IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 22, 2019, accepted March 15, 2019, date of publication March 19, 2019, date of current version June 5, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2906116

Multi-Branch Ensemble Learning Architecture Based on 3D CNN for False Positive Reduction in Lung Nodule Detection

HAICHAO CAO[1], HONG LIU[1], ENMIN SONG[1], (Senior Member, IEEE), GUANGZHI MA[1], XIANGYANG XU[1], RENCHAO JIN[1], TENGYING LIU[1], AND CHIH-CHENG HUNG[2]

[1]School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China
[2]Laboratory for Machine Vision and Security Research, Kennesaw State University, Kennesaw, GA 30144, USA

Corresponding author: Hong Liu (hl.cbib@gmail.com)

This work was supported in part by the National Key R&D Program of China under Grant 2017YFC0112804, and in part by the National Natural Science Foundation of China under Grant 81671768.

ABSTRACT It is critical to have accurate detection of lung nodules in CT images for the early diagnosis of lung cancer. In order to achieve this, it is necessary to reduce the false positive rate of detection. Due to the heterogeneity of lung nodules and their similarity to the background, it is difficult to distinguish true lung nodules from numerous candidate nodules. In this paper, in order to solve this challenging problem, we propose a Multi-Branch Ensemble Learning architecture based on the three-dimensional (3D) convolutional neural networks (MBEL-3D-CNN). The method combines three key ideas: 1) constructing a 3D-CNN to make the maximum utilization of spatial information of lung nodules in the 3D space; 2) embedding a multi-branch network architecture in the 3D-CNN that is well adapted to the heterogeneity of lung nodules, and; 3) using ensemble learning to effectively improve the generalization performance of the 3D-CNN model. In addition, we use offline hard mining operations to make the network capable of handling those indistinguishable positive and negative samples. The proposed method was tested on the dataset LUNA16 in our experiments. The experimental results show that MBEL-3D-CNN architecture can achieve better screening results.

INDEX TERMS Computer-aided diagnosis, ensemble learning, false positive reduction, offline hard mining, 3D CNN.

I. INTRODUCTION

In order to improve the survival rate of lung cancer patients, it is essential for early detection and analysis of lung nodules in computed tomography (CT) images. According to the statistics, among the mortality rate of all cancers is 19.5% [1], [2], the incidence of lung cancer accounts for 66.67% with the 18% of five-year survival rate [3]. Because the lung cancer generally evolves from the pulmonary nodules, if suspicious nodules can be found earlier, this detection will help doctors to develop effective treatment options for patients, and minimize the risk of death from lung cancer [4], [5]. In order to detect lesions as accurately as possible, traditional manual reading methods require doctors to observe and analyze CT images with great concentration. However, this manual detection method is not only time-consuming and labor-intensive, but also easy to miss and misdiagnose, especially for small nodules less than 10mm in diameter [6], [7]. With an increasing number of CT images, in order to avoid cumbersome manual operations, it is of great clinical significance to develop a pulmonary nodule detection system for assisting doctors in diagnosis [8].

In general, the lung nodule detection consists of two steps: 1) finding suspicious candidate nodules on CT images (lung nodule detection phase), and 2) screening out true lung nodules from a large number of candidate nodules (false positive reduction phase). For an initial detection of candidate nodules, many methods have been proposed in recent years, such as the traditional methods based on threshold [9], [10], morphology [11], [12] and energy optimization [13], [14]. In addition, methods based on machine learning [15]–[19] and convolutional neural networks (CNN) [20]–[23] are widely used. In the detection phase, to ensure a high recall

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

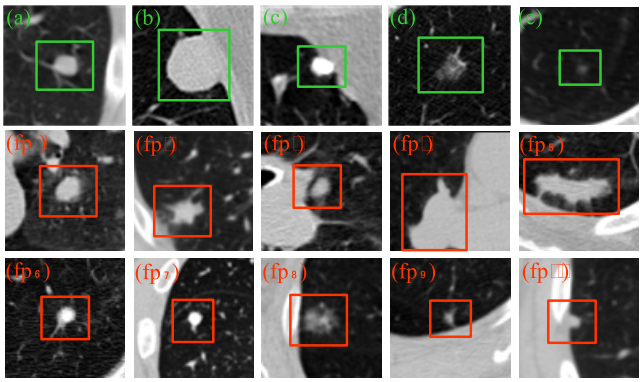


FIGURE 1. An example shows images of lung nodules with different locations and shapes in CT: (a) common isolated nodule, (b) juxtapleural nodule, (c) calcific nodule, (d) ground-glass opacity (GGO) nodule, and (e) small nodule with a diameter of 4.4 mm. Images of (fp1-10) show ten hard negative samples.

rate, the decision criteria adopted by the system are usually very simple and slack, which will introduce a large number of candidate nodules into the false positive reduction phase. Therefore, the false positive reduction phase is a particularly important component of the pulmonary nodule detection system. At present, many researchers have devoted their study of false positive reduction to ensure that the false positive rate of nodules is minimized in the case of high sensitivity detection of nodules [24].

In the process of reducing false positives, two problems are often encountered. Firstly, the heterogeneity of lung nodules in CT images (size, shape, location and texture of each nodule are different, as shown in Fig. 1), which is very unfavorable for building a robust screening model. For example, for the juxtapleural nodule (Fig. 1(b)), the calcific nodule (Fig. 1(c)), and the ground-glass opacity nodule (Fig. 1(d)), their intensities (features) are quite different. In addition, the intensity of the juxtapleural nodule and that of the lung wall are almost the same, but the intensity of the calcified nodule is generally higher than that of the lung wall, and the intensity of the ground-glass opacity nodule is much smaller than that of the lung wall. Secondly, due to the highly similarity of visual characteristics between the positive lung nodules and the non-positive lung samples (negative samples) in some cases, it is difficult to find a method to obtain a satisfactory screening result. Typically, for small pulmonary nodules (Fig. 1(e)), such nodules are more difficult to be distinguished due to their very similar appearance to the surrounding noise in terms of characteristics. Furthermore, in order to more intuitively show the similarity between hard negative samples and positive samples, we list ten hard negative samples in the second and third rows of Fig. 1(Fig. 1(fp1-10)).

In this study, we developed a 3D CNN-based neural network for false positive reduction. In order to solve the above two challenges, we propose a Multi-Branch Ensemble Learning architecture based on 3D Convolutional Neural Networks (MBEL-3D-CNN). The architecture consists of three modules: the multi-branch network model based on 1) 3D-VggNet [25], 2) 3D-InceptionResNet [26], and 3) 3D-DenseNet [27]. Specifically, in order to make the proposed networks adapt well to the heterogeneity of lung nodules, we have designed three different sizes of inputs for three multi-branch networks, and each branch structure in each multi-branch networks corresponding to a different size as an input. In addition, for the similarity problem encountered between lung nodules and hard negative samples, we first use the offline hard mining operation to make the model be able to handle those indistinguishable negative samples, and then further improve the generalization performance of the model using the ensemble learning. Please be noted that for the convenience of description, we refer to ''InceptionResNet'' as ''IResNet''.

Our technical contributions in this work include the following three aspects.

(1) In order to make full use of the 3D spatial information of lung nodules, we designed three 3D-CNNs based on VggNet, IResNet and DenseNet, respectively. These 3D-CNNs can extract different high-level features from CT images to characterize lung nodules.

(2) A multi-branch network architecture was designed, and in order to effectively learn the differences among various types of lung nodules, we constructed three different branch structures, each of which corresponds to a different input size.

(3) Not only does the offline hard mining operation be used to excavate hard negative samples (not belonging to the tissue or background of the lung nodules), but it is also based on this idea to find positive samples (lung nodules) that are difficult to be distinguished. In addition, we further improve the generalization ability of the model using the ensemble learning.

II. RELATED WORKS

Many researchers have devoted their effort to developing an effective and robust method for reducing false positive lung nodules. These methods can generally be divided into traditional machine learning algorithms and methods based on two-dimensional (2D) or three-dimensional (3D) CNNs.

