. TRAINING THE DETECTION MODEL

The aim of the RPN is to find the ROIs from all boxes. When training, we assigned a binary class label to each box. A box was set as a positive sample if it had an Intersection over Union (IoU) higher than an empirical value of 0.7 with any ground-truth box. And a box was selected as a negative sample if its IoU with all ground-truth boxes was lower than an empirical value of 0.4. Other boxes that were neither positive nor negative were ignored during the training. Because an image only contains few nodules, there is a great disparity in the proportion of positives and negatives. To alleviate the data imbalance issue, the total samples per image were constrained to 256 during the training (extra negatives are randomly ignored).

The target of the RCN is to reduce most FPs from the ROIs. Similar to the RPN training, we assigned a binary class label to each ROI. If an ROI had an IoU higher than an empirical value of 0.7 with any ground-truth box, it was set as a positive sample. An ROI was set as a negative sample if its IoU with all ground-truth boxes was lower than an empirical value of 0.6. Other boxes that were neither positive nor negative were ignored during the training. Besides, we jittered the ground-truth boxes and appended them to the ROIs as the positive samples.

Both RPN and RCN, if trained independently, will modify their convolutional layers in different ways. Therefore, the authors of Faster-RCNN developed three training schemes for sharing convolutional layers: Alternating training, approximate joint training and non-approximate joint training. In our implementation, we adopted the approximate joint training (end-to-end training) for simplicity. We implemented our framework on Google Tensorflow® and used the momentum optimizer to minimize the total loss of both RPN and RCN on a NVIDIA GTX 1080 GPU, with a total of 200k iterations and with a learning rate of 0.01 which was decreased by 9.9% per every 50k iterations. The loss function for the classification and regression layers is identical to the Faster-RCNN [24] framework. The input image was up-sampled such that its shorter side has 512 pixels, which can enlarge nodules and thus benefit the detection.

D. CANDIDATE DETECTION ON VOLUMETRIC DATA

The trained model is only used for detection of a 3D image. In practice, candidate detection should be performed on a CT volume data that is loaded from all the DICOM files of a scan. To achieve the goal, we detect candidates from the volume data slice by slice, from the first axial slice to the last of the scan. For each slice under detection, we first build a 3D image by concatenating the slice and the slices adjacent to its two sides. Then, we crop a sub-image from the 3D image and convert the CT numbers of the sub-image into gray values with an L/W setting of −700 HU/1000 HU. Finally, the sub-image is fed into the pre-trained model to produce candidates.

The sub-image is cropped based on a coarse lung region segmentation of the slice to make the detection focus on the lung regions. The procedure for the segmentation is shown in Fig. 4. The original image is first smoothed by using a 5 × 5 gaussian blur operator, followed by binarization with a threshold value of −300 HU. Connected components are then computed from the binary image through a labeling operation. The components, which are connected to the image boundary, are removed and the remaining ones are the lung regions. Finally, a bounding box (see the red rectangle in Fig. 4) of all remaining regions is computed as the area for cropping the sub-image. We can see that the segmentation is very simple. This is an advantage compared to most existing CAD systems which require an accurate lung segmentation. In many cases, accurate lung segmentation is very complex and not robust due to the appearance of lesions.

Slice-by-slice detection is adopted to reduce the FNs in the candidate detection. Using this scheme, a nodule may have more than one opportunity to be detected since it may cover several slices in the volumetric data.

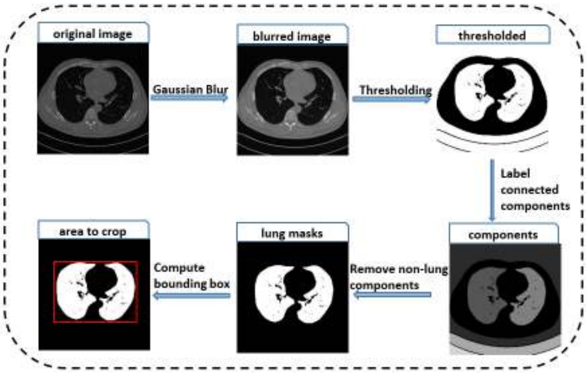


FIGURE 4. Major steps in the lung region segmentation. The original image is first smoothed by using a 5 × 5 gaussian blur operator, followed by binarization with a −300 HU threshold. Connected components are then computed from the binary image by using a labeling algorithm. The components connected to the image boundary are removed and the remaining ones are the lung regions. Finally, the bounding box (see the red rectangle) of all lung regions is adopted as the crop area for the sub-image.

After detection of all slices is completed, we can obtain a set of 2D boxes, each of which has six properties: x_1, y_1, x_2, y_2, z and *prob*, where (x_1, y_1) and (x_2, y_2), the up-left and bottom-right points of a box, indicate the location of a candidate in the axial slice (xy coordinates), z is the slice location, and *prob* is a score which indicates the possibility that the box encloses a nodule. A box is classified as a candidate if its score *prob* is larger than a user-defined threshold T_1 as shown in Fig.3 (c).

