IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 22, 2019, accepted March 13, 2019, date of publication March 18, 2019, date of current version April 5, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2905574

Lung Nodule Detection With Deep Learning in 3D Thoracic MR Images

YANFENG LI[1], LINLIN ZHANG[1], HOUJIN CHEN[1], AND NA YANG[2]
[1]School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100044, China
[2]School of Information and Communication Engineering, North University of China, Taiyuan 030051, China

Corresponding author: Houjin Chen (hjchen@bjtu.edu.cn)

This work was supported in part by the Fundamental Research Funds for the Central Universities under Grant 2018JBM004, and in part by the National Natural Science Foundation of China under Grant 61872030 and Grant 61571036.

ABSTRACT Early detection of lung cancer is crucial in reducing mortality. Magnetic resonance imaging (MRI) may be a viable imaging technique for lung cancer detection. Numerous lung nodule detection methods have been studied for computed tomography (CT) images. However, to the best of our knowledge, no detection methods have been carried out for the MR images. In this paper, a lung nodule detection method based on deep learning is proposed for thoracic MR images. With parameter optimizing, spatial three-channel input construction, and transfer learning, a faster R-convolution neural network (CNN) is designed to locate the lung nodule region. Then, a false positive (FP) reduction scheme based on anatomical characteristics is designed to reduce FPs and preserve the true nodule. The proposed method is tested on 142 T2-weighted MR scans from the First Affiliated Hospital of Guangzhou Medical University. The sensitivity of the proposed method is 85.2% with 3.47 FPs per scan. The experimental results demonstrate that the designed faster R-CNN network and the FP reduction scheme are effective in the lung nodule detection and the FP reduction for MR images.

INDEX TERMS Lung nodule detection, T2-weighted MR images, faster R-CNN, FP reduction.

I. INTRODUCTION

Lung cancer has the highest rate of mortality among all type of cancers [1]. The 5-year survival rate for all stages of lung cancer is only 16%. When lung cancer is detected at a localized stage, the 5-year survival rate will increase to 52% [2]. At present, early detection of lung cancer is the cornerstone to reduce mortality [3]. At present, computed tomography (CT) is a widely used technique for lung nodule detection. Recently, with the development of magnetic resonance imaging (MRI) technique, MRI has also been used in lung disease diagnosis [4]. Though CT is superior to MRI in small nodule detection, MRI has its own advantages. Compared with CT, MRI is a non-radiation examination. There were no significant differences in malignant nodule detection between these two techniques [5]. MRI might be a valuable tool in malignant nodule detection and lung cancer screening [6]. Moreover, MRI can provide not only morphological, but also functional information, such as physiological, pathophysiological, and molecular information. Thus MRI plays an important role in lung disease diagnosis.

MRI has multiple modalities and each MRI scan has many slices, making the dataset of each case very huge. Moreover, the diagnosis sensitivity is affected by radiologist's experience and visual fatigue. Thus computer-aided detection (CAD) for MR images is demanding [7]. Pulmonary nodule detection is the basis for nodule measurement and classification. The study on automatic pulmonary nodule detection is vital.

In general, nodule detection methods consist of four steps: preprocessing, lung parenchyma segmentation, nodule detection and false positive (FP) reduction [8]. Sousa *et al.* [9] first extracted the thorax and lung parenchyma based on region growing. Then a lung reconstruction method using rolling-ball algorithm was designed in order to refine the parenchyma contour. After that, dense structures inside the lung parenchyma were selected. Finally a support vector machine (SVM) classifier was trained to classify the dense structure as nodule or FP. This method was tested on 33 CT exams and obtained 0.42 FPs/scan with 0.15 false negative (FN) rate. Cascio *et al.* [10] developed a detec-

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

2169-3536 © 2019 IEEE. Translations and content mining are permitted for academic research only.
Personal use is also permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

tion method for internal and juxta pleural nodules. First nodule candidate was detected and segmented based on 3D mass–spring model. Then seven features described by shape, size and intensity were extracted. Finally an FP reduction algorithm was designed. This method was tested on a set of 84 scans. The detection sensitivity is 88% at 2.5 FPs/scan. Choi and Choi [11] proposed a 3D shape-based feature descriptor to detect pulmonary nodules in CT scans. First lung volume was segmented combining thresholding with 3D connected component labeling. Then nodule candidates were detected through multi-scale dot enhancement filtering. Finally 3D shape-based feature descriptors were extracted and an SVM classifier was trained. This method achieved 97.5% sensitivity with 6.76 FPs/scan. Santos *et al.* [12] developed a detection method for small lung nodule. Region growing technique was first used to segment the pulmonary parenchyma. Then the Gaussian mixture models and the Hessian matrix were combined to get the nodule candidate. Finally Tsallis's and Shannon's entropy measurements were extracted, and SVM was trained to classify each candidate as nodule or non-nodule. This method was tested on 28 CT scans. The detection sensitivity was 90.6% and the specificity was 85%. Filho *et al.* [13] first extracted and reconstructed the pulmonary parenchyma. Then nodule candidates were detected. Finally, shape and texture features were extracted and an SVM classifier was trained. Javaid *et al.* [14] developed a detection method for juxta vascular and juxta pleural nodules. 110 CT scans were selected to test the detection performance. The sensitivity of this method was 91.65% with 3.19 FPs/scan.

Most previous nodule detection methods require handcrafted feature extraction. Recently convolution neural network (CNN) which automatically discovers features, shows promising results in many pattern recognition tasks. This inspired the application of CNN in automated pulmonary nodule detection [15]–[21]. Jiang *et al.* [15] first detected the nodule candidate through traditional method. Then a patch was cropped for each candidate and CNN was used to classify it as normal or nodule. The sensitivity of this method was 80.06% with 4.7 FPs/scan. Huang *et al.* [18] gave a nodule detection method based on 3D CNN. This method has two steps: nodule candidate detection through local geometric-model based filtering and candidate classifying using 3D CNN. Li *et al.* [20] constructed a shallow CNN to classify each suspicious region as normal or nodule for CT scans. Li *et al.* [21] employed three CNNs with different input sizes and different depths to classify image patches as nodule or normal.

Numerous lung nodule detection methods have been studied for CT images. For MR images, few detection methods have been carried out. CT has better image contrast and less air artifacts. Due to the image differences between CT and MR, nodule detection methods for CT images may be not suitable for MR images. In this paper, we propose a lung nodule detection method for thoracic MR images based on deep learning. For CT nodule detection, most deep learning based methods consist of two steps. First nodule candidate was extracted and cropped. Then a deep learning network was designed to classify the candidate as nodule or normal. As the size of lung nodule differs a lot, cropping each region with a fixed size may be not reasonable. To solve this problem, Faster R-CNN is designed for lung nodule detection in this paper. Faster R-CNN takes the whole image as input and no candidate extraction is needed. Moreover, it can detect objects with different sizes as multiple anchors are introduced. To add spatial information in the input image, a spatial three-channel input is constructed. To overcome the problem of overfitting, transfer learning is applied in parameter optimization. As Faster R-CNN does not consider anatomical characteristics, many FP regions exist in the detection results. To reduce FPs and preserve true nodule, a FP reduction scheme is designed based on the anatomical characteristics of lung nodule.