In traditional machine learning methods, experts often design features that can characterize lung nodules for distinguishing false-positive nodules [5], [28]–[30]. For example, Jacobs et al. defined 128 nodule features, in addition to shape, intensity and texture features, introduced a new set of context features, and then used sequential forward floating selection (SFFS) procedures [31] and linear discriminant classifiers for feature selection and classification [32]. Aghabalaei *et al.* [18] designed a set of spectral, texture and shape features to characterize nodules, and then used support vector machine (SVM) classifiers to classify candidate nodules. Soudeh *et al.* first reconstructed the 3D region of each candidate nodule by using the sparse field method, and then extracted a set of 2D and 3D related features from the segmented candidate mask, and the RUSBoost algorithm is used to distinguish true lung nodules from candidate nodules [33].

In the neural network method based on 2D convolution, high-level semantic features automatically extracted by convolutional neural networks (CNN) are used to replace artificially designed features such as shape and texture features [34]–[38]. For example, in order to map 3D contextual information of lung nodules into a 2D space, Haeil *et al.* use averaging multiple slice of candidate nodules according to the Gaussian distribution to produce a 2D image patch for training [39]. To further mine more 3D information in a 2D convolution, Setio *et al.* designed the input to include not only the axial, coronal, and sagittal views of the lung nodules, but also six views of diagonals. In other words, there are nine views stacked as inputs to a 2D CNN [40]. Based on the idea of extracting depth features, there is also a method proposed by Ling *et al.* in combining the high-level semantic features and the traditional features of manual design to reduce the false positives of lung nodules [22].

Although the above 2D-CNN method has utilized the 3D spatial information of the lung nodule, it has not been fully exploited and explored. In general, the method based on 3D-CNN in the field of medical image processing research is still in its infancy. For the 3D CNN-based method for reducing the false positive lung nodules in a CT image, so far we only found two papers published in the literature. One is based on the 3D-CNN, a simple and efficient multi-level information coding strategy proposed by Dou et al. [41]. The other is a multi-view based 3D-CNN method proposed by Dobrenkii *et al.* Their method includes three steps of receptive field selection, feature extraction and advanced feature fusion and classification [42].

The proposed MBEL-3D-CNN method in this study is different from the previous methods in three aspects: 1) The method of the 3D-CNN in its infancy is taken as the basic architecture; 2) A non-shared multi-branch network architecture is proposed, which can simultaneously learn multi-view and multi-scale features of CT images; 3) Based on the idea of expert consultation, we trained three independent models based on VggNet, IResNet and DenseNet, respectively, and weighted the prediction results of these three models with an ensemble learning.

III. METHODS

The proposed theoretical methods are described in detail below. This section is divided into three subsections: Section A sketches the proposed network architecture, Section B describes the design of a multi-branch architecture, Section C describes the proposed offline hard mining, and Section D gives our training procedure.

A. NETWORK ARCHITECTURE

Our proposed 3D CNN-based ensemble learning framework consists of three network models, namely VggNet-based, IResNet-based and DenseNet-based 3D multi-branch network models. For a given candidate lung nodule, we extract three different scales of 3D data blocks containing the lung nodules as inputs to the three network models, and obtain the final prediction result by fusing the output probabilities of the three models. Fig. 2 shows the overall architectural diagram of the proposed false positive reduction method.

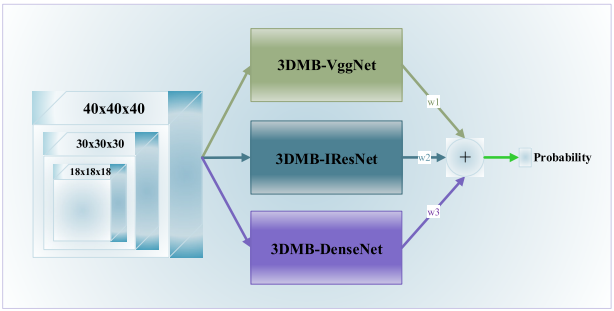


FIGURE 2. The proposed 3D CNN-based ensemble learning framework: 3DMB-VggNet, 3DMB-IResNet and 3DMB-DenseNet represent 3D multi-branch VggNet, IResNet and DenseNet, respectively. The final prediction result is a weighted average of the output probabilities of the three modules, where w1, w2, and w3 are all one-third.

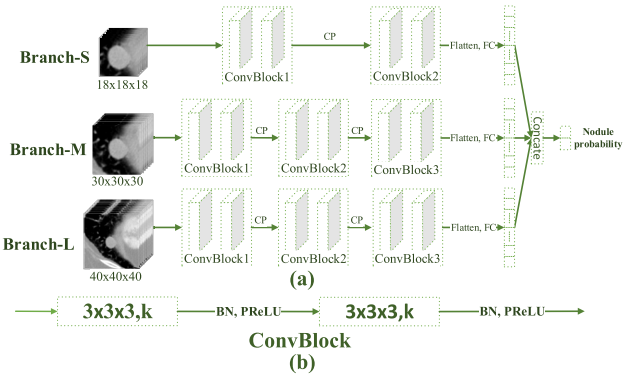


FIGURE 3. (a) The proposed 3D multi-branch VggNet (3DMB-VggNet) and (b) the diagram of 3D convolution block (ConvBlock). Parameter k indicates the number of channels. In addition, the CP and Concate represent the Central Pooling and Concatenate operation, respectively.

TABLE 1. Network parameters of the 3DMB-VggNet. Building blocks are shown in brackets with the numbers of blocks stacked. Downsampling is done using the Central Pooling. The stride size of the convolution operation is one. The symbol "*" indicates that there is no such operation.

Layer name	Branch-S	Branch-M	Branch-L
Input size	18×18×18	30×30×30	40×40×40
ConvBlock1	[3×3×3, 64]×2	[3×3×3, 64]×2	[3×3×3, 64]×2
ConvBlock2	[3×3×3, 64]×2	[3×3×3, 64]×2	[3×3×3, 64]×2
ConvBlock3	*	[3×3×3, 64]×2	[3×3×3, 64]×2
Output size	9×9×9	8×8×8	10×10×10
FC	150	200	250
Probability	Concatenate, 2-d fc, softmax		

1) 3D MULTI-BRANCH VggNet

Fig. 3 shows the architecture of the proposed 3D multi-branch VggNet (3DMB-VggNet). Table 1 lists the corresponding network parameters.

The network incorporates deep branches of three different structures, and the inputs for each branch are different.

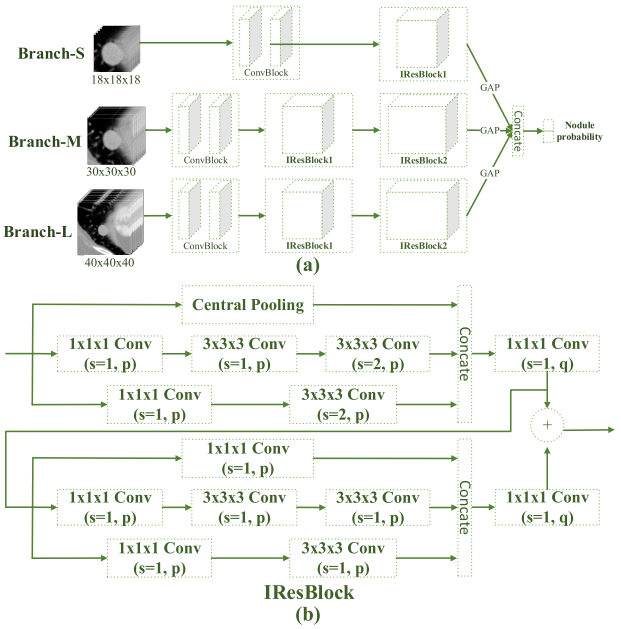


FIGURE 4. (a) The proposed 3D multi-branch InceptionResNet (3DMB-IResNet) and (b) The diagram of the 3D InceptionResNet block (IResBlock). Parameters p and q indicate the number of channels. The GAP and Concate represent the Global Average Pooling operation and Concatenate operation, respectively. In addition, the ConvBlock is the same as that described in Fig. 3.

TABLE 2. Network parameters of the 3DMB-IResNet. Building blocks are shown in brackets with the number of blocks stacked. The symbol "*" indicates there is no such operation.

Layer name	Branch-S	Branch-M	Branch-L
Input size	18×18×18	30×30×30	40×40×40
ConvBlock_x	[3×3×3, 96]×2	[3×3×3, 64]×2	[3x3x3, 3]× 2
IResBlock1	p=32, q=128	p=24, q=96	p=16, q=64
IResBlock2	*	p=32, q=128	p=24, q=96
Output size	9×9×9	8×8×8	10×10×10
Probability	Global Average Pooling, Concatenate, 2-d fc, softmax		

TABLE 3. Network parameters of the 3DMB-DenseNet. Building blocks are shown in brackets with the number of blocks stacked. The growth rate for the networks is 32. Where TL MEANS transition layer.

Layer name	Branch-S	Branch-M	Branch-L
Input size	18×18×18	30×30×30	40×40×40
ConvBlock_x	[3×3×3, 96]×2	[3×3×3, 64]×2	[3×3×3, 32]×2
TL	1×1×1 Conv, Central Pooling		
DenseBlock1_x	$\begin{bmatrix}1\times1\times1\ Conv\\3\times3\times3\ Conv\end{bmatrix}\times5$	$\begin{bmatrix}1\times1\times1\ Conv\\3\times3\times3\ Conv\end{bmatrix}\times4$	$\begin{bmatrix}1\times1\times1\ Conv\\3\times3\times3\ Conv\end{bmatrix}\times3$
TL	*	1x1x1 Conv, Central Pooling	
DenseBlock2_x	*	$\begin{bmatrix}1\times1\times1\ Conv\\3\times3\times3\ Conv\end{bmatrix}\times7$	$\begin{bmatrix}1\times1\times1\ Conv\\3\times3\times3\ Conv\end{bmatrix}\times6$
Output size	9×9×9	8×8×8	10×10×10
Probability	Global Average Pooling, Concatenate, 2-d fc, softmax		

The Branch-S of 3DMB-VggNet consists of four convolutional layers, a central pooling layer [43] and a fully connected layer. The four convolution layers of the Branch-S are divided into two groups, each group consisting of two stacked convolution layers (as shown in Fig. 3 (b)). The Branch-M of 3DMB-VggNet consists of six convolutional layers, two central pooling layers and one fully connected layer. The six convolution layers of the Branch-M are divided into three groups, and each group is composed of two stacked convolution layers. The overall structure of the Branch-L of 3DMB-VggNet is similar to that of Branch-M, except that the number of neurons included in the fully connected layer is different. In addition, in order to alleviate the gradient vanishing and speed up the convergence, the batch normalization operation [44] is used, and after each convolution, we use the nonlinear parameter rectification linear unit (PReLU) as the activation function [45].

2) 3D MULTI-BRANCH InceptionResNet

Fig. 4 shows the architecture of the proposed 3D multi-branch InceptionResNet (3DMB-IResNet). Table 2 lists the corresponding network parameters.

As shown in Fig. 4, the network incorporates deep branches of three different structures, and the inputs corresponding to each branch are also different. The Branch-S of 3DMB-IResNet contains 15 convolutional layers, a central pooling layer and a global average pooling layer. The 15 convolutional layers of the Branch-S are divided into two categories; one is a ConvBlock which is composed of two stacked convolution layers (as shown in Fig. 3 (b)), and the other only contains an IResBlock (as shown in Fig. 4(b)). The Branch-M of 3DMB-IResNet consists of 28 convolutional layers, two central pooling layers and one global average pooling layer. The 28 convolutional layers of the Branch-M are also divided into two categories; the first one is the same as the Branch-L, and the second one is the IResBlock group consisting of two stacked IResBlock. The overall structure of the Branch-L of 3DMB-IResNet is similar to that of Branch-M. The difference is the number of feature maps for each convolutional layer (please see Table 2 for details). In addition, the batch normalization operations and PReLU are also used.

3) 3D MULTI-BRANCH DenseNet

We present the architecture of proposed 3D multi-branch dense network (3DMB-DenseNet) in Fig. 5. Table 3 lists its corresponding network parameters.

The overall architecture of the network, similar to that of 3DMB-IResNet, also incorporates deep branches of three different structures. The Branch-S of 3DMB-DenseNet consists of 13 convolutional layers, a central pooling layer and a global averaging pooling layer. The 13 convolutional layers of the Branch-S are divided into two categories; one is a convolution block composed of two stacked convolution layers (same as those shown in Fig. 3 (b)), and the other is a dense block group composed of five stacked dense blocks (as shown in Fig. 5(b)). The Branch-M of 3DMB-DenseNet consists of 26 convolutional layers, two central pooling layers

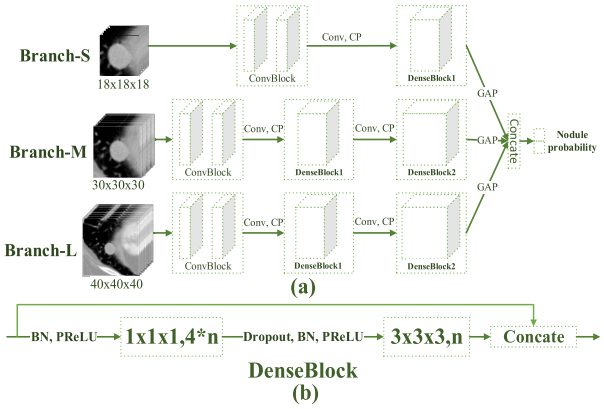


FIGURE 5. (a) The proposed 3D multi-branch DenseNet (3DMB-DenseNet) and (b) The diagram of the 3D dense block (DenseBlock). The parameter n indicates the number of channels. The CP, GAP and Concate represent the Central Pooling, Global Average Pooling operation and Concatenate operation, respectively. The ConvBlock is the same as that described in Fig. 3.

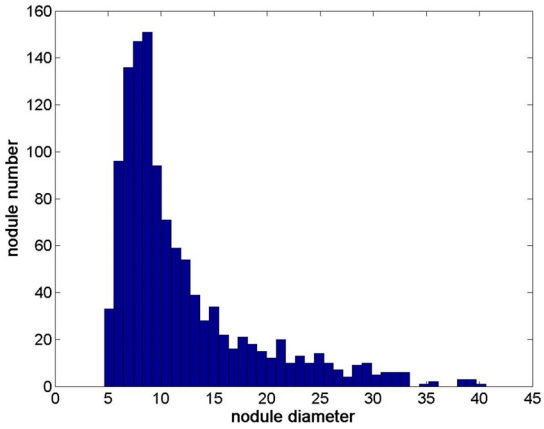


FIGURE 6. Distribution of the sizes of lung nodules, with diameters measured in voxels.

and a global averaging pooling layer. The 26 convolutional layers of the Branch-M are divided into three categories; the first one is the same as the Branch-S, the second one is the dense block group consisting of four stacked dense blocks, and the third one is a dense block group consisting of seven stacked dense blocks. The overall structure of the Branch-L of 3DMB-DenseNet is similar to that of Branch-M, except that the number of convolutional layers is different. Specifically, Branch-L containing 24 convolutional layers, which are similar to Branch-M, are classified into three categories, and its second and third categories consisting of three and six stacked dense blocks, respectively. The rest of operations such as the fully connected layer and the activation functions are the same as those used in the 3DMB-IResNet.