E. MERGING 2D BOXES

The target is to locate each 3D candidate from the 2D boxes. A candidate may have more than one 2D box. Thus, a clustering method was designed to find the boxes that belong to each candidate. Figure 5 shows the pseudo code for the clustering. We assign an ID to each 2D box and cluster the boxes with the same ID together. In a cluster B_k that has more than one box, each box b_i at least has another box b_j such that they are satisfied with the relation as,

$$\left|cp\left(b_i\right) - cp\left(b_j\right)\right| \leq T \quad \mathrm{i} \neq \mathrm{j}, \tag{1}$$

where $cp\left(\cdot\right)$ denotes the central point of a box and $|\cdot|$ means the Euclidean distance. T is a user-defined threshold and we set it to 3 mm (which should detect a nodule of 3 mm minimum size) in our implementation.

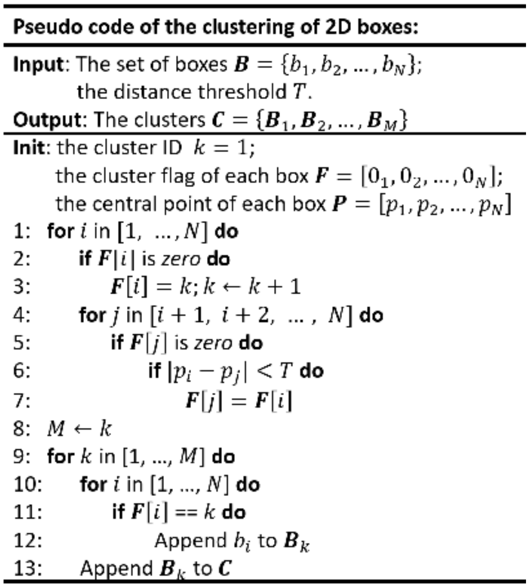
After all boxes are clustered, we compute the xy location of a 3D candidate by averaging the corner points of all boxes clustered to this candidate, as formulated in equations (2) to (5),

$$x_1' = \frac{\sum_{i=1}^{N} x_1^i}{N}, \tag{2}$$

$$x_2' = \frac{\sum_{i=1}^{N} x_2^i}{N}, \tag{3}$$

$$y_1' = \frac{\sum_{i=1}^{N} y_1^i}{N}, \tag{4}$$

$$y_2' = \frac{\sum_{i=1}^{N} y_2^i}{N}, \tag{5}$$



```
Pseudo code of the clustering of 2D boxes:
Input: The set of boxes $\boldsymbol{B} = \{b_1, b_2, \ldots, b_N\}$;
       the distance threshold $T$.
Output: The clusters $\boldsymbol{C} = \{\boldsymbol{B}_1, \boldsymbol{B}_2, \ldots, \boldsymbol{B}_M\}$
Init: the cluster ID $k = 1$;
      the cluster flag of each box $\boldsymbol{F} = [0_1, 0_2, \ldots, 0_N]$;
      the central point of each box $\boldsymbol{P} = [p_1, p_2, \ldots, p_N]$
for $i$ in $[1, \ldots, N]$ do
    if $\boldsymbol{F}|i|$ is zero do
        $\boldsymbol{F}[i] = k; k \leftarrow k + 1$
    for $j$ in $[i + 1, i + 2, \ldots, N]$ do
        if $\boldsymbol{F}|j|$ is zero do
            if $|p_i - p_j| < T$ do
                $\boldsymbol{F}|j| = \boldsymbol{F}|i|$
$M \leftarrow k$
for $k$ in $[1, \ldots, M]$ do
    for $i$ in $[1, \ldots, N]$ do
        if $\boldsymbol{F}[i] == k$ do
            Append $b_i$ to $\boldsymbol{B}_k$
    Append $\boldsymbol{B}_k$ to $\boldsymbol{C}$
```

FIGURE 5. The pseudo code for clustering the 2D boxes. Each box is first assigned with an ID. Then, the boxes with the same ID are clustered together.

where N is the number of 2D boxes of the candidate. We get the z location of the 3D candidate by computing the minimum and maximum slice locations from its boxes, as formulated in equations (6) to (7),

$$z'_1 = \min(\{z_i\}\, i = 1, 2, \ldots, N), \tag{6}$$

$$z'_2 = \max(\{z_i\}\, i = 1, 2, \ldots, N), \tag{7}$$

where (x'_1, y'_1, z'_1) and (x'_2, y'_2, z'_2) make up the two diagonal points of a bounding cube which encloses the 3D candidate.

F. FALSE POSITIVE REDUCTION

For each candidate, we extracted two 3D patches of size $K \times K \times 7$ and $2K \times 2K \times 7$, respectively, from the central point of the bounding cube, where 7 is the channels (seven adjacent axial slices) and K is the larger value of the height and width of the axial-plane of the cube. The voxel value of each patch is converted from the CT number with the L/W setting -600 HU/1000 HU.