The rest of this paper is organized as follows: Section 2 presents the dataset. Section 3 describes the proposed method. Experimental setup is explained in Section 4. Results and discussions are given in Section 5. Conclusions are summarized in Section 6.

II. DATASET

The data used in this paper is T2-weighted MR scans from the First Affiliated Hospital of Guangzhou Medical University. All MR scans were performed on a 3T MR imager. Subjects were placed in the supine position. This dataset includes 142 T2W-MR scans. Each scan consists of 13 to 33 slices, with a 7.7mm spatial resolution along the axial direction. The total number of slices is 3403. In one scan, the size of each slice is 576×576. In the other scans, the size of each slice is 640×640. The spatial resolution within each slice ranges from 0.6597 mm to 0.6719 mm. For the 3403 slices, 800 slices are with lung nodule. Some slices contain more than one lung nodule region, resulting in a total number of 862 nodule regions. According to the position, lung nodules can be divided into isolated nodule, juxta vascular nodule and juxta pleural nodule. Examples of these nodules are shown in Fig. 1.

III. METHOD

To avoid candidate extraction and be less dependent on scale, Faster R-CNN is designed for lung nodule detection. Then an FP reduction scheme is proposed based on the anatomical characters of lung nodule.

A. LUNG NODULE DETECTION

1) FASTER R-CNN

Faster R-CNN is composed of two modules [22]. The first module is a Region Proposal Network (RPN), which can generate proposed regions for each image. The second module is the Fast R-CNN detector, which classifies the region proposals. To save time, RPN and Fast R-CNN share convolutional layers.

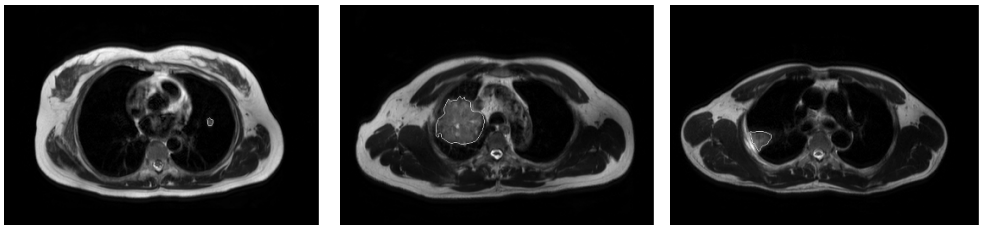

FIGURE 1. Example of lung nodules in T2W-MR. (a) is an isolated lung nodule. (b) is a juxta vascular nodule. (c) is a juxta pleural nodule.

In RPN, the convolution layers of a pre-trained network are first followed by a 3×3 convolutional layer. Then two 1×1 convolutional layers are added for classification and regression, respectively. To deal with different scales and aspect ratios of objects, anchors are introduced in the RPN. Each anchor is associated with a scale and an aspect ratio. The default setting of anchors are 3 scales (128×128, 256×256 and 512×512 pixels) and 3 aspect ratios (1: 1, 1: 2, and 2: 1), leading to $k = 9$ anchors at each location. From Fig.1, a nodule generally occupies a small portion in the image. Thus the anchor scale should be adjusted.

2) FEATURE EXTRACTION MODELS

Different feature extraction models can be used in Faster R-CNN. In this paper, three of the most impressive CNN models are tested. One is VGG16 [23]. The other two are residual network (ResNet) [24] with different architectures. ResNet uses a residual learning framework to ease the training of deep networks. It takes a standard deep CNN and add shortcut connections which bypass few convolutional layers at a time.

VGG16: VGG16 is composed of 13 convolutional layers and 4 pooling layers. For each convolutional layer, the kernel size is 3×3 and the stride is 1. For each pooling layer, the kernel size is 2×2 and the stride is 2.

ResNet50: ResNet50 is one of the residual models. The architecture of ResNet50 is shown in Table 1. When ResNet50 is used as the feature extraction model in Faster R-CNN, the output of Conv4_x is the feature map.

ResNet101: ResNet101 is another architecture of ResNet, shown in Table 1. It is the same with ResNet50, except the Conv4_x layer. We also use the output of Conv4_x as the feature map in Faster R-CNN.

3) DETECTION USING FASTER R-CNN

Directly using MR images to train the parameters of feature extraction model will get overfitting. To solve this problem, transfer learning is employed. First a pre-trained model from natural images is employed as the initial parameters. Then fine-tuning is performed to get the final parameters value.

Natural image is a three-channel image, including R, G and B channels. A T2W-MR slice is one channel gray level image. To fine-tune the existing trained models, we should construct a three-channel input for each MR slice. One construction manner is using the gray level image in each channel and the three channels are the same. In this construction, each slice in one MR scan is separately treated and the spatial information is discarded. When spatial information is discarded, air artifacts may be easily detected as lung nodule, shown in Fig. 2(a) and 2(d).

Compared with lung nodule, air artifacts show different characteristics. In one situation, air artifact can only be seen in one slice. The consecutive slices for image in Fig. 2(a) are shown in Fig. 2(b) and (c). It can be seen that air artifact can only be seen in Fig. 2(a). In the other situation, air artifacts in consecutive slices show different shapes at the same position, shown in Fig. 2(e) and (f). Based on the difference between air artifact and nodule, a three-channel input adding spatial information is designed. In the spatial three-channel input, one channel is the gray level image. The other two channels are its two consecutive slices. For the top or the bottom slice, the two consecutive slices are the following two slices or the upper two slices. For the other slices, the two consecutive slices are one following slice and one upper slice. Thus spatial information is added in the input image.

TABLE 1. The architectures of ResNet50 and ResNet 101.