B. MULTI-BRANCH ARCHITECTURE

In the false positive reduction method for 3D CNN-based lung nodule detection, a cube centered on the center location of the candidate nodule is extracted and fed to the CNN network for training. It is critical to select a size of training samples for achieving the good recognition performance of the CNN network. The size of the receptive field of the network will also have a significant impact on the final predictions. If the size of the training sample is too small, the network will not be able to obtain sufficient contextual information for training, and for larger size nodules, it will not be able to adapt. On the other hand, if the size of the training sample (i.e. a patch) is too large, the input patch may contain a lot of redundant information, even including noise, especially when the number of training samples is limited, which will greatly reduce the performance of the network. Due to the difficulty in selecting an appropriate receptive field that can obtain rich contextual information without introducing noise, we have designed a multi-branch network architecture with multiple scales to meet this challenge. Meanwhile, the classification layer of the proposed multi-branch network architecture combines the learned features of different branches to improve the discriminating ability of the network.

To determine the specific size for an input sample, we analyzed the size distribution of the lung nodules. Fig. 6 shows the diameter distribution of lung nodules in the annotations.csv file provided by LUNA16. Note that data resampling is performed to ensure that the ratio of spacing in the X, Y, and Z directions is 1:1:1. Through the analysis of Fig. 6, we have chosen three different input sizes. Firstly, we can see that the horizontal-axis coordinate corresponding to the peak of nodule diameters is the location with a diameter of nine. Based on this observation, we designed the first branch Branch-S, whose input size is $18 \times 18 \times 18$. For a smaller lung nodule, this size can contain appropriate contextual information without introducing too much redundant information. Secondly, we can see that most of the nodules are less than 15 in diameter. Based on this, a larger size of $30 \times 30 \times 30$ is used as the input for the second branch Branch-M. This scale of lung nodules occupies a large proportion, which can provide rich contextual information for small nodules and appropriate contextual information for medium-sized nodules. For a large nodule, it can also obtain its main part and exclude some edge (noise) areas. Finally, we constructed a Branch-L with an input size of $40 \times 40 \times 40$, which basically covers all sizes of the nodules. This scale can provide rich context information for medium-sized nodules, but may introduce noise to small nodules. However, compared to previous two sizes, this size can better handle a larger diameter of lung nodules.

C. OFFLINE HARD HINING

The ratio of positive and negative samples in the original data set is approximately 1:490. In order to solve the problem of imbalanced data consisting of positive samples (true lung nodules) and negative samples (detecting wrong lung nodules); we adopted two methods to alleviate this problem. One method is to reduce the proportion of positive and negative

samples by means of data augmentation. The other method is to gradually select positive and negative samples for training based on the idea of hard mining.

Specifically, we first randomly select 150,000 negative samples from the training set, and then augment the positive samples so that the ratio of positive and negative samples of the initial training is 1:1. After the training is completed, we use this initial false positive reduction model to predict all positive and negative samples in the training set. It should be noted that, assuming that the model needs to be trained for N epochs to converge, we only use the model corresponding to the N-th epoch to predict the training set. In addition, the initially obtained model (the model of the first iteration) refers to the model after training the N epoch on the training set, and the training set is not predicted in the process. Only in the latter iterations, the training set will be predicted by the model obtained in the previous iteration.

To ensure that the proportion of positive and negative samples is balanced at each iteration (second and subsequent iterations), we adopted the following strategy: if the number of positive samples of the misclassification is small, the positive sample is expanded. Otherwise, the negative sample will be expanded. In addition, during each iteration, our learning rate will continue to decline. The purpose of this is to prevent over-fitting and slowly approximate the optimal solution at a gradually decreasing pace. We designed two methods to expand the positive samples; one is to randomly sample the sample points within the radius of the lung nodule, and then extract the data block centered at that point, and the other is to rotate 90°, 180° and 270° in the axial plane. In the offline hard mining process, if we need to expand the negative samples, the rotation and oversampling can be used.

D. TRAING PROCEDURE

During the training procedure, the initialization was done using Xavier method [46] and the model parameters were updated using the SGD optimizer. Our initial learning rate was set to 0.001. After that, each epoch's learning rate is decreased to 95% of the previous learning rate. Since there is no suitable pre-training weights that can be used in our proposed models, we set a relatively high learning rate to train the model from scratch. The batch size for 3DMB-VggNet, 3DMB-IResNet and 3DMB-DenseNet is 64, 40, and 64, respectively. The momentum size is 0.9 for all three nets. In addition, the training and validation of these three network architectures are independent of each other, and the training objectives are maximizing the probability of correct classification. We achieve this by minimizing the cross-entropy loss of the training samples. For an input sample with a true label y, the loss function is defined in equation (1):

$$L = -\frac{1}{N}\sum_{n=1}^{N}[y_n log(y'_n) + (1 - y_n) log(1 - y'_n)] \qquad (1)$$

where y' represents the predicted probability of the model and N is the number of samples.

IV. DATA AND EXPERIMENTS

We describe the information about the dataset used in our experiments in section A. Section B gives the evaluation criteria and Section C explains the environment for the experiments.

A. DATA

The data set used in our experiments is from the LUNA16 Challenge, which includes 551,065 candidate nodules and 1,120 true lung nodules [47]. The LUNA16 Challenge screened 888 CT data from a publicly available LIDC dataset containing 2610 lung nodules, each of which was labeled by up to four experienced radiologists [48]. During this labeling process, each radiologist classifies the identified lesions into three categories: non-nodule (other tissues or backgrounds), nodules larger than 3-mm in diameter, and nodules less than 3-mm in diameter. Then, a nodule with a diameter greater than 3-mm marked by at least three radiologists is selected as the gold standard. Finally, nodules that are less than 3-mm in diameter and marked by only one or two radiologists will be ignored.

B. EVALUATION CRITERIA

We use the metrics of the LUNA16 competition to evaluate our proposed approach. Based on the 10-fold cross validation, the evaluation is performed by measuring the detection sensitivity of the algorithm and the false positive rate corresponding to each scan. If the center of the predicted candidate nodule is within the radius of the real nodule, the prediction is considered correct. If an unrelated candidate nodule is detected, it is ignored. At the same time, the 95% confidence interval is also calculated using the bootstrapping algorithm [49]. The competition performance metric (CPM) was defined as the average sensitivity of seven predefined false positive rates (These seven values are 0.125, 0.25, 0.5, 1, 2, 4, and 8) [50], and calculated as follows.

$$CPM = \frac{1}{N}\sum_{i=\{0.125,0.25,0.5,1,2,4,8\}} Recall_{fpr=i} \qquad (2)$$

where the value of "N" is seven, "fpr" represents the average number of false positives per scan, and "$\text{Recall}_{\text{fpr=i}}$" represents the corresponding recall rate at fpr=i.

C. ENVIRONMENT FOR THE EXPERIMENTS

We ran our experiments on a server with an Intel(R) Xeon(R) processor with 125GB memory. The integrated development tool we used was PyCharm and the encoding language was Python 3.6. In addition, we built our network architecture based on the Keras deep learning framework. When we trained the model, we used the NVIDIA GTX-1080Ti GPU (11GB video memory) for acceleration. Based on our experiments, the time required for the 3DMB-VggNet, 3DMB-IResNet and 3DMB-DenseNet models to converge was approximately 36, 30 and 34 hours, respectively.

TABLE 4. The sensitivities of models under different false positive rates are shown in the percentage including the CPM. Note that, "IResNet" indicates "InceptionResNet". "*Net_Best" indicates the optimal performance based on *Net. For example, "3DMB-VggNet_Best" represents the optimal performance obtained based on 3DMB-VggNet, which corresponds to "3DMB-VggNet_IS_CP" in Table 5. In addition, "3DMB-IResNet_Best" corresponds to "3DMB-IResNet_IS_CP-OHM_2" in Table 6, and "3DMB-DenseNet_Best" corresponds to "3DMB-DenseNet_IS_CP" in Table 5.