To classify candidates into nodules and non-nodules, we trained two inception-v4 neural networks (denoted as FPR-1 and FPR-2 in Fig. 3(d)) independently on Google Tensorflow® using the patches of size $K \times K \times 7$ and $2K \times 2K \times 7$, respectively. In our implementation, we constrained the input size of the two networks to 64×64 (the channels are same with the patches) which means each patch must be resized before being fed into the networks. Under this restriction, the patch size of a small nodule will be amplified, which highlights the features of the nodule and makes it easier for prediction.

After two scores for each candidate are obtained separately from the two inception-v4 models, a final score is computed by fusing the two scores, as formulated in equation (8).

$$P_{fusion} = \gamma \times P_1 + (1 - \gamma) \times P_2, \tag{8}$$

where P_1 and P_2 are the predicted scores from FPR-1 and FPR-2, respectively. γ is the coefficient of the value between (0, 1] that is used to control the weights of the two scores in the fusion and the value was empirically set to 0.5 in our experiments. A candidate is classified as a nodule if its final score P_{fusion} is larger than a user-defined threshold T_2 as shown in Fig. 3(d).

Normally, other classifiers can be used for the false positive reduction, *e.g.*, the deep residual network [30], which is one of the state-of-the-art CNNs-based networks for image recognition. In our study, we chose the inception-v4 networks because inception-v4 can achieve better performance than the residual networks while retaining its computational efficiency, which has been demonstrated in the reference [31].

As discussed in reference [12], the size of the patches, called the receptive field, play a crucial role in the FP reduction. If the size of the receptive field is too small, only limited contextual information is exploited for training the models and then the discrimination capacity would be deficient in coping with large variations in detection targets. On the other hand, if the receptive field is too large, more redundant information may be included making it hard to train good models, especially when the number of training samples is quite limited. In practice, it is very hard to determine the optimal receptive field, so the authors of [12] designed three neural networks for different receptive fields to handle this issue, while the authors of [13] used multi-view convolutional networks (nine networks in their implementation). In our system, we only use two networks and the advantage is that the receptive fields for the networks are adaptive to the size of the nodules since the patches are cropped according to the bounding cubes.

To train the two models, we collected positive and negative samples from the candidates. A candidate was cropped as a positive sample if its location was inside the radius of a ground-truth nodule, otherwise it was cropped as a negative one. To address the data-imbalance issue, we augmented the positives by rotating each patch to 0, 90, 180 and 270 degrees. We trained the models on a NVIDIA GTX 1080 GPU using the momentum optimizer with a decay of 0.9 to reduce the focal loss [28] in 50k iterations. We used a learning rate of 0.01, decayed per every 10k iterations using an exponential rate of 0.1.

IV. EXPERIMENTS

A. EVALUATION ON THE LUNA-16 DATASET

In the LUNA-16 challenge, results were evaluated by measuring sensitivity and average FPs per scan. A detected candidate was counted as a true positive if its location was inside the radius of a ground-truth nodule. The Free-Response Receiver Operating Characteristic (FROC) curve [41] was adopted to analysis the overall performance of a CAD system. Based on the FROC curve, a competition performance metric (CPM)

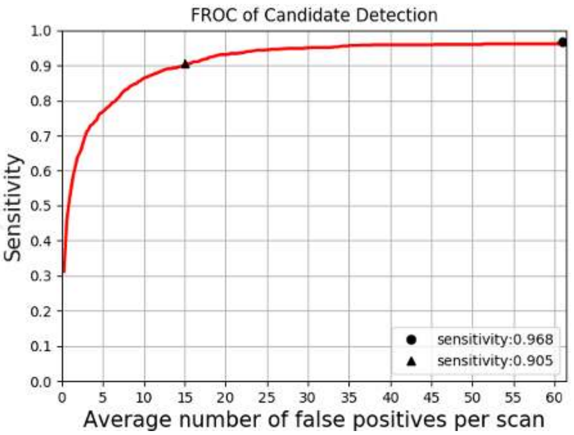


FIGURE 6. FROC curves of the proposed candidate detection on the LUNA-16 datasets. The proposed system achieved a maximum sensitivity of 96.8% at about 60 FPs per scan.

TABLE 1. Candidate detection of different systems on the luna-16 datasets.

System	Combination	Sensitivity	Candidates/scan
Ours		0.968	60.23
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
Combination 1	●○○○●	0.959	590.2
Combination 2	●●○○●	0.977	820.0
Combination 3	●●●○●	0.983	825.3
Combination 4	●●●●●	0.983	850.2

score of each system was calculated for comparison as the average sensitivity at seven predefined FPs rates: 1/8, 1/4, 1/2, 1, 2, 4 and 8 FPs per scan. We also used FROC and CPM to evaluate the proposed system and compared it to the state-of-the-art systems that were submitted to the challenge.