Layer name	ResNet50	ResNet101
Conv1	7×7,64, stride 2	
	3×3 max pool, stride 2	
Conv2_x	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$
Conv3_x	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$
Conv4_x	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times23$
Conv5_x	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$
	average pool, 1000-d fc, softmax	

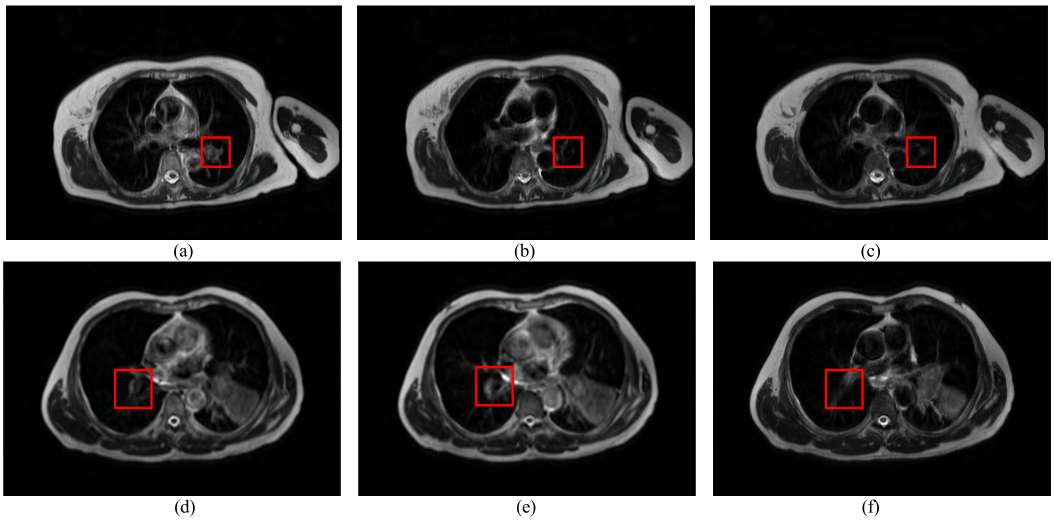

FIGURE 2. Examples of air artifacts. For (a) and (d), region in the red rectangle is the air artifact region. (b) and (c) are two consecutive slices for (a). For air artifact in (a), (b) and (c) do not display air artifact region indicated by the red rectangle. (e) and (f) are two consecutive slices for (d). Air artifact in (d), (e) and (f) display different shapes indicated by the red rectangle.

When training the RPN, two labels should be assigned to each anchor. The positive label denotes the nodule region. The negative label denotes the normal region. Anchor that has no overlap with any ground-truth box is assigned to negative label. Two kinds of anchors are assigned to positive label.

1. The anchor with the highest Intersection-over-Union (IoU) overlap with a ground truth box.

2. The anchor that has an IoU overlap higher than 0.6 with any ground-truth box.

The loss function of RPN for an image consists of two parts: classification loss L_{cls} and regression loss L_{reg}. It is defined as

$$L(\{p_i\},\{t_i\}) = \frac{1}{N_{\text{cls}}}\sum_i L_{\text{cls}}(p_i,p_i^*)+\lambda\frac{1}{N_{\text{reg}}}\sum_i p_i^* L_{\text{reg}}(t_i,t_i^*) \tag{1}$$

In (1), i is the index of the anchor. p_i is the predicted probability of anchor i being an object. p_i^* is the ground truth label for anchor i. If anchor i is positive, $p_i^* = 1$. If anchor i is negative, $p_i^* = 0$. t_i is the predicted bounding box, represented by $[t_{\text{x}}, t_{\text{y}}, t_{\text{w}}, t_{\text{h}}]$. t_i^* is the ground truth bounding box associated with a positive anchor, represented by $[t_{\text{x}}^*, t_{\text{y}}^*, t_{\text{w}}^*, t_{\text{h}}^*]$.

L_{cls} is the loss function of classification for two classes, defined as:

$$L_{\text{cls}}(p_i,p_i^*) = -\log[p_i p_i^* + (1-p_i^*)(1-p_i)] \tag{2}$$

L_{reg} is the loss function for regression, defined as:

$$L_{\text{reg}}(t_i,t_i^*) = \sum_i \text{smooth}_{L1}(t_i - t_i^*) \tag{3}$$

smooth $_{\text{L1}}$ is defined in (4).

$$\text{smooth}_{L1}(x) = \begin{cases} 0.5x^2, & \text{if } |x| < 1; \\ |x| - 0.5, & \text{otherwise.} \end{cases} \tag{4}$$

The classification loss and the regression loss are normalized by N_{cls} and N_{reg} and weighted by parameter λ.

The four coordinates for t_i and t_i^* are as follows:

$$\begin{aligned} t_{\text{x}} &= \frac{x - x_{\text{a}}}{w_{\text{a}}}, \quad t_{\text{y}} = \frac{y - y_{\text{a}}}{h_{\text{a}}}, \\ t_{\text{w}} &= \log(w/w_{\text{a}}), \quad t_{\text{h}} = \log(h/h_{\text{a}}), \\ t_{\text{x}}^* &= \frac{x^* - x_{\text{a}}}{w_{\text{a}}}, \quad t_{\text{y}} = \frac{y^* - y_{\text{a}}}{h_{\text{a}}}, \\ t_{\text{w}}^* &= \log(w^*/w_{\text{a}}), \quad t_{\text{h}}^* = \log(h^*/h_{\text{a}}). \end{aligned} \tag{5}$$

In (5), x, y, w, and h denote the box's center coordinates and its width and height. x, x_{a}, and x^* are for the predicted box, anchor box, and ground-truth box respectively (likewise for y; w; h). The output of Faster R-CNN is the predicted nodule position and the probability of being a nodule.

B. FP REDUCTION

Though spatial three-channel input is constructed for Faster R-CNN, many false positives (FPs) still exist. One type of FP is the tissue outside the lung parenchyma, indicated by the rectangle regions in Fig. 3(a). The other type of FP is the air artifact region inside the lung parenchyma, indicated by the rectangle region in Fig. 3(b). An FP reduction scheme based on the anatomical characters of lung nodule is proposed.

Lung nodules has two common characteristics:

1. Larger lung nodules can be seen in consecutive slices.

2. Lung nodules are inside the lung parenchyma.

For the first characteristic, two elements are important. The first element is deciding each region $\boldsymbol{R}_j$ belong to large or small. The second element is how to judge one large region is consecutive or not. For the first element, a fix threshold T_{s} is used to justify $\boldsymbol{R}_j$ belong to large or small. For the second element, the area overlap ratios between $\boldsymbol{R}_j$ and its corresponding regions in the two consecutive slices are used. If these two

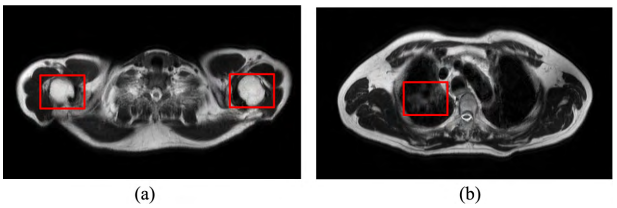

FIGURE 3. Examples of FP regions detected by Faster R-CNN.

area overlap ratios are both less than T_r, then $\boldsymbol{R}_j$ is regarded as large isolated region and should be removed. Thus the first criterion can be defined below.

Criterion One: If a detected region with an area larger than T_s does not show larger area overlap ratios with its two consecutive slices, this region should be removed.

For the second characteristic, the criterion to remove FP is to justify if the detected region is inside the lung parenchyma or not. However for lung with a juxta heart nodule or a juxta pleural nodule, the lung parenchyma may not be segmented as a whole. In this situation, the nodule is not inside the lung parenchyma and will be removed. For a balance, the second criterion is defined below.