Method	0.125	0.25	0.5	1	2	4	8	CPM
3DMB-VggNet_Best	67.5	72.9	77.7	84.8	88.0	91.1	95.7	82.5
3DMB-IResNet_Best	55.8	69.1	80.1	84.6	90.3	92.8	96.0	81.2
3DMB-DenseNet_Best	61.2	69.7	76.3	83.6	91.4	95.0	98.1	82.2
MBEL-3D-CNN_Best	72.9	77.9	85.8	91.5	93.7	95.2	97.6	87.8

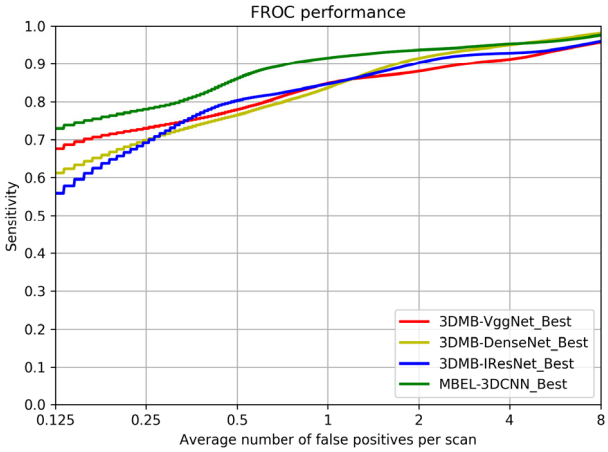


FIGURE 7. The FROC curves for three networks and their integration.

V. RESULTS AND DISCUSSION

We give the overall performance of the proposed networks in section A. Section B describes the ablation study of the proposed networks and Section C provides a comparison with other false positive reduction methods.

A. OVERALL PERFORMANCE

To show the overall performance of the proposed network, we give a further quantitative analysis on the performance of three different network architectures (i.e. 3DMB-VggNet, 3DMB-IResNet, and 3DMB-DenseNet). The FROC (Free-response Receiver Operating Characteristic) curves for the three networks and their integration are shown in Fig. 7. By comparing the curves, in the case of a false positive rate of 8 per scan (as shown in the curves), the sensitivity of each of three networks can obtain more than 95%, which indicates that 3D-CNN can extract discriminating features from CT images.

Table 4 lists the sensitivity of these three networks with different false positive rates specified. It can be seen that the recall rate of three networks is less than 85% when the false positive rate of each scan is one. In addition, in order to increase the difficulty of the challenge, the organizer of LUNA16 Challenge added several extremely low false positive rates (0.5, 0.25, 0.125 false positives per scan) to the evaluation parameters, which is critical to improve the automated level of the current computer-aided diagnosis system. In this case, our multi-model integration strategy demonstrates the ability to reduce false positives while maintaining satisfactory sensitivity. For example, when each scan limits to 0.125 false positives, the sensitivity of 3DMB-VggNet, 3DMB-IResNet, and 3DMB-DenseNet is only 67.5%, 55.8%, and 61.2%, respectively, but the sensitivity of the integrated model can reach 72.9%. These experiments show that different network architectures can complement each other, and their integration can significantly improve the performance.

B. ABLATION STUDY

To validate the effectiveness of the components in the MBEL-3D-CNN architecture, we designed a 2D VggNet-based ablation experiment. Experimental results are shown in Table 5. It should be noted that the training of each model in Table 5 has not used the offline hard mining operation. These models extend the number of positive samples to the scale of negative samples by performing rotation (rotating 90°, 180° and 270° on the axial plane), translation (translating one voxel along each axis) and oversampling (oversampling four times based on the results of rotation and translation) to achieve data balance during training. The training procedure and network parameters are the same as described in Section III.

1) EFFECTIVENESS OF THE SPATIAL INFORMATION IN 3D SPACE

In Table 5, 2D-VggNet_MP is a Vgg architecture based on a 2D-CNN with an input size of 32*32*3, and its corresponding score CPM is 33.8%. Similarly, 3D-VggNet_MP is a Vgg architecture based on a 3D-CNN with an input size of 32*32*10, and its corresponding score CPM is 77.6%. Compared with 2D-VggNet_MP, its performance is improved by 2.3 times. By comparing these two results, we can see that although the model parameters have increased, the performance has been greatly improved, which verifies the effectiveness of using more spatial information.

2) EFFECTIVENESS OF THE ISOTROPIC OPERATION

In Table 5, the network architecture and training mode of 3D-VggNet_IS_MP and 3D-VggNet_MP are identical.

TABLE 5. Ablation experiment. Please note that, "MP" indicates max pooling, "CP" indicates central pooling, "IS" indicates isotropic operation, "IResNet" indicates "InceptionResNet", and "ParaA" means the amount of parameters that need to be trained.

Method	ParaA	0.125	0.25	0.5	1	2	4	8	CPM
2D-VggNet_MP	0.5×10^7	4.20	8.30	16.6	33.2	48.1	56.5	69.6	33.8
3D-VggNet_MP	1.4×10^7	62.4	74.4	83.5	87.2	90.3	93.4	77.6	52.3
3D-VggNet_IS_MP	1.4×10^7	56.9	66.6	77.4	82.9	85.6	90.3	92.3	78.9
3D-VggNet_IS_CP	1.4×10^7	52.6	65.7	78.1	86.7	89.2	92.9	96.2	80.3
3DMB-VggNet_IS_CP	4.4×10^7	67.5	72.9	77.7	84.8	88.0	91.1	95.7	82.5
3DMB- IResNet_IS_CP	2.4×10^7	53.8	64.2	72.6	82	86.9	92.5	96.0	78.3
3DMB-DenseNet_IS_CP	4.2×10^7	61.2	69.7	76.3	83.6	91.4	95.0	98.1	82.2
MBEL-3D-CNN	*	67.5	77.3	82.4	86.3	90.0	94.4	96.5	84.9

TABLE 6. Performance metrics for each network by using OHM operations. Please note that, "IS" indicates an isotropic operation, "CP" indicates central pooling, "IResNet" indicates "InceptionResNet", and "OHM_n" represents the n-th iteration of the OHM operation.

Method	0.125	0.25	0.5	1	2	4	8	CPM
3DMB-VggNet_IS_CP-OHM_1	47.6	58.2	65.8	82.3	86.1	89.4	92.7	74.6
3DMB-VggNet_IS_CP-OHM_2	52.9	64.6	76.6	84.7	88.1	93.0	97.8	81.0
3DMB-IResNet_IS_CP-OHM_1	56.4	65.0	71.5	81.9	87.7	90.9	94.7	78.3
3DMB-IResNet_IS_CP-OHM_2	55.8	69.1	80.1	84.6	90.3	92.8	96.0	81.2
3DMB-DenseNet_IS_CP-OHM_1	47.8	60.5	77.6	86.6	91.4	95.4	95.7	79.3
3DMB-DenseNet_IS_CP-OHM_2	49.1	65.7	76.6	83.1	90.0	93.8	95.2	79.2
MBEL-3D-CNN-OHM	62.5	75.6	87.1	90.7	92.6	95.3	95.9	85.7

The difference is that 3D-VggNet_IS_MP performs an isotropic operation on the input data, that is, the spacing of the CT images in the x, y, and z directions is uniformly spaced by the linear interpolation, so that the ratio of spacing in the x, y, and z directions is 1:1:1. Since we the spacing is uniform in all directions, we change the input size of 3D-VggNet_IS_MP to 32*32*32. According to the results of the second and third rows in Table 5, the isotropic operation is useful for model training.

3) EFFECTIVENESS OF THE CENTRAL POOLING

In Table 5, the difference between 3D-VggNet_IS_CP and 3D-VggNet_IS_MP is the pooling mode of the network. The 3D-VggNet_IS_MP uses 3D max pooling while the 3D-VggNet_IS_CP uses 3D central pooling which is an extension of a 2D center pooling proposed by Wang *et al.* [43]. By comparing the results of 3D-VggNet_IS_CP and 3D-VggNet_IS_MP, it can be proved that the 3D central pooling is effective.