Figure 6 shows the curve of the proposed candidate detection. We can see that the system achieved a maximum sensitivity of 96.8% at about 60 FPs per scan. With only 15 FPs per scan, the system still approached a sensitivity of about 90%.

In clinical practice, a suitable threshold value of T_1 should be chosen to achieve lowest rates of FPs per scan at maximum sensitivity. Larger value of T_1 can lead to fewer FPs but at the cost of lower sensitivity. Our experiments demonstrated that the best result was achieved for the LUNA-16 datasets when T_1 was set to about 0.4. However, a smaller value is suggested in clinical applications to ensure high sensitivity. For the LUNA-16 datasets, the average FPs per scan is only about 60 when T_1 is set to 0.1, which is still a low level compared to most existing systems. For all experiments in this study, T_1 was set to 0.1.

For the candidate detection, it is hard to make a comprehensive comparison with state-of-the-art systems, because most published works focused on false positive reduction. Therefore, we compared our results to five existing systems that were used in the challenge to produce candidates for the FPRED track. All five systems were implemented using traditional methods. Table 1 summarizes the comparison of our candidate detection to the five systems and their potential combinations (in the table, black circles indicate combined methods, *e.g.*, in the row *Combination 4*, the five black circles mean the combined results from all five systems). The results for the five systems and their combinations are from table 1 of [18]. Table 1 shows that, for any individual system, the highest sensitivity was achieved by the proposed system. When the results of multiple existing systems were combined, the sensitivity was remarkably improved and the highest reached 98.3% which is higher than the proposed system. However, the average number of candidates per scan was substantially increased as well. For example, when the five systems were combined (see the row *Combination 4* in Table 1), the average number of candidates per scan was 850.2 which is about 14 times our result. Another drawback of the combination of multiple systems is that it requires more computational time because of the multiple detections per scan.

For the FP reduction, we compared our system to nine state-of-the-art systems, which achieved top-9 results in the NDET track [38]. All systems were developed using deep-learning techniques. However, no details could be obtained since these systems had no published papers yet and only limited descriptions. Table 2 shows the results. The average score (CPM) of the proposed system is lower than the *zhongliu_xie* system which reached the fifth place in the competition. Inspection of the table shows that if FP rates are smaller than 1 FP per scan, our results are much worse than the highest sensitivity achieved by the top-9 systems (see the results shown in bold). When FP rates are larger than 0.5 FPs per scan, the gaps become much smaller and the maximum gap is only 0.028 between our system and the *JianPeiCAD* at the rate of 1 FP per scan. It is hard to conduct deep discussions on the algorithms used in these systems, since fewer details can be obtained. A major difference between the proposed system and these existing systems is the data source. The models in the proposed system were trained from the TIANCHI AI datasets and the independent datasets, while the models of the existing systems were trained from the LUNA16 datasets using 10-fold cross-validation.

In clinical practice, the threshold value of T_2 also has a significant impact on the performance of the proposed system. Larger value of T_2 can generate fewer FPs per scan, but it also can cause lower sensitivity. An empirical value of 0.8 was adopted in our study which met a compromise of sensitivity and FPs rate according to the suggestion of our cooperative doctors.

B. EVALUATION ON AN INDEPENDENT DATASET

In the LUNA-16 challenge, only 1186 nodules were provided for both training and validation, which were selected from

TABLE 2. The results of the false positive reduction of the proposed system and other systems on the luna-16 datasets.

System	Date	0.125	0.25	0.5	1	2	4	8	CPM
Ours	Nov. 2018	0.788	0.847	0.895	0.934	0.952	0.959	0.963	0.903
PAtech	2 Jan. 2018	**0.919**	0.923	0.933	0.953	0.969	0.977	**0.983**	**0.951**
JianPeiCAD	22 Dec. 2017	0.881	**0.941**	**0.960**	**0.962**	0.967	0.969	0.970	0.950
LUNA16FONOVACAD	28 Nov. 2017	0.917	0.929	0.943	0.956	0.958	0.962	0.964	0.947
iFLYTEK-MIG	17 Aug. 2017	0.842	0.905	0.937	0.961	**0.979**	**0.981**	0.982	0.941
zhongliu_xie	29 Sep. 2017	0.805	0.858	0.919	0.952	0.967	0.976	0.977	0.922
iDST-VC	13 Jul. 2017	0.740	0.826	0.890	0.928	0.953	0.969	0.973	0.897
qfpxfd	27 May 2017	0.744	0.853	0.887	0.922	0.938	0.945	0.948	0.891
CASED	15 Jun. 2017	0.771	0.828	0.872	0.906	0.929	0.942	0.961	0.887
3DCNN_NDET	22 Jun. 2017	0.731	0.810	0.874	0.918	0.938	0.946	0.957	0.882