Criterion Two: If a detected region displays in a slice without lung parenchyma, this region should be removed.

1) LUNG PARENCHYMA SEGMENTATION

For each MR slice, testing lung parenchyma exist or not is enough. Thus a rough lung parenchyma segmentation method is implemented. First the thorax is segmented. Then the lung parenchyma candidate is extracted in each slice. Finally, the non-lung parenchyma tissues are removed by 3D connected component labeling.

The region outside the thorax usually has very low intensity, which makes thresholding effective to segment the thorax. 3D Otsu threshold and shape analysis are combined to get the thorax region [25]. First, 3D Otsu is used to find the optimal threshold for an MR scan. Then, the 2D hole-filling operator is applied slice by slice. Finally, slice with incorrect segmentation is found based on Hausdorff distance and re-segmented using iterative thresholding. For the MR slices in Fig. 4(a), the thorax segmentation results are indicated as red boundary shown in Fig. 4(b).

To get the lung parenchyma candidate, each MR slice is segmented by a low threshold T. The segmented result is denoted as $\boldsymbol{BW}$ and expressed by (6).

$$BW_{i,j} = \begin{cases} 1, & \text{if } I_{i,j} < T. \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

In (6), $\boldsymbol{I}_{i,j}$ is the gray value of pixel (i, j). T is the threshold. Based on the gray value of lung parenchyma, a suitable value for T is 10. To remove the influence of the background, $\boldsymbol{BW}$ is multiplied with the thorax segmentation result and $\boldsymbol{BC}$ is obtained. The lung parenchyma candidate results are shown in Fig. 4(c). In the parenchyma candidate, some non-lung tissues (indexed by the arrow in Fig. 4(c)) still remain.

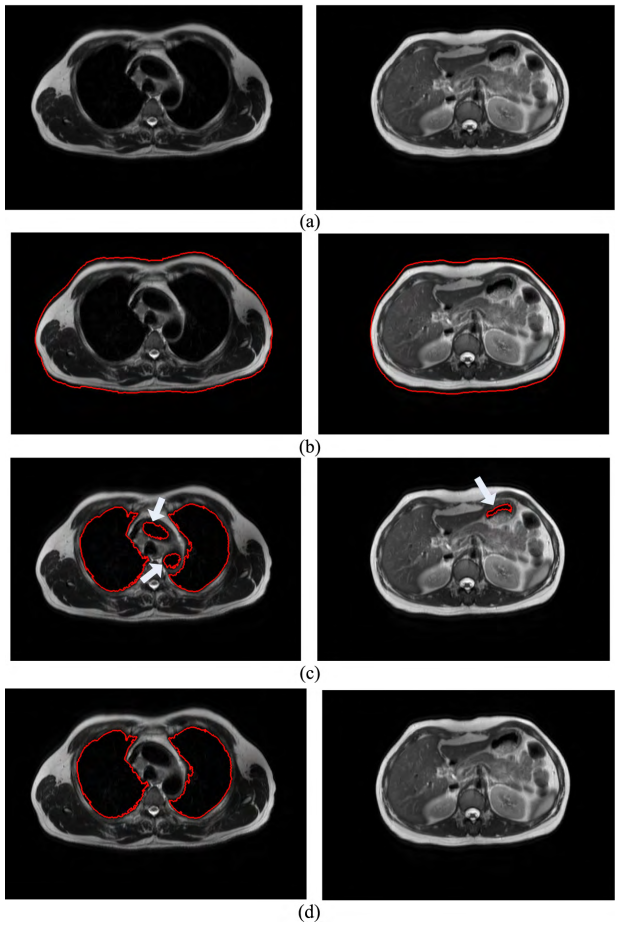

FIGURE 4. Lung parenchyma segmentation. (a) is the MR slice. (b) is the thorax segmentation result. (c) is the lung parenchyma candidate result. (d) is the final lung parenchyma result.

We employ 3D connected component analysis to remove these tissues. Treating the MR scan as a whole, 3D connected component labeling is applied. Then the lung volume V, is selected from the labeled volumes depending on the volume size, expressed by (7).

$$V = \begin{cases} V_1 \cup V_2, & \text{if } V_2 > 0.5V_1; \\ V_1, & \text{otherwise;}. \end{cases} \tag{7}$$

In (7), V_1 and V_2 are the largest and second largest volumes among the labeled volumes.

As testing lung parenchyma exist or not in each slice is enough, V is used as the final segmentation result. Fig. 4(d) shows the parenchyma segmentation for two slices. In can be seen that non-lung parenchyma tissues are removed.

2) FP REDUCTION PROCEDURE

The whole FP reduction scheme is detailed as follows.

Step 1: If no lung parenchyma exist in slice i, then the detected regions in slice i are regarded as FPs and removed. If slice i contains lung parenchyma, then go to Step 2. This procedure can remove FPs in Fig. 3(a).

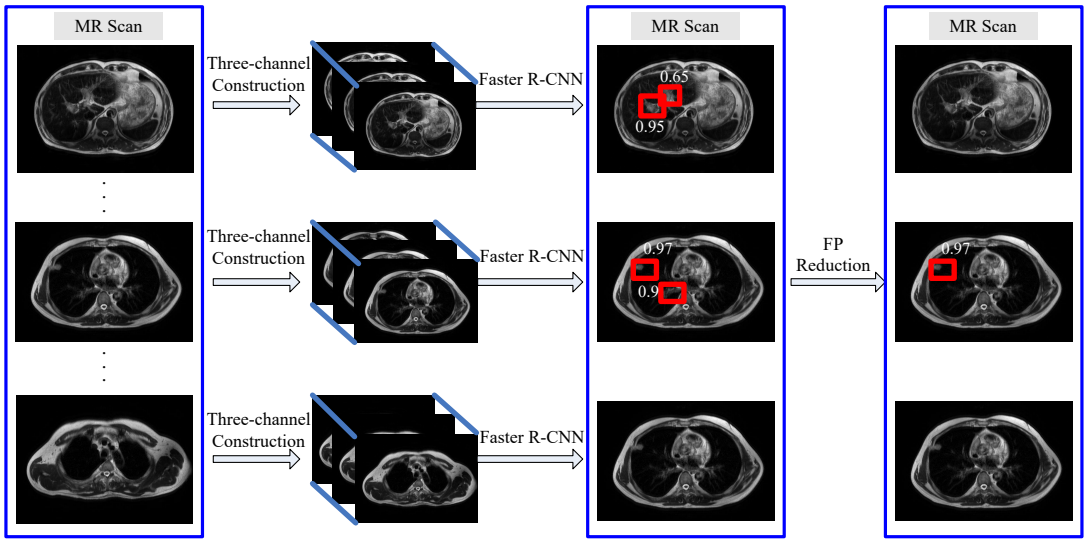

FIGURE 5. The whole nodule detection procedure for an MR scan.