4) EFFECTIVENESS OF THE MULTI-BRANCH

In Table 5, 3DMB-VggNet_IS_CP is based on 3D-VggNet_IS_CP by adding two different branches; one is a smaller scale branch with input 24 * 24 * 24 and the other is a larger scale branch with an input of 40 * 40 * 40. Compared to the single-branch used in 3D-VggNet_IS_CP, the performance of multi-branch is increased by 2.2 percentage points, which verifies the effectiveness of the multi-branch architecture.

5) EFFECTIVENESS OF THE ENSEMBLE LEARNING

The last four rows of Table 5 verify the effectiveness of the ensemble learning approach. The network architectures of 3DMB-IResNet_IS_CP and 3DMB-DenseNet_IS_CP are different from that of 3DMB-VggNet_IS_CP. Other than that, they are similar. The MBEL-3D-CNN is the result of their weighted average performance of three networks. Comparing the experimental results of the last four rows in Table 5, the integrated result MBEL-3D-CNN is 2.4 percentage points higher than the best performing network of 3DMB-VggNet_IS_CP.

6) EFFECTIVENESS OF THE OFFLINE HARD MINING OPERATION

Finally, in order to verify the effectiveness of the Offline Hard Mining (OHM) operation we show the results of the OHM operation in Table 6. By observing the results in Table 5 and 6, it can be seen that the performance of 3DMB-IResNet_IS_CP using OHM operation has been improved. Although 3DMB-VggNet_IS_CP-OHM_2 and 3DMB-DenseNet_IS_CP-OHM_2 are less effective than those without OHM operation (3DMB-VggNet_IS_CP and 3DMB-DenseNet_IS_CP in Table 5), their improvement after integrated with 3DMB-IResNet_IS_CP-OHM_2 is better than MBEL-3D-CNN (see Table 5). This shows that the

TABLE 7. A comparison on the quantitative results of various false positive reduction methods. Among them, "Ours" indicates the result of the 10-fold cross validation of our proposed method. Please note that, the unit of recall rate and final score CPM is "%". The best performance is shown in bold.

Method	Data Type	0.125	0.25	0.5	1	2	4	8	CPM
Lee [39]	2D	26.2	035.9	47.2	58.6	67.4	71.3	76.6	54.8
Roth [36]	2.5D	49.9	59.8	66.6	70.5	75.6	80.3	82.9	69.4
Dobrenkii [42]	3D	51.7	60.2	72.0	78.8	82.2	83.9	85.6	73.5
Setio [40]	2.5D	63.6	72.7	79.2	84.4	87.6	90.5	91.6	81.4
Dou [41]	3D	67.7	73.7	81.5	84.8	87.9	90.7	92.2	82.7
Ours	**3D**	**69.5**	**80.1**	**85.1**	**91.4**	**93.9**	**94.9**	**96.1**	**87.3**

overall performance is improved by using the OHM operation. Moreover, by comparing the last row of Table 6 with other rows, the effectiveness of the ensemble learning approach can be further verified. Please note that after the second iteration of the OHM operation, the performance is hardly improved again. Hence, we only performed two iterations.

C. COMPARISON WITH OTHER METHODS

In order to verify the superiority of the proposed method, we compare it with the false positive reduction method published in recent years. Table 7 gives a comparison on the quantitative results of six false positive reduction methods including ours. The unit of recall rate and final score CPM is in the percentage (i.e. %). According to the experimental results in Table 7, the 2D CNN-based method, proposed by Lee *et al.* [39], only utilizes the local information of the nodule, and the final score CPM is only 54.8% which is not efficient. The methods, proposed by Roth *et al.* [36] and Setio *et al.* [40], feed multiple slices of different views and different angles together into the network for training, which enables the network to see more spatial information. These two methods have some improvement over the single view and single angle method. However, the method proposed by Setio has the best CPM with 81.4%. In order to make the maximized utilization of the spatial information of nodules, Dobrenkii *et al.* [42] and Dou *et al.* [41], respectively, proposed a false positive reduction method based on 3D-CNN, and they obtained better results. Our proposed method based on 3D-CNN performs better than five existing false positive reduction methods in our experiments as shown in Table 7. It should be noted that the experimental results described in Sections "OVERALL PERFORMANCE" and "ABLATION STUDY" above are the results of one-fold of the randomly selected from the ten-fold cross validation results. The experimental results shown in Table 7 are the results of the 10-fold cross validation.

D. THE REASONS FOR CHOOSING VggNet, IResNet AND DenseNet IN ENSEMBLE LEARNING

We chose these three network architectures for integration for the following two reasons. 1) These three network models are classic and widely used network architecture. Meanwhile, they are also the classification network used by the top teams in the ImageNet competition. 2) We have implemented six common network architectures: VggNet, ResNet, Inception-V1 (IncepV1), Inception-V3 (IncepV3), InceptionResNet (IResNet) and DenseNet. Then, we selected the three network structures (VggNet, IResNet and DenseNet) through experimental verification. The specific experiment is described below.

TABLE 8. The CPM scores for the six network models.

	VggNet	ResNet	DenseNet	IncepV1	IncepV3	IResNet
CPM	82.5	78.9	82.2	72.5	80.0	81.2

TABLE 9. The corresponding CPM scores for the integration of related models. Please noted that the symbol "$*$" indicates that the corresponding model is used for integration, and the "E_N" indicates the Nth experiment.

	DenseNet	IResNet	VggNet	ResNet	IncepV1	IncepV3	CPM
E_1	*						82.2
E_2	*	*					84.6
E_3	*	*	*				87.8
E_4	*	*	*	*			87.8
E_5	*	*	*	*	*		87.9
E_6	*	*	*	*	*	*	87.4

TABLE 10. Model selection experiments. Please noted that the symbol "*" indicates that the corresponding model is used for integration, and the "EN" indicates the Nth experiment.

	DenseNet	ResNet	VggNet	ResNet	IncepV1	IncepV3	CPM
E_1	*	*	*				87.4
E_2	*	*		*			86.5
E_3	*		*	*			87.8
E_4		*	*	*			87.6
E_5			*	*	*		86.6
E_6			*	*		*	85.9
E_7			*		*	*	85.3
E_8				*	*	*	85.2

The corresponding experimental data is shown in Table 8 and Table 9. Table 8 shows the CPM scores corresponding to the six networks, and Table 9 shows the CPM scores of their integrations. It should be noted that the symbol "*" indicates that the corresponding model is used for integration, and the "E_N" indicates the Nth experiment.

In order to improve the generalization ability of the classification method, we combine multiple models according to the idea of "expert consultation" to simulate the public

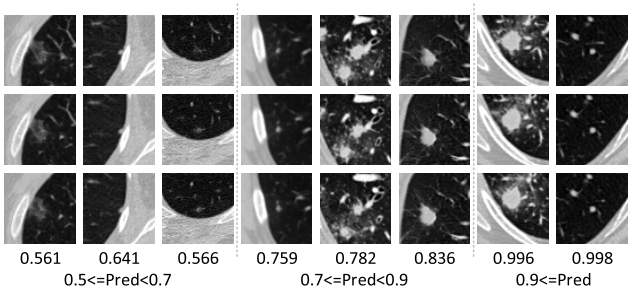


FIGURE 8. The partial samples randomly selected from FP samples.

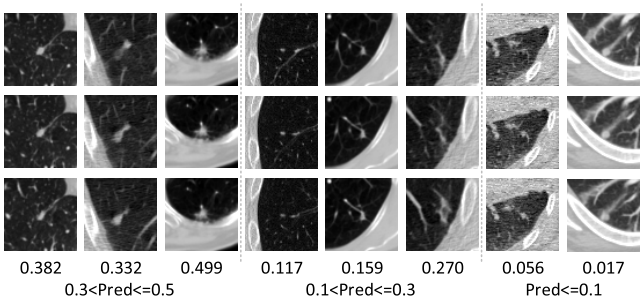


FIGURE 9. The partial samples randomly selected from FN samples.

decision-making process of experts. As for the question of how many models should be integrated, we conducted the following experiments, as shown in Table 9.

According to the results of Table 9, we can see that the integration from the two models to the integration of the three models, the corresponding performance growth is the largest. Although continuing to increase the number of models used for integration helps to further improve overall performance, the magnitude of the increase is very weak (when the number of integrations is large, the overall performance may even decline), and at the same time the cost is high, especially in the training of the model. Based on this, we have designed the following experiment to select the three best models for integration, so as to achieve the mutual compensation of the defects between the models. The relevant experimental results are shown in Table 10, where the symbol "*" indicates that the corresponding model is used for integration, and the "E_N" indicates the Nth experiment. In addition, it should be noted that in the results shown in Table 9 and Table 10, each model has the same importance, that is, if it is the integration of N models, the weight of each model is 1/N when integrated.

From the experimental results shown in Table 10, it is not difficult to draw the following conclusions. For the task of reducing the false positive of lung nodules, the integration of three models with different network structures and certain representation capabilities can achieve satisfactory results (Compare the results of the fourth row "E3" of Table 10 with the other rows of Table 10). On the contrary, the integration of three models with similar network structures and weak representation ability is not effective (Compare the results of the last row "E8" of Table 10 with the other rows of Table 10).

E. CAUSE ANALYSIS OF ERROR DETECTION

We used 0.5 as the threshold and obtained 31 false positive (FP) samples and 32 false negative (FN) samples by predicting the testing set. Fig. 8 and Fig. 9 show the randomly selected partial FP samples and FN samples, respectively. Among them, each column of Fig. 8 and Fig. 9 shows the upper, middle and lower three-layer images centered on the candidate points. Meanwhile, the FP samples are classified into [0.5, 0.7), [0.7, 0.9) and [0.9,) according to the predicted values. Similarly, the FN samples are classified into (, 0.1), (0.1, 0.3] and (0.3, 0.5) according to the predicted values.

From the observation of Fig. 8, it is found that plaques or flocs having a high contrast with the surrounding pixels appear in the center of the window of most of the FP samples, and their shapes and intensity are very similar to the real lung nodules. This makes it difficult for convolutional networks to distinguish between such samples. In addition, as can be seen from Fig. 8, the larger the bright plaque appearing in the center of the sample or the higher the contrast with the surrounding pixels, the greater the probability that the model predicts it as a positive sample.

It is not difficult to find from the example shown in Fig. 9 that most of the FN samples are lung nodules that adhere to vascular tissue or the lung wall. When the lung nodules adhere to the vascular tissue or the lung wall, the boundary of the lung nodules become blurred, making it difficult for the convolutional network to extract more effective features from the image, which may result in the classifier not being able to adapt to such samples. Furthermore, by observing the three probability intervals of the FN sample, we can find that the more blurred the visual, the lower the probability that the model will predict it as a positive sample.

VI. CONCLUSION

Our goal in this study is to adapt the heterogeneity of nodules for lung nodule detection. We proposed a multi-branch framework based on 3D-CNN for performing the detection. For the similarity between lung nodules and non-nodules, we use an OHM operation for the model to handle those indistinguishable positive and negative samples. Simultaneously, based on the idea of "expert consultation", we build multiple independent networks to simulate different expert's behaviors, and then use ensemble learning to fuse the prediction result from each of multiple models to have the best decision of the experts. The ensemble learning makes our proposed networks more stable. We also verified the various components and overall performance of the proposed network architecture through the ablation study and made a comparison on several false positive reduction methods based on our experimental results. According to the experimental results shown in Table 5, 6 and 7, our approach has achieved competitive results. In the future work, we plan to develop a candidate

lung nodule detection method based on the Deconvolutional Single Shot Detector (DSSD) network architecture [51], and combine it with our method to hopefully achieve a superior lung nodule detection algorithm.

ACKNOWLEDGMENT

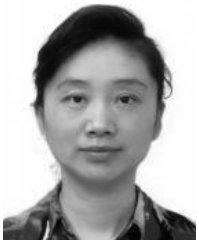
The authors acknowledge the National Cancer Institute and the Foundation for the National Institutes of Health and their critical role in the creation of the free publicly available LIDC-IDRI Database used in this study.