The highest score of each column is shown in bold

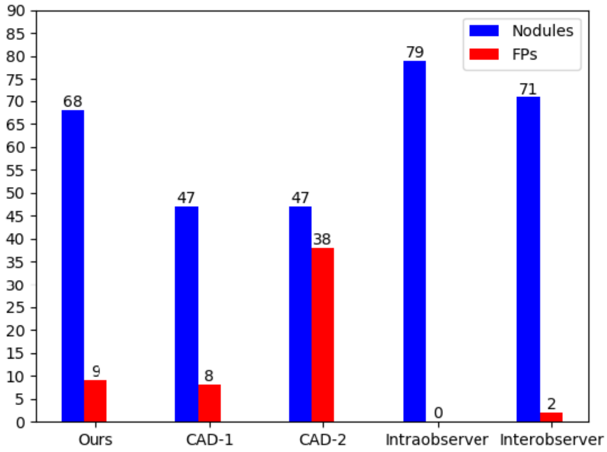


FIGURE 7. The results of the 30 validation scans. The proposed system is superior to both CAD-1 and CAD-2 in the detection of nodules. All systems are inferior to manual screenings.

the LIDC-IDRI database based on the criteria as introduced in part B of Sec. II. Besides, Non-nodules and remaining nodules were referred to as irrelevant findings and were ignored during the evaluation. In clinical applications, the scenarios may be much more complex. Thus, we collected 30 scans from the cooperative hospital for further evaluation. We compared our system to two commercial CAD systems, CAD-1 [43] and CAD-2 [44], which have been widely accepted by doctors in many top hospitals in China as a tool for nodule screening. Both commercial systems were developed in recent years using deep learning techniques. The details are not available because of confidentiality rules. In addition, we compared our system to human observers: intraobserver and interobserver.

Figure 7 shows the statistical results. All results were counted according to the ground-truth (the first set of findings of radiologist ***A*** as introduced in part C of Sec. II). Coincidentally, both CAD-1 and CAD-2 screened out 47 nodules from a total of 90 nodules, respectively. The sensitivity is only about 52.2%. CAD-1 generates only 8 FPs while CAD-2 produced 38 FPs. The proposed system found 68 nodules and 9 FPs. The sensitivity (about 75.6%) is much higher than both CAD-1 and CAD-2. The number of FPs is close to CAD-1. By analyzing the FNs, we found that most of the FNs produced in all systems were the small nodules of size smaller than 5 mm. The proposed system outperforms both CAD-1 and CAD-2 in the detection of small nodules, which is the major reason that the proposed system has higher sensitivity. From the results of intraobserver and interobserver, it can be concluded that FNs and FPs are unavoidable from even manual screenings (in clinical practice, there is no golden-standard between human observers for identification of nodules). However, all automatic detections are inferior to the manual screenings. By contrast, the sensitivity 75.6% of the proposed system is more comparable to the intraobserver and interobserver (87.8% and 78.9%, respectively). One should note that the radiologists found nodules intentionally from the 30 validation scans in one session for our study. Therefore, it is possible that they were more careful and consistent in their findings than in a normal clinical job.

V. DISCUSSION

In this study, a CAD system was developed for fast and accurate detection of pulmonary nodules in CT images. In contrast to most existing works that focus on false positive reduction, we concentrate on the candidate detection in our study. A deep object detection model was designed and trained from 3D images. The major advantages of the proposed model can be summarized as:

(1) The proposed model can simultaneously detect nodules and masses with a broad spectrum of appearance, regardless of their types, sizes and locations. Especially, it has superior performance in the detection of small nodules most often seen in the clinic that are difficult to find even by experienced radiologists, *e.g.*, juxta-vascular nodules of type ground glass.

(2) The proposed model can simplify the design of the false positive reduction. In our candidate detection, each candidate is indicated with a bounding cube which not only provides location information but also offers the rough size information. The size information is very useful for design of the false positive reduction, since the effects of positive reduction heavily reply on the receptive-fields.

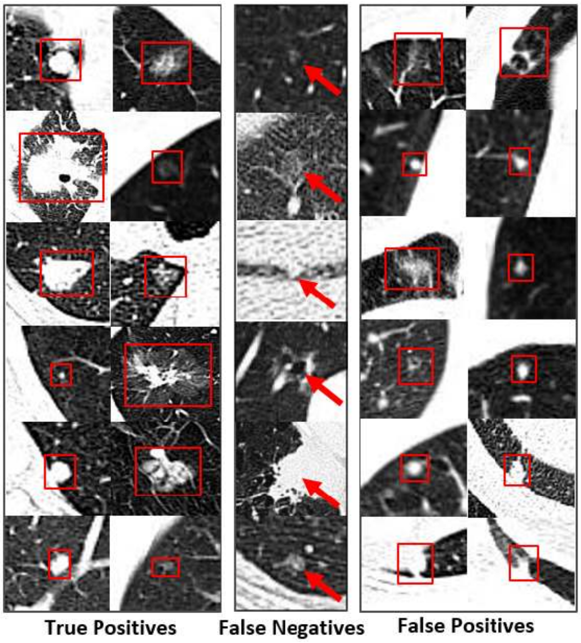


FIGURE 8. Examples of true positives (the left sets), false negatives (the middle sets) and false positives (the right sets).

(3) The proposed candidate detection is an independent module, so it can be incorporated with any other false positive reduction methods. It also can be used to implement automatic tools that generate candidates for training false positive reduction. Besides, it can be adopted as a potential solution for other similar clinical application.

(4) The proposed candidate detection is very fast. For a slice of size 512 × 512, the detection time is only about 30 milliseconds when running the system on a Geforce GTX 1080 GPU.

After candidate detection, we combined two inception-v4 networks of different receptive fields as a classifier for FP reduction. As demonstrated in Sec. IV, the classifier can effectively reduce most FPs while keeping a relatively high sensitivity. However, due to the specificity of scans, it is impossible to find all nodules without any FPs. Figure 8 shows some examples of true positives, false negatives and false positives. These examples show that the proposed system should detect nodules of varied appearance (see the left sets in Fig. 8). Yet, it may fail to find some nodules (see the red arrows in the middle sets in Fig. 8). In addition, some nodule-like tissues (see the right sets in Fig. 8) may be recognized as nodules. In our experiments, we found that most false negatives were screened out in the candidate detection but missed in the FP reduction. Collecting more samples for training the classifier should alleviate this issue. Another potential solution is to use more complex FP reduction methods, *e.g.*, the 3D CNNs used in [12] and [16].

Nodule detection is the first task in our study. In the future, we will focus on estimation of the nodule's size [42], type (ground-glass, part solid, solid, benign and malignant, etc.) and location (18 pulmonary segments). These properties of a nodule should provide useful information for a doctor to help in establishing more accurate treatment plans.

VI. CONCLUSION

We presented a CAD system for nodule detection in CT images and demonstrated the importance of candidate detection in the system. Candidate detection with high sensitivity and low ratios of FPs effectively reduced the complexity of the design for FP reduction. We achieved promising results using two convolutional neural networks as classifiers for the FP reduction. Our experiments show that the proposed system is an effective lung cancer screening tool in clinical applications.

ACKNOWLEDGMENT

(Jun Wang and Jiawei Wang are co-first authors.)

REFERENCES

[1] J. K. Field, M. Oudkerk, J. H. Pedersen, and S. W. Duffy, "Prospects for population screening and diagnosis of lung cancer," *Lancet*, vol. 382, no. 9893, pp. 732–741, Aug. 2013.

[2] E. T. Scholten *et al.*, "Computed tomographic characteristics of interval and post screen carcinomas in lung cancer screening," *Eur. Radiol.*, vol. 25, no. 1, pp. 81–88, Jan. 2015.

[3] M. Liang *et al.*, "Low-dose CT screening for lung cancer: Computer-aided detection of missed lung cancers," *Radiology*, vol. 281, no. 1, pp. 279–288, Oct. 2016.

[4] C. Jacobs *et al.*, "Automatic detection of subsolid pulmonary nodules in thoracic computed tomography images," *Med. Image Anal.*, vol. 18, pp. 374–384, Feb. 2014.

[5] A. A. Setio, C. Jacobs, J. Gelderblom, and B. van Ginneken, "Automatic detection of large pulmonary solid nodules in thoracic CT images," *Med. Phys.*, vol. 42, no. 10, pp. 5642–5653, Oct. 2015.

[6] K. Murphy, B. van Ginneken, A. M. R. Schilham, B. J. de Hoop, H. A. Gietema, and M. Prokop, "A large-scale evaluation of automatic pulmonary nodule detection in chest CT using local image features and k-nearest-neighbour classification," *Med. Image Anal.*, vol. 13, no. 5, pp. 757–770, Oct. 2009.

[7] X. Ye, X. Lin, J. Dehmeshki, G. Slabaugh, and G. Beddoe, "Shape-based computer-aided detection of lung nodules in thoracic CT images," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 7, pp. 1810–1820, Jul. 2009.

[8] F. Zhang *et al.*, "Lung nodule classification with multilevel patch-based context analysis," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 4, pp. 1155–1166, Apr. 2014.

[9] F. Ciompi *et al.*, "Bag-of-Frequencies: A descriptor of pulmonary nodules in computed tomography images," *IEEE Trans. Med. Imag.*, vol. 34, no. 4, pp. 962–973, Apr. 2015.

[10] W. Shen, M. Zhou, F. Yang, C. Yang, and J. Tian, "Multi-scale convolutional neural networks for lung nodule classification," in *Information Processing in Medical Imaging*, vol. 9123. Cham, Switzerland: Springer, 2015, pp. 588–599.

[11] S. Hussein, K. Cao, Q. Song, and U. Bagci. (2017). "Risk stratification of lung nodules using 3D CNN-based multi-task learning." [Online]. Available: https://arxiv.org/abs/1704.08797

[12] Q. Dou, H. Chen, L. Yu, J. Qin, and P.-A. Heng, "Multilevel contextual 3-D CNNs for false positive reduction in pulmonary nodule detection," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 7, pp. 1558–1567, Jul. 2017.

[13] A. A. A. Setio *et al.*, "Pulmonary nodule detection in CT images: False positive reduction using multi-view convolutional networks," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1160–1168, May 2016.

[14] H. Jin, Z. Li, R. Tong, and L. Lin, "A deep 3D residual CNN for false-positive reduction in pulmonary nodule detection," *Med. Phys.*, vol. 45, no. 5, pp. 2097–2107, May 2018.

[15] Q. Dou, H. Chen, Y. M. Jin, H. Lin, J. Qin, and P. A. Heng, "Automated pulmonary nodule detection via 3D convnets with online sample filtering and hybrid-loss residual learning," *Medical Image Computing and Computer Assisted Intervention-MICCAI* (Lecture Notes in Computer Science), vol. 10435, M. Descoteaux, Eds. *et al.* Berlin, Germany: Springer, 2017, pp. 630–638.

[16] J. Ding, A. Li, Z. Hu, and L. Wang. (2017). "Accurate pulmonary nodule detection in computed tomography images using deep convolutional neural networks." [Online]. Available: https://arxiv.org/abs/1706.04303

[17] S. Chen *et al.*, "Automatic scoring of multiple semantic attributes with multi-task feature leverage: A study on pulmonary nodules in CT images," *IEEE Trans. Med. Imag.*, vol. 36, no. 3, pp. 802–812, Mar. 2017.

[18] A. A. A. Setio *et al.*, "Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge," *Med. Image Anal.*, vol. 42, pp. 1–13, Dec. 2017.

[19] W. Zhu, C. Liu, W. Fan, and X. Xie, "DeepLung: Deep 3D dual path nets for automated pulmonary nodule detection and classification," in *Proc. WACV*, Lake Tahoe, NV, USA, Mar. 2018, pp. 673–681.

[20] W. Zhu, Y. S. Vang, Y. Huang, and X. Xie, "DeepEM: Deep 3D ConvNets with EM for weakly supervised pulmonary nodule detection," in *Proc. MICCAI*, Granada, Spain, Sep. 2018, pp. 812–820.

[21] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI* (Lecture Notes in Computer Science), vol. 9351. Cham, Switzerland: Springer, 2015, pp. 234–241.

[22] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. CVPR*, Boston, MA, USA, Jun. 2015, pp. 3431–3440.

[23] R. Girshick. "Fast R-CNN." [Online]. Available: https://arxiv.org/abs/1504.08083

[24] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. NIPS*, Montreal, Canada, 2015, pp. 91–99.

[25] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," in *Proc. ICCV*, Venice, Italy, Oct. 2017, pp. 2980–2988.

[26] W. Liu *et al.*, "SSD: Single shot multibox detector," in *Proc. ECCV*, Amsterdam, The Netherlands, 2016, pp. 21–37.

[27] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. CVPR*, Honolulu, HI, USA, Jul. 2017, pp. 2117–2125.

[28] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in *Proc. ICCV*, Venice, Italy, Oct. 2017, pp. 2980–2988.

[29] T. Kong, A. Yao, Y. Chen, and F. Sun, "HyperNet: Towards accurate region proposal generation and joint object detection," in *Proc. CVPR*, Las Vegas, NV, USA, Jun. 2016, pp. 845–853.

[30] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. CVPR*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778.

[31] C. Szegedy, S. loffe, V. Vanhoucke, and A. Alemi. (2016). "Inception-v4, inception-ResNet and the impact of residual connections on learning," [Online]. Available: https://arxiv.org/abs/1602.07261

[32] K. Simonyan and A. Zisserman. (2014). "Very deep convolutional networks for large-scale image recognition." [Online]. Available: https://arxiv.org/abs/1409.1556

[33] TIANCHI challenge. (2017). *First Season: Lung Nodule Detection*. [Online]. Available: https://tianchi.aliyun.com/competition/introduction.html

[34] Y. Wei, Y. Zhang, and Q. Yang. (2017). "Learning to transfer." [Online]. Available: https://arxiv.org/abs/1708.05629

[35] A. R. Zamir, A. Sax, W. Shen, L. Guibas, J. Malik, and S. Savarese, "Taskonomy: Disentangling task transfer learning," in *Proc. CVPR*, Salt Lake City, UT, USA, Jun. 2018, pp. 3712–3722.

[36] A. Diba *et al.*, "Temporal 3D convnets: New architecture and transfer learning for video classification," [Online]. Available: https://arxiv.org/abs/1711.08200

[37] D. S. Kermany *et al.*, "Temporal 3D ConvNets: New architecture and transfer learning for video classification," *Cell*, vol. 172, no. 5, pp. 1122–1131, 2018.