Step 2: For each region $\boldsymbol{R}_j$ in slice i, if its area is less than T_s, $\boldsymbol{R}_j$ will be reserved. If its area is larger than T_s, then go to Step 3. In this experiment, $T_s = 500$.

Step 3: Get the detection result in slice $i-1$ and slice $i+1$, $\boldsymbol{S}_{i-1}$ and $\boldsymbol{S}_{i+1}$. Compute the area overlap ratios between $\boldsymbol{R}_j$ and $\boldsymbol{S}_{i-1}$, and $\boldsymbol{R}_j$ and $\boldsymbol{S}_{i+1}$. If these two area overlap ratios are both less than T_r, then $\boldsymbol{R}_j$ is regarded as FP and removed. In this experiment, $T_r = 0.2$. This procedure can remove FPs in Fig. 3(b).

An example of the whole nodule detection procedure for an MR scan is shown in Fig. 5.

IV. EXPERIMENTAL SETUP

For the 142 T2W-MR scans, 97 scans are randomly selected as training set and 45 scans are used as testing set. For the training set, the total number of slices is 2332. The number of slices with at least one nodule is 549 and the number of nodules is 604. When training the Faster R-CNN, just the slices with lung nodule are used. Thus the image number for training Faster R-CNN is 549. In this experiment, a transfer learning method is used to train the network. For the testing set, the total number of slices is 1071. The number of slices with at least one nodule is 251 and the number of nodules is 258. All the testing images are used to test the performance.

The typical anchor sizes for Faster R-CNN are 128×128, 256×256 and 512×512. Lung nodule occupies a small part in one slice image. Thus the typical anchor sizes are not suitable for lung nodule detection. In this experiment, the nodule size in the training images are used to optimize the anchor sizes. Based on analyzing the longer side length of nodules in the training images, the optimized anchor size are set as 32×32, 64×64 and 128×128 in order to cover all nodules. The typical number of anchors for computing the loss function of a mini-batch is 256. As most images contain one nodule, the number of positive anchors is small for each image. Thus the number of anchors of a mini-batch is adjusted to 128. Other parameters are set as follows: batch size (1), number of training iterations (12000), learning rate (10e-3), weight decay (10e-4) and momentum (0.9).

To evaluate the performance of nodule detection, free receiver operating characteristic (FROC) curve is used. The FROC curve shows the relationship between the true positive rate (TPR) and the false positives per images (FPs/I) as the decision threshold varies.

V. RESULTS AND DISCUSSION

A. PERFORMANCE OF FEATURE EXTRACTION MODEL

Different feature extraction models can be used in Faster R-CNN. In order to test which model is more suitable, VGG16 [23], ResNet50 [24] and ResNet101 [24] are compared. We divide the training set into two parts. One part is used for training Faster R-CNN and the scan size is 68. The other part is used to get the validation result and the scan size is 29. The FROC curves for different feature extraction models under the validation dataset are shown in Fig. 6. It can be seen that ResNet101 module gets the best performance compared with the other two modules. ResNet101 is the

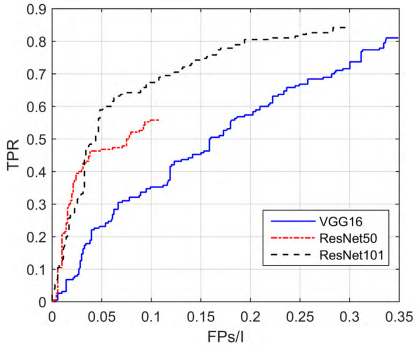

FIGURE 6. FROC curves for different feature extraction models.

deepest model among the three feature extraction models. A deeper feature extraction model can get deeper features and has larger receptive field. Larger receptive field may lead to better detection performance.

B. PERFORMANCE OF LUNG NODULE DETECTION

1) ADDING SPATIAL INFORMATION

To eliminate the influence of air artifacts, the spatial three-channel input is designed for Faster R-CNN. To validate the effectiveness of spatial three-channel, another input using gray level image in each of the three channels is compared. This three-channel input is called gray three-channel. For these two inputs, separate Faster R-CNN networks under ResNet101 are trained using the whole training set. The image-based FROC curves are drawn in Fig. 7. It can be seen that spatial three-channel designed in this paper gets better lung nodule detection performance compared with gray three-channel. At 75%, 80% and 85% sensitivity, FPs/I for spatial three-channel are 0.17, 0.19 and 0.28. For gray three-channel, the corresponding FPs/I are 0.19, 0.23 and 0.31. The performance difference between the two inputs is statistically significant ($p = 0.017 < 0.05$) by t-test analysis.

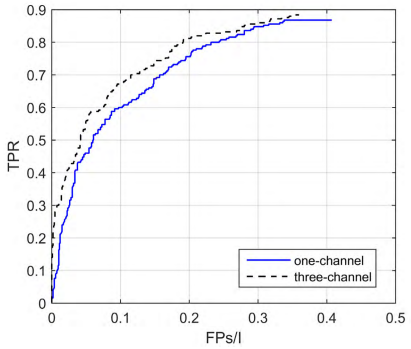

FIGURE 7. FROC curves for different input constructions.

Some detection examples between these two inputs are shown in Fig. 8. In Fig. 8(b), Faster R-CNN with gray three-channel detect both the lung nodule and an FP region. When using Faster R-CNN with spatial three-channel, this FP region is not detected and the true nodule is preserved, shown in Fig. 8(c). In Fig. 8(e), the organ tissue is detected as lung nodule using Faster R-CNN with gray three-channel. For Faster R-CNN with spatial three-channel, this FP region is not detected, shown in Fig. 8(f). It may be concluded that adding spatial information in Faster R-CNN input can remove some FP regions without eliminating true nodules. As each MR scan is 7.7mm spatial resolution along the axial direction, spatial three-channel can not increase small nodule detection rate.

2) FP REDUCTION COMPARISON

Secondly, the performance of the proposed FP reduction scheme is tested. Two criteria are proposed to remove FPs. Let C1 represent Criterion one and C2 represent Criterion two. The image-based FROC curves for Faster R-CNN, Faster R-CNN with C1 and Faster R-CNN with C1+C2 are drawn in Fig. 9. It can be seen that the designed FP reduction scheme gets better lung nodule detection performance compared with Faster R-CNN. Using t-test analysis, the performance difference between Faster R-CNN with C1+C2 and Faster R-CNN is statistically significant ($p < 0.05$).

We also test the performance difference between C1 and C1+C2. From Fig. 9, C1+C2 shows better detection result and the improvement is statistically significant ($p < 0.05$). It may be concluded that different FP regions are separately removed by C1 and C2 criterion.

What's more, FP reduction methods based on handcrafted features are implemented for a comparison. Three popular texture features: anchor graph hashing (AGH), local ternary patterns (LTP) and histogram of gradient (HOG) are selected. FP reduction methods based on handcrafted features are as follows.

Step 1: For the training dataset, two sets containing both normal and nodule regions are cropped. One set of normal and nodule regions is with 64×64 size and denoted as $\boldsymbol{D}_1$. The other set of normal and nodule regions is with 128×128 size and denoted as $\boldsymbol{D}_2$.

Step 2: For each regions in $\boldsymbol{D}_1$, use AGH (or LTP or HOG) to extract features. All the features compose $\boldsymbol{F}_1$. For each regions in $\boldsymbol{D}_2$, use AGH (or LTP or HOG) to extract features. All the features compose$\boldsymbol{F}_2$.

Step 3: For the detected regions by Faster R-CNN, if the region size is smaller than 64×64, a 64×64 region is cropped based on the detected centroid. Otherwise, a 128×128 region is cropped based on the detected centroid. The cropped region is denoted as $\boldsymbol{R}_\mathrm{q}$.

Step 4: For $\boldsymbol{R}_\mathrm{q}$, use AGH (or LTP or HOG) to extract features and $\boldsymbol{f}_\mathrm{q}$ is obtained.

Step 5: If $\boldsymbol{R}_\mathrm{q}$ is with 64×64 size, compute the L2 distance between $\boldsymbol{f}_\mathrm{q}$ and each element in $\boldsymbol{F}_1$. Otherwise, compute the L2 distance between $\boldsymbol{f}_\mathrm{q}$ and each element in $\boldsymbol{F}_2$.

Step 6: Regions with the *Num* minimum distances are similar regions with $\boldsymbol{R}_\mathrm{q}$. If there are less than $0.3Num$ nodule regions in the similar regions, $\boldsymbol{R}_\mathrm{q}$ is regarded as normal and will be removed. In this experiment, $Num = 25$.

The image-based FROC curves for the proposed FP reduction schemes and FP reduction methods based on AGH, LTP and HOG are drawn in Fig. 10. It can be seen that the proposed scheme is better in FP reduction compared with handcrafted features. And the performance differences are statistically significant ($p < 0.05$). For handcrafted features, the discriminating performance orders are LTP, HOG and AGH. However, their performance differences are not statistically significant ($p > 0.05$). From the results, we can conclude that using handcrafted features to remove normal regions detected by Faster R-CNN may be not suitable. The reason may be that the normal regions detected by Faster R-CNN show similar image characteristics with nodule regions. The handcrafted features are extracted from the image gray value, which has already involved in Faster

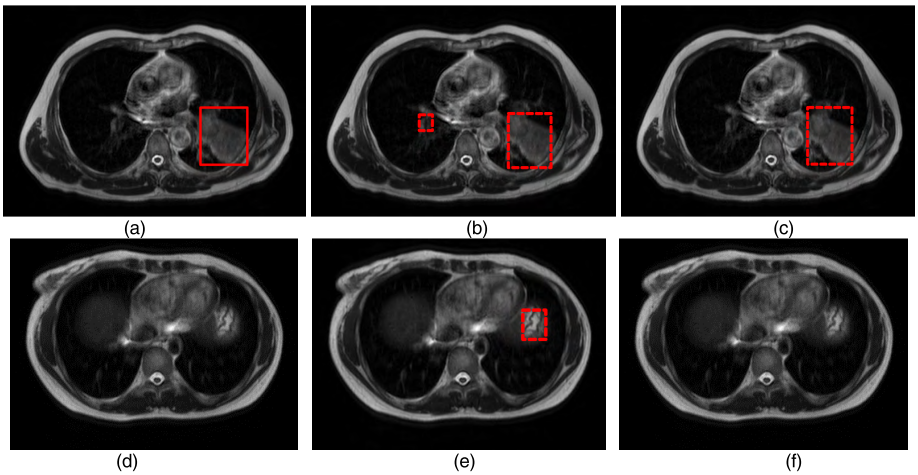

FIGURE 8. Nodule detection results comparison between different input constructions. (a) and (d) are the ground truth images. (b) and (e) are the detection results with gray three-channel input. (c) and (f) are the detection results with spatial three-channel input. Regions in the rectangle are the true nodule regions. Regions in the dashed rectangle are the detected nodule regions.

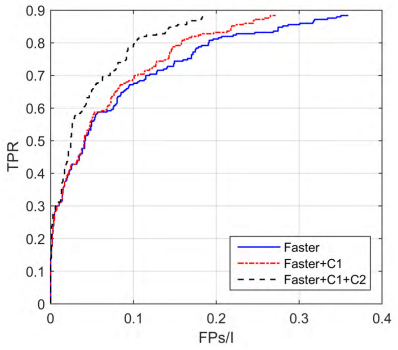

FIGURE 9. FROC curves for FP reduction schemes.

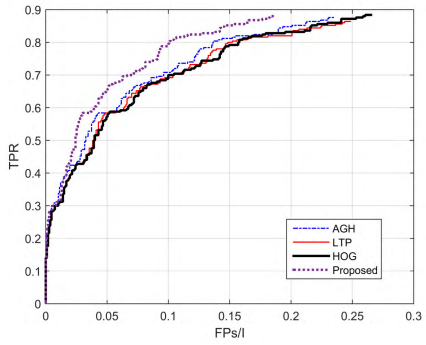

FIGURE 10. FROC curves between the proposed FP reduction scheme and schemes based on handcrafted features.

R-CNN. Thus no obvious improvement is obtained. The proposed FP reduction scheme employ anatomical characteristics, which is not used in Faster R-CNN. Thus the improvement is statistically significant.

3) COMPARISON WITH EXISTING METHODS

Finally, a comparison between the proposed method and existing nodule detection methods is shown. As no methods have been carried out on MR images, some handcrafted feature based methods [8], [14] and deep learning based methods [20], [21] for CT or chest radiographs (CXRs) nodule detection are selected. As the image characteristics of CT, CXRs and T2W-MR are different, the comparison is implemented from two patterns.

The first pattern is implementing existing methods and comparing the results on our MRI dataset. For existing methods, nodule candidate was first detected then features were extracted to classify the candidate as normal or nodule. As candidate boundary should be segmented for feature extraction in [8] and [14], two texture features (LTP and HOG) are also tested for lung nodule detection. The comparison results are shown in Table 2. From Table 2, the proposed method gets better detection results compared with existing nodule detection methods on T2W-MR modality.

TABLE 2. Comparison with existing methods on nodule detection under our dataset.

	Sensitivity=80%, FPs/scan	Sensitivity=85%, FPs/scan
LTP	14.7	17.4
HOG	16.3	19.8
Sousa [8]	17.8	21.7
Javaid [14]	18.7	20.7
Li [20]	22.4	25.2
Li [21]	14.8	19.9
Proposed	**2.38**	**3.47**

The second pattern is comparing the results of different methods on their own dataset. The comparison results are shown in Table 3. From Table 3, nodule detection for T2W-MR modality is slightly lower than the best performance using CT modality. The reason may be the lower resolution of T2W-MR. From the difference between Table 2 and Table 3, it may be concluded that candidate detection and feature extraction methods for CT (or CXRs) images are

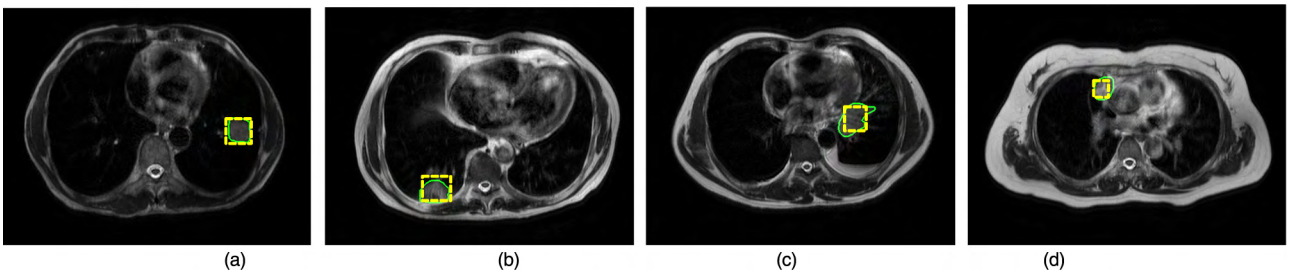

FIGURE 11. Examples of Lung nodule detection. Ground truth is indicated by the green boundary. Region in the yellow rectangle is the detected result.

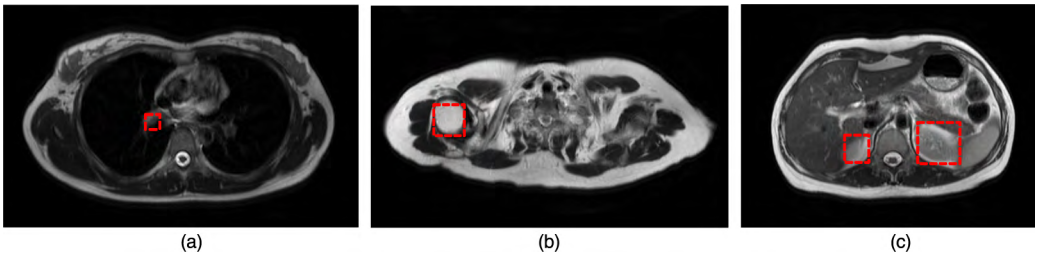

FIGURE 12. Examples of the detected FP. Region in the red rectangle is the detected FP region.

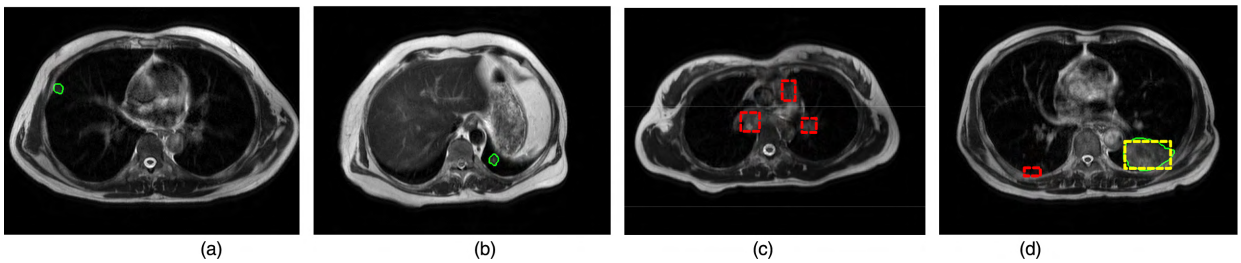

FIGURE 13. Examples of incorrect detection. Ground truth is indicated by the green boundary. Region in the dashed rectangle is the detected result. In (a) and (b), true nodule is not detected. In (c) and (d), region in red rectangle is the FP region. Region in yellow rectangle is the true positive region.

TABLE 3. Comparison with existing methods on nodule detection under different datasets.

	Dataset	Sensitivity (%)	FPs/scan
Santos [8]	CT	80.5	1.17
Javaid. [14]	CT	91.65	3.19
Li [20]	CT	87.1	4.622
Li [21]	XCRs	84	2.0
Proposed	MRI	85.2	3.47

not suitable for MR images. The reason may be the image differences between different modalities.

C. DISCUSSION

We present a novel deep learning based method for lung nodule detection in 3D thoracic T2W-MR images. To the best of our knowledge, it is the first attempt to detect lung nodules in thoracic MR images. Different from nodule detection methods in CT, the proposed method takes the whole image as input and no candidate extraction is needed. Moreover, it can detect nodules with different sizes and types. For isolated nodule, juxta heart nodule and juxta pleural nodule, the proposed method can get correct detection results, shown in Fig. 11(a)-(c). Besides, some challenging nodules can also be detected. For nodule connected with the organ tissue shown in Fig. 11(d), the designed Faster R-CNN method can get correct detection result.

Though Faster R-CNN can detect most of the lung nodules, many FP regions are also detected, shown in Fig. 12. To reduce FPs and preserve true nodule, an FP reduction scheme using the anatomical characters of lung nodule is designed. For air artifacts displaying in less than two slices (shown in Fig. 12(a)) and regions detected in a slice without lung parenchyma (shown in Fig. 12(b) and (c)), it can be removed by the FP reduction scheme.

Some problems also exist in the detection result. Firstly, some small and low contrast nodules are not detected by Faster R-CNN. For nodules indicated by the green boundary in Fig. 13(a) and (b), the proposed method misses these nodules. Second some air artifacts and juxta heart tissues

may be falsely detected as nodules. For regions in the red rectangle shown in Fig. 13(c) and (d), the proposed method falsely detect them as nodules. To alleviate these problems, one manner is to design a filter method to improve the image quality and remove air artifacts. The second manner is adding multi-scale in the whole detection system to increase the detection rate of small and low contrast nodules.

VI. CONCLUSION

In this paper, a lung nodule detection method for thoracic MR images is proposed based on deep learning. With optimized parameter, spatial three-channel input and transfer learning, Faster R-CNN is designed for lung nodule detection. This detection scheme can avoid candidate extraction and be less dependent on scale. As Faster R-CNN does not consider anatomical characteristics, many FP regions exist in the detection results. To reduce FPs and preserve true nodule, an FP reduction scheme based on the anatomical characteristics of lung nodule is designed. Experimental results show that the designed Faster R-CNN can detect most of the nodules and the proposed FP reduction scheme can obviously reduce FP regions. Two observations can be concluded. First, adding spatial information in input image can remove some FP regions without eliminating true nodules for 3D medical images. Second, FP reduction scheme based on anatomical characteristics may be better than image features in Faster R-CNN detection.

REFERENCES

[1] B. W. Stewart and C. P. Wild, *World Cancer Report 2014. International Agency for Research on Cancer*. Geneva, Switzerland: World Health Organization, vol. 505, 2014.