REFERENCES

[1] J. Ferlay *et al.*, "Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012," *Int. J. cancer*, vol. 136, no. 5, pp. E359–E386, 2015.
[2] K. Bhavanishankar and M. V. Sudhamani, "Techniques for detection of solitary pulmonary nodules in human lung and their classifications—A survey," *Int. J. Cybern. Inf.*, vol. 4, no. 1, pp. 27–40, 2015.
[3] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2018," *CA, Cancer J. Clin.*, vol. 68, no. 1, pp. 7–30, Jan. 2018.
[4] C. I. Henschke, "Early lung cancer action project: Overall design and findings from baseline screening," *Cancer*, vol. 89, pp. 2474–2482, Dec. 2000.
[5] 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, pp. 390–406, Jun. 2010.
[6] 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.
[7] M. Woźniak, D. Połap, G. Capizzi, G. L. Sciuto, L. Kośmider, and K. Frankiewicz, "Small lung nodules detection based on local variance analysis and probabilistic neural network," *Comput. Methods Programs Biomed.*, vol. 161, pp. 173–180, Jul. 2018.
[8] I. R. S. Valente, P. C. Cortez, E. C. Neto, J. M. Soares, V. H. C. de Albuquerque, and J. M. R. Tavares, "Automatic 3D pulmonary nodule detection in CT images: A survey," *Comput. Methods Programs Biomed.*, vol. 124, pp. 91–107, Feb. 2016.
[9] Y. Chen, C. Liu, W. Peng, and S. Xia, "Thyroid nodule detection using attenuation value based on non-enhancement CT images," in *Proc. 2nd IET Int. Conf. Biomed. Image Signal Process. (ICBISP)*, 2017, pp. 1–4.
[10] L. Lu, Y. Tan, L. H. Schwartz, and B. Zhao, "Hybrid detection of lung nodules on CT scan images," *Med. Phys.*, vol. 42, no. 9, pp. 5042–5054, 2015.
[11] M. P. Paing and S. Choomchuay, "A computer aided diagnosis system for detection of lung nodules from series of CT slices," in *Proc. 14th Int. Conf. Elect. Eng./Electron., Comput., Telecommun. Inf. Technol. (ECTI-CON)*, Jun. 2017, pp. 302–305.
[12] A. Gupta, O. Martens, Y. Le Moullec, and T. Saar, "Methods for increased sensitivity and scope in automatic segmentation and detection of lung nodules in CT images," in *Proc. IEEE Int. Symp. Signal Process. Inf. Technol. (ISSPIT)*, Dec. 2015, pp. 375–380.
[13] A. A. Rezaie and A. Habiboghli, "Detection of lung nodules on medical images by the use of fractal segmentation," *Int. J. Interact. Multimedia Artif. Intell.*, vol. 4, no. 5, pp. 15–19, 2017.
[14] B. Wang *et al.*, "Pulmonary nodule detection in CT images based on shape constraint CV model," *Med. Phys.*, vol. 42, pp. 1241–1254, Mar. 2015.
[15] G. Aresta, A. Cunha, and A. Campilho, "Detection of juxta-pleural lung nodules in computed tomography images," *Proc. SPIE*, vol. 10134, Mar. 2017, Art. no. 101343N.
[16] A. O. de Carvalho Filho, A. C. Silva, A. C. de Paiva, R. A. Nunes, and M. Gattass, "3D shape analysis to reduce false positives for lung nodule detection systems," *Med. Biol. Eng. Comput.*, vol. 55, no. 8, pp. 1199–1213, 2017.
[17] B. R. Froz, A. O. de Carvalho Filho, A. C. Silva, A. C. de Paiva, R. A. Nunes, and M. Gattass, "Lung nodule classification using artificial crawlers, directional texture and support vector machine," *Expert Syst. Appl.*, vol. 69, pp. 176–188, Mar. 2017.
[18] E. A. Khordehchi, A. Ayatollahi, and M. R. Daliri, "Automatic lung nodule detection based on statistical region merging and support vector machines," *Image Anal. Stereology*, vol. 36, no. 2, pp. 65–78, 2017.
[19] E. E. Nithila and S. S. Kumar, "Automatic detection of solitary pulmonary nodules using swarm intelligence optimized neural networks on CT images," *Eng. Sci. Technol., Int. J.*, vol. 20, no. 3, pp. 1192–1202, 2017.
[20] 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.
[21] 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, Nov. 2016, Art. no. 6215085.
[22] L. Fu, J. Ma, Y. Ren, Y. S. Han, and J. Zhao, "Automatic detection of lung nodules: False positive reduction using convolution neural networks and handcrafted features," *Proc. SPIE*, vol. 10134, Mar. 2017, Art. no. 101340A.
[23] H.-C. Shin *et al.*, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, May 2016.
[24] I. Sluimer, A. Schilham, M. Prokop, and B. van Ginneken, "Computer analysis of computed tomography scans of the lung: A survey," *IEEE Trans. Med. Imag.*, vol. 25, no. 4, pp. 385–405, Apr. 2006.
[25] K. Simonyan and A. Zisserman. (2014). "Very deep convolutional networks for large-scale image recognition." [Online]. Available: https://arxiv.org/abs/1409.1556
[26] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi. (2017). "Inception-v4, inception-ResNet and the impact of residual connections on learning." [Online]. Available: https://arxiv.org/abs/1602.07261
[27] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2017, pp. 2261–2269.
[28] M. Javaid, M. Javid, M. Z. U. Rehman, and S. I. A. 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.
[29] P. Cao *et al.*, "A $\ell 2$, 1 norm regularized multi-kernel learning for false positive reduction in Lung nodule CAD," *Comput. Methods Programs Biomed.*, vol. 140, pp. 211–231, Mar. 2017.
[30] G. Cao, Y. Liu, and K. Suzuki, "A new method for false-positive reduction in detection of lung nodules in CT images," in *Proc. 19th Int. Conf. Digit. Signal Process.*, Aug. 2014, pp. 474–479.
[31] P. Pudil, J. Novovičová, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognit. Lett.*, vol. 15, no. 11, pp. 1119–1125, 1994.
[32] 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.
[33] S. Saien, H. A. Moghaddam, and M. Fathian, "A unified methodology based on sparse field level sets and boosting algorithms for false positives reduction in lung nodules detection," *Int. J. Comput. Assist. Radiol. Surgery*, vol. 13, pp. 397–409, Mar. 2018.
[34] F. Ciompi *et al.*, "Automatic classification of pulmonary peri-fissural nodules in computed tomography using an ensemble of 2D views and a convolutional neural network out-of-the-box," *Med. Image Anal.*, vol. 26, no. 1, pp. 195–202, Dec. 2015.
[35] S. S. Ramachandran, J. George, S. Skaria, and V. V. Varun, "Using YOLO based deep learning network for real time detection and localization of lung nodules from low dose CT scans," *Proc. SPIE*, vol. 10575, Feb. 2018, Art. no. 105751I.
[36] H. R. Roth *et al.*, "Improving computer-aided detection using convolutional neural networks and random view aggregation," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1170–1181, May 2016.
[37] A. Teramoto, H. Fujita, O. Yamamuro, and T. Tamaki, "Automated detection of pulmonary nodules in PET/CT images: Ensemble false-positive reduction using a convolutional neural network technique," *Med. Phys.*, vol. 43, no. 6, pp. 2821–2827, 2016.
[38] W. Sun, X. Huang, T.-L. B. Tseng, and W. Qian, "Automatic lung nodule graph cuts segmentation with deep learning false positive reduction," *Proc. SPIE*, vol. 10134, Mar. 2017, Art. no. 101343M.
[39] H. Lee, H. Lee, M. Park, and J. Kim, "Contextual convolutional neural networks for lung nodule classification using Gaussian-weighted average image patches," *Proc. SPIE*, vol. 10134, Mar. 2017, Art. no. 1013423.
[40] 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. 1160–1169, May 2016.

[41] 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.
[42] A. Dobrenkii, R. Kuleev, A. Khan, A. R. Rivera, and A. M. Khattak, "Large residual multiple view 3D CNN for false positive reduction in pulmonary nodule detection," in *Proc. IEEE Conf. Comput. Intell. Bioinf. Comput. Biol. (CIBCB)*, Aug. 2017, pp. 1–6.
[43] S. Wang *et al.*, "Central focused convolutional neural networks: Developing a data-driven model for lung nodule segmentation," *Med. Image Anal.*, vol. 40, pp. 172–183, Aug. 2017.
[44] S. Ioffe and C. Szegedy. (2015). "Batch normalization: Accelerating deep network training by reducing internal covariate shift." [Online]. Available: https://arxiv.org/abs/1502.03167
[45] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1026–1034.
[46] G. Xavier and B. Yoshua, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. Artif. Intell. Statist.*, 2010, pp. 249–256.
[47] 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.
[48] S. 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, 2011.
[49] S. Sivaganesan, "An introduction to the bootstrap," *SIAM Rev.*, vol. 36, no. 4, pp. 677–678, 1994.
[50] M. Niemeijer, M. Loog, M. D. Abramoff, M. A. Viergever, M. Prokop, and B. van Ginneken, "On combining computer-aided detection systems," *IEEE Trans. Med. Imag.*, vol. 30, no. 2, pp. 215–223, Feb. 2011.
[51] C.-Y. Fu, W. Liu, A. Ranga, A. Tyagi, and A. C. Berg. (2017). "DSSD: Deconvolutional single shot detector." [Online]. Available: https://arxiv.org/abs/1701.06659

HAICHAO CAO received the bachelor's degree from Nanchang Hangkong University, in 2015. He is currently pursuing the Ph.D. degree with the Huazhong University of Science and Technology. His research areas include digital image processing, medical image analysis, machine learning, and deep learning.

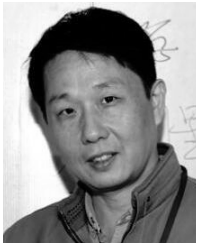


HONG LIU received the Ph.D. degree in electrical engineering and computer from Teesside University, U.K. She is currently a Professor with the School of Computer Science and Technology, Huazhong University of Science and Technology, China. Her current research interests include medical image processing, pattern recognition, and algorithmic design and analysis.

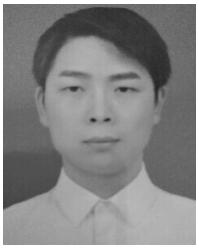
ENMIN SONG received the Ph.D. degree in electrical engineering and computer from Teesside University, U.K. After completing his Ph.D., he was a Postdoctoral Researcher with the University of California at San Francisco. He is currently a Professor with the School of Computer Science and Technology, Huazhong University of Science and Technology, China. His current research interests include medical image processing and medical image information analysis.

GUANGZHI MA is currently an Associate Professor with the School of Computer Science, Huazhong University of Science and Technology. His current research interests include medical image processing, artificial intelligence, and data mining.

XIANGYANG XU is currently an Associate Professor with the School of Computer Science, Huazhong University of Science and Technology. His current research interests include medical image processing and analysis, and machine learning.



RENCHAO JIN is currently a Professor with the School of Computer Science, Huazhong University of Science and Technology. His current research interests include medical image processing and analysis, artificial intelligence, and computer vision.



TENGYING LIU is currently pursuing the master's degree with the School of Computer Science, Huazhong University of Science and Technology. His current research interests include medical image processing and analysis, and deep learning.

CHIH-CHENG HUNG was with Intergraph Corporation for research and development of the remote sensing image processing software, from 1990 to 1993. He is also associated with the Laboratory for Machine Vision and Security Research, Kennesaw State University (KSU), Kennesaw, GA, USA, where he is currently a Professor of computer science. His research interests include image processing, pattern recognition, and artificial intelligence.