[38] C. Jacobs, A. A. A. Setio, A. Traverso, and B. van Ginneken. (2017). *LUNA-16: Lung Nodule Analysis*. [Online]. Available: https://luna16.grand-challenge.org

[39] A. G. Armato, III, *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.*, vol. 38, no. 2, pp. 915–931, Feb. 2011.

[40] *ImageJ*. Accessed: Jan. 11, 2018. [Online]. Available: https://imagej.nih.gov/ij/

[41] P. M. DeLuca, A. Wambersie, and G. F. Whitmore, "Receiver operating characteristic analysis in medical imaging: Contents," *J. Int. Commission Radiat. Units Meas.*, vol. 8, no. 1, pp. 1–62, Apr. 2008.

[42] M. M. Farhangi, H. Frigui, A. Seow, and A. A. Amini, "3-D active contour segmentation based on sparse linear combination of training shapes (SCoTS)," *IEEE Trans. Med. Imag.*, vol. 36, no. 11, pp. 2239–2249, Nov. 2017.

[43] *YITU Healthcare: Care.aiTM*. Accessed: Jun. 13, 2018. [Online]. Available: http://www.yitutech.com/corebusiness/2.html

[44] *Infervision: AI-CTTM*. Accessed: Jun. 13, 2018. [Online]. Available: http://www.infervision.com/Infer/product-en

[45] M. Everingham, L. Van Gool, C. K. I. Wililams, J. Winn, and A. Zisserman, "The pascal visual object classes (VOC) challenge," *Int. J. Comput. Vis.*, vol. 88, pp. 303–338, 2010. [Online]. Available: http://www.pascal-network.org/challenges/VOC/voc2007/index.html

[46] Y.-D. Zhang, Y. Zhang, X.-X. Hou, H. Chen, and S.-H. Wang, "Seven-layer deep neural network based on sparse autoencoder for voxelwise detection of cerebral microbleed," *Multimedia Tools Appl.*, vol. 77, no. 9, pp. 10521–10538, May 2018.

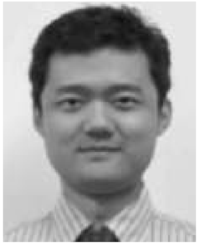
[47] W. Zhu *et al.*, "AnatomyNet: Deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy," *Med. Phys.*, vol. 46, no. 2, pp. 576–589, Feb. 2018.

JUN WANG received the M.D. degree in computer science from Southeast University, China. He is currently pursuing the Ph.D. degree in biological physics with the Institute of Translational Medicine, Zhejiang University, China. He has work experience of five years in MinFound Medical Systems Co., Ltd. as an Algorithm Engineer and Senior Software Engineer. His main research interests include deep learning and medical image processing.

JIAWEI WANG received the master's degree in imaging and nuclear medicine from Zhejiang University, China, in 2008. From 2008 to 2011, he was a Resident with the Department of Radiology, The Second Affiliated Hospital, Zhejiang University School of Medicine, where he has been an Attending Doctor, since 2012. He is engaged in CAD of medical images.

YAOFENG WEN received the B.S.E. degree from the Nanjing University of Posts and Telecommunications in China, in 2001, and the M.S.E. degree and the Ph.D. degree in BME from Zhejiang University in China, in 2005 and 2009, respectively. He has held various engineering positions at Huawei, OminiVision, and Accel Semiconductor. He was the Co-Founder of Bellnet Inc., in Wuxi, China. He is currently a Researcher with the Shanghai Industrial Technology Institute.

HONGBING LU received the Ph.D. degree from Zhejiang University, China, in 1995, where he is currently an Associate Professor with the College of Computer Science and Technology. His current research interests include big data, machine learning, and intelligent controller.



TIANYE NIU received the Ph.D. degree from the University of Science and Technology of China, in 2009. From 2009 to 2013, he held a postdoctoral position in the medical physics program with the Georgia Institute of Technology, USA. He is currently a Faculty Member with the Institute of Translational Medicine, Zhejiang University, and the Affiliated Sir Run Run Shaw Hospital, Zhejiang University School of Medicine.

JIANGFENG PAN is currently with Jinhua Municipal Central Hospital as a Chief Physician and an Associate Director of the Medical Imaging Department. He has over ten years' experience and specializes in imaging diagnosis of thoracopathy. Every year, he leads his team to conduct pulmonary nodule screening over 30 000 patients using low-dose CT.

DAHONG QIAN received the B.S.E. degree from Zhejiang University, China, in 1988, the M.S.E. degree from The University of Texas at Austin, Austin, TX, USA, in 1991, and the Ph.D. degree in computer science from Harvard University, Cambridge, MA, in 2002. He has held various engineering and management positions at Maxim, Analog Devices, and OmniVision. He was the Co-Founder of InnoPhase, San Diego, CA. He is also a Professor with the Medical Robot Institute, Shanghai Jiao Tong University.