[2] *Cancer Facts & Figures*, Amer. Cancer Soc., Atlanta, GA, USA, 2012.

[3] 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.

[4] J. Biederer *et al.*, "MRI of the lung (3/3)–current applications and future perspectives," *Insights Imag.*, vol. 3, no. 4, pp. 373–386, 2012.

[5] H. Koyama *et al.*, "Quantitative and qualitative assessment of non-contrast-enhanced pulmonary MR imaging for management of pulmonary nodules in 161 subjects," *Eur. Radiol.*, vol. 18, pp. 2120–2131, Oct. 2008.

[6] J. C. Sieren, Y. Ohno, H. Koyama, K. Sugimura, and G. McLennan, "Recent technological and application developments in computed tomography and magnetic resonance imaging for improved pulmonary nodule detection and lung cancer staging," *J. Magn. Reson. Imag.*, vol. 32, no. 6, pp. 1353–1369, 2010.

[7] T. Messay, R. C. Hardie, and S. K. Rogers, "A new computationally efficient CAD system for pulmonary nodule detection in CT imagery," *Med. Image Anal.*, vol. 14, no. 3, pp. 390–406, 2010.

[8] I. R. S. Valente, P. C. Cortez, E. C. Neto, J. M. Soares, V. H. C. de Albuquerque, and J. M. R. S. Tavaresd, "Automatic 3D pulmonary nodule detection in CT images: A survey," *Comput. Methods Programs Biomed.*, vol. 124, pp. 91–107, Feb. 2016.

[9] J. R. F. da Silva Sousa, A. C. Silva, A. C. de Paiva, and R. A. Nunes, "Methodology for automatic detection of lung nodules in computerized tomography images," *Comput. Methods Programs Biomed.*, vol. 98, no. 1, pp. 1–14, 2010.

[10] D. Cascio, R. Magro, F. Fauci, M. Iacomi, and G. Raso, "Automatic detection of lung nodules in CT datasets based on stable 3D mass–spring models," *Comput. Biol. Med.*, vol. 42, no. 11, pp. 1098–1109, 2012.

[11] W.-J. Choi and T.-S. Choi, "Automated pulmonary nodule detection based on three-dimensional shape-based feature descriptor," *Comput. Methods Programs Biomed.*, vol. 113, no. 1, pp. 37–54, 2014.

[12] A. M. Santos, A. O. de Carvalho Filho, A. C. Silva, A. C. de Paiva, R. A. Nunes, and M. Gattass, "Automatic detection of small lung nodules in 3D CT data using Gaussian mixture models, Tsallis entropy and SVM," *Eng. Appl. Artif. Intell.*, vol. 36, pp. 27–39, Nov. 2014.

[13] A. O. de Carvalh Filho, W. B. de Sampaio, A. C. Silva, A. C. de Paiva, R. A. Nunes, and M. Gattass, "Automatic detection of solitary lung nodules using quality threshold clustering, genetic algorithm and diversity index," *Artif. Intell. Med.*, vol. 60, no. 3, pp. 165–177, 2014.

[14] M. Javaid, M. Javid, M. Z. Rehman, and S. I. Shah, "A novel approach to CAD system for the detection of lung nodules in CT images," *Comput. Methods Programs Biomed.*, vol. 135, pp. 125–139, Oct. 2016.

[15] H. Jiang, H. Ma, W. Qian, M. Gao, and Y. Li, "An automatic detection system of lung nodule based on multigroup patch-based deep learning network," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 4, pp. 1227–1237, Jul. 2018.

[16] N. Tajbakhsh and K. Suzuki, "Comparing two classes of end-to-end machine-learning models in lung nodule detection and classification: MTANNs vs. CNNs," *Pattern Recognit.*, vol. 63, pp. 476–486, Mar. 2017.

[17] R. Golan, C. Jacob, and J. Denzinger, "Lung nodule detection in CT images using deep convolutional neural networks," in *Proc. Int. Joint Conf. Neural Netw.*, Jul. 2016, pp. 243–250.

[18] X. Huang, J. Shan, and V. Vaidya, "Lung nodule detection in CT using 3D convolutional neural networks," in *Proc. IEEE 14th Int. Symp. Biomed. Imag.*, Apr. 2017, pp. 379–383,.

[19] Q. Dou, H. Chen, L. Yu, Jing 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.

[20] W. Li, P. Cao, D. Zhao, and J. Wang, "Pulmonary nodule classification with deep convolutional neural networks on computed tomography images," *Comput. Math. Methods Med.*, vol. 2016, Dec. 2016, Art. no. 6215085.

[21] C. Li, G. Zhu, X. Wu, and Y. Wang, "False-positive reduction on lung nodules detection in chest radiographs by ensemble of convolutional neural networks," *IEEE Access*, vol. 6, pp. 16060–16067, Mar. 2018.

[22] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

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

[24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

[25] Y. Li, H. Chen, and X. Wei, "An automated lung parenchyma segmentation method for 3D thoracic MR images," in *Proc. Int. Conf. Softw. Eng. Inf. Syst.*, 2015, pp. 330–336.

YANFENG LI was born in Hebei, China, in 1988. She received the B.S. degree in communication engineering and the Ph.D. degree in circuit and system from Beijing Jiaotong University, Beijing, China, in 2010 and 2015, respectively.

Since 2015, she has been an Associate Professor with the School of Electronic and Information Engineering, Beijing Jiaotong University. She is the author of two books and more than 20 articles. Her research interests include image processing, pattern recognition, medical image analysis, and video analysis and understanding.

LINLIN ZHANG was born in Shandong, China, in 1994. He received the B.S. degree in biomedical engineering from the Shandong University of Science and Technology, Qingdao, China, in 2017. He is currently pursuing the degree with the School of Electronic and Information Engineering, Beijing Jiaotong University.

His research interests include pattern recognition and medical image analysis.

HOUJIN CHEN was born in Anhui, China, in 1965. He received the B.S. degree in signal engineering from Lanzhou Jiaotong University, Lanzhou, China, in 1986, and the M.S. and Ph.D. degrees in communication and control from Beijing Jiaotong University, Beijing, China, in 1989 and 2003, respectively.

From 1991 to 1995 and from 1995 to 2000, he was a Lecturer and an Associate Professor, respectively, with the School of Electronic and Information Engineering, Beijing Jiaotong University, where he has been a Professor, since 2000. He has authored three books and more than 100 articles. His research interests include image processing, pattern recognition, and medical image analysis.

NA YANG was born in Shanxi, China, in 1977. She received the B.S. and M.S. degrees in electronic information engineering from the North University of China, Taiyuan, in 2005, and the Ph.D. degree in traffic information engineering and control from Beijing Jiaotong University, Beijing, China, in 2013.

She held a postdoctoral position with the Institute of Computing Technology, Chinese Academy of Sciences, in 2013. Since 2005, she has been an Associate Professor with the School of Information and Communication Engineering, North University of China. Her research interests include image classification and medical image analysis.