IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 12, 2021, accepted December 13, 2021, date of publication December 23, 2021,
date of current version January 4, 2022.

Digital Object Identifier 10.1109/ACCESS.2021.3137317

Pulmonary Nodule Detection Using 3-D Residual U-Net Oriented Context-Guided Attention and Multi-Branch Classification Network

HAIYING YUAN, YANRUI WU, JUNPENG CHENG, ZHONGWEI FAN, AND ZHIYONG ZENG
Faculty of Information Technology, Beijing University of Technology, Beijing 100020, China
Corresponding author: Haiying Yuan (yhyingcn@gmail.com)

This work was supported in part by the Beijing Natural Science Foundation under Grant 4172010, and in part by the National Natural Science Foundation of China under Grant 61001049.

ABSTRACT Accurate detection of pulmonary nodules on chest computed tomography scans is crucial to early diagnosis of lung cancer. To address the thorn problems on low detection sensitivity and high false-positive rate caused by heterogeneity and morphological complexity of 3-D nodule features, a computer-aided detection system is developed to increase the detection sensitivity and classification accuracy of pulmonary nodules. The contributions include: (1) Nodule candidate detection: 3-D Residual U-Net model is improved to detect candidate nodules, which constructs 3-D context-guided module to extract local and global nodule features by setting the dilated convolution with different dilation rates. Furthermore, channel attention mechanism is used to dynamically adjust the channel features, which enhances the generalization and expression ability of the detection-network to effectively learn 3-D spatial context features. (2) False-positive reduction: multi-branch classification network is designed for multi-task learning. Image reconstruction task is performed to retain more microscopic nodules information from convolutional neural network (CNN) hierarchy. Moreover, each branch deals with the feature map at corresponding depth layers, and then all branches' feature maps are combined together to perform nodule classification task. Numerous experimental results show that the proposed system is perfectly qualified for pulmonary nodules detection on Lung Nodules Analysis 2016 dataset, which achieves detection sensitivity up to 94.0% and competition performance metric (CPM) score up to 0.959.

INDEX TERMS 3-D context-guided attention module, multi-task learning, multi-branch classification, convolutional neural network, pulmonary nodule detection.

I. INTRODUCTION

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]. Computed Tomography (CT) images with high resolution and rapid acquisition are widely used in pulmonary nodules detection, so as to realize early screening and intelligent diagnosis of lung cancer. With the rapid increase of disease screening requirements and the availability of CT images under the background of intelligent medical technology, radiologists are facing great pressure from massive image data processing. Hence, computer-aided detection (CAD) systems are applied to assist radiologists to diagnose pulmonary nodules [4], [5]. It improves diagnostic efficiency and accuracy by detecting suspicious nodules for preliminary screening, while reducing missed diagnosis and misdiagnosis of pulmonary nodules.

CAD system typically includes: (1) Nodule candidate detection stage. Candidate nodules are detected from a large number of CT images, all true positive nodules should be included as much as possible, and many false-positive candidates are generated in the process [6]–[8]. (2) False-positive reduction stage [9]–[11]. Candidate nodules are classified by distinguishing true positive nodules from false-positive nodules. Fig. 1 shows some true positive and false-positive nodules. Some nodules have large variations in shapes, size and location, such as calcific nodules, isolated nodule, ground

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

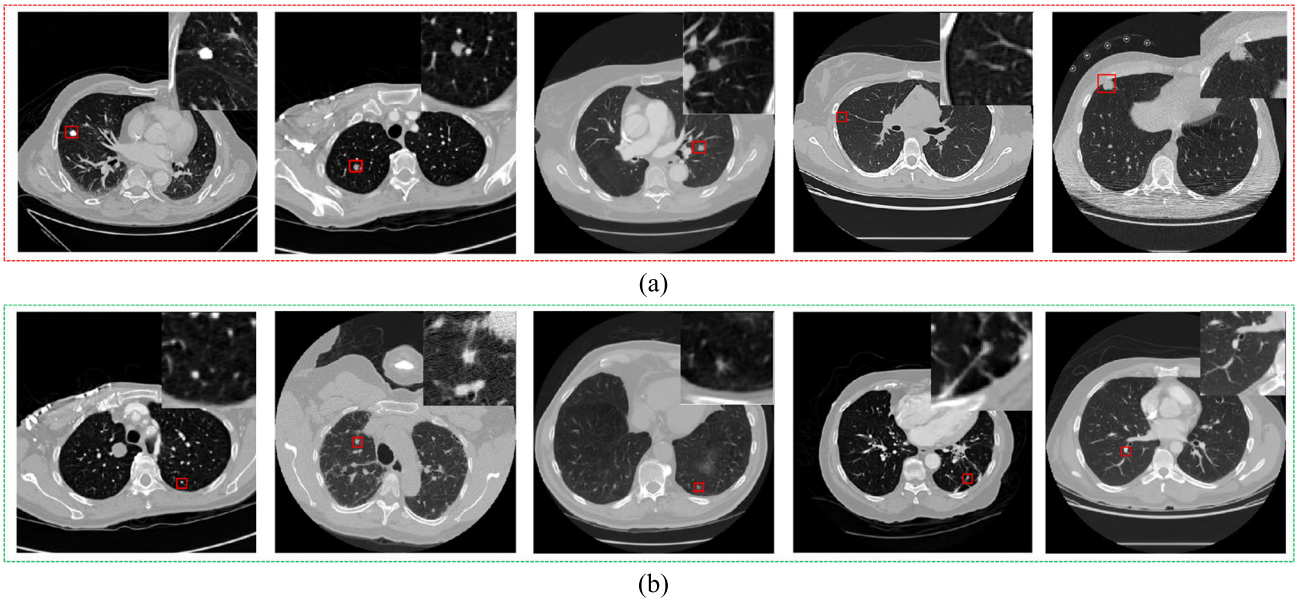


FIGURE 1. Examples of the nodules with various sizes, shapes, and locations in LUNA16 dataset. (a) True positive pulmonary nodules. (b) False-positive pulmonary nodules.

glass opacity nodules, small nodule and juxta-pleural nodules as shown in Fig.1(a). Some false- positive nodules are similar in appearance to true positive nodules, which can easily be mistaken for suspicious nodules.

Due to different appearance, type, malignancy, size, internal structure and location of pulmonary nodules, it is a great challenge to design a CAD system involving different levels to efficiently detect candidate nodules. Zhang *et al.* [6] detected candidate nodules by using multi-scale Laplacian of Gaussian (LoG) filters and densely dilated 3-D deep convolutional neural network (DCNN) to classify candidate nodules, but it contained more false-positive nodules in the nodule candidate detection. Zhu *et al.* [7] used 3-D Faster R-CNN and dual-path network (DPN) to improve pulmonary nodule detection, but the network was complex and the average false-positive rate was high. Gong *et al.* [8] combined residual and squeeze-and-excitation (SE) module to improve the detection effect. However, the convolution kernel has fixed 3*3*3 size, which only extract the features with fixed receptive field and single feature information. Aim to these thorn problems, the fixed 3*3*3 convolution kernel was replaced by the dilated convolution combination with different dilation rates, which reduced the complexity of DPN and extracted rich spatial context features with different dimensions and receptive fields, and the key feature information was preserved well. Furthermore, considering that the extracted features have redundant interference and the key features are not significant, the channel attention mechanism is used for adaptive adjustment to further suppress redundant information and enhance the key features. And then more representative nodal features containing rich context information can be obtained.

Zhai *et al.* [9] designed a multi-task classification network using nine 2D views to reduce false positives for pulmonary nodules. However, the 3-D spatial contextual feature had not been fully utilized in CT images. Dou *et al.* [10] used three 3-D input branches with different sizes and receptive fields to separately train and fused them for false positives reduction, resulting in too many network parameters, which is not conducive to network convergence. Hence, Zuo *et al.* [11] used single-branch feature layers with different depths and different receptive fields to reduce false positives. It reduced the number of parameters. However, it did not optimize the feature extraction network, and may lack the microscopic features of the nodules. Aim to these thorn problems, 3-D CNN is used to fully extract the spatial context features of nodules. Combining CT image reconstruction task and nodule classification task, a multi-task learning method was designed to fully learn the microscopic features of image spatial context. Moreover, the number of parameters is greatly reduced because the different depths of the single branch network are utilized.

To sum up, an efficient 3-D computer-aided system is designed to extract the representative pulmonary nodules features for detection. The main contributions are summarized as follows: (1) In order to fully extract the abundant three-dimensional spatial features of pulmonary nodules, 3-D context-guided module (3-D CGM) containing low-dimensional and high-dimensional local information and spatial context information is designed in the detection model based on 3-D Residual U-Net. (2) Channel attention mechanism is applied to further enhance the channel correlation between the feature maps for the representative features extraction. (3) Multi-branch classification network

of multi-task learning is designed to classify candidate nodules, and the microscopic feature information of shared feature network is adjusted through image reconstruction task. (4) The feature maps with different depths and levels in shared feature network are applied for feature fusion to improve the recognition effect.

In the following, related work and our method are described in Section II and Section III, respectively. Section IV reports the experimental results. Some key issues are discussed in Section V. Conclusions are drawn in Section VI.

II. RELATED WORK

Lots of research works are devoted to developing an effective and robust classified method for detecting lung nodules. They are generally divided into traditional machine learning method and deep learning method (two-dimensional (2D) or three-dimensional (3-D) CNNs).

The traditional CAD methods mainly apply machine learning algorithm to manually design contour features, shape features and texture features of pulmonary nodules, and then input them into the classifier for classification. Lassen *et al.* [12] used the region growing method to detect the candidate nodules, and selected the regional growth points of the initial seeds by artificial labeling. However, some nodules are indistinguishable from the surrounding tissues because of the variety of nodule types. Zhang *et al.* [13] proposed a method based on rule and Support Vector Machine (SVM). The shape features of Region of Interest (ROI) were firstly calculated, and then vessels were removed using a rule-based method. Next, grayscale and texture features in chest CT scans were calculated. Finally, the shape, grayscale and texture features were input to SVM classifier for nodule candidate detection. Gong *et al.* [14] designed a candidate nodule detection algorithm based on 3-D tensor filtering and local feature analysis. Contour correction is performed by 3-D horizontal segmentation method, and then a random forest classifier and 19 image features are used to reduce false positives. Farahani *et al.* [15] proposed an algorithm to integrate three classifiers including multi-layer perceptron, k-nearest neighbor and SVM. Firstly, the whole lung was segmented from CT image, and the shape features such as roundness, ellipticity and eccentricity were calculated, which were input into three classifiers for lung diagnosis. Arulmurugan and Anandakumar [16] performed computer-aided diagnosis of pulmonary nodules based on wavelet features and artificial neural network classifiers. Javaid *et al.* [17] first broke the connection between nodule candidates and vessels or thorax, and then used the k-means algorithm to cluster the nodule candidates, and finally extracted visual features, including the weighted centroid distance and mean intensity difference, to train an SVM classifier with the RBF kernel. These methods achieved relatively good results in the early detection of lung nodules, and provide an idea for subsequent computer-aided diagnosis. Research shows that traditional pulmonary nodules detection involves in a number of complex preconditioning operations. Moreover, manual feature extraction often requires some medical background. Hence, these hand-crafted features are affected by limited representation capability and are insufficient to deal with the large variations of lung nodules.

With the emergence of a large number of annotated data from medical images and the rapid improvement of computer computing ability, deep learning technology has developed rapidly in medical lesion detection system. Poongodi *et al.* [18] used RNN and CNN to diagnose the coronavirus rapidly and accurately. Roth *et al.* [19] used deep learning-based 2DCNN technology to detect lung nodules. First, scale transformation, random translation and rotation sampling are performed, and then ConvNet is used to extract expression features to reduce false positives. It belongs to the early application of combining 2D CNN with medical imaging. However, due to the limited performance of ConvNet, the effect is not optimal. With the development of pattern recognition and artificial intelligence, many 2D CNN (Faster-RCNN) models have emerged. Ding *et al.* [20] used 2D Faster R-CNN and VGG-16 to detect candidate nodules. Talha *et al.* [21] applied adaptive thresholding technique (OTSU) and 2DCNN semantic segmentation to accurately detect the lung nodule, and the selected optimal features were applied to 9 classifiers. Xie *et al.* [22] designed an enhanced 2D CNN architecture to reduce false positives, and the sensitivity was 73.4% and 74.4% for 1/8 and 1/4 false positives of per scan, respectively. Albahli *et al.* [23] applied GAN to generate synthetic data for training the data as the amount of the data is limited. And then different 2DCNN models were used to diagnose the cardiothoracic diseases. Albahli *et al.* [24] provided a synthetic data augmentation in three 2D CNNs (DenseNet121, InceptionResNetV2, and ResNet152V2) architectures for the detection of 14 chest-related diseases. After training and validation, an average ROC-AUC score of 0.80 was obtained. Setio *et al.* [25] designed the input to include not only the axial, coronal, and sagittal views of the lung nodules, but also six views of diagonals.

Compared with traditional methods, the detection effect of 2DCNN has been greatly improved. However, detecting pulmonary nodules from volumetric CT scans is a 3-D object detection problem in essence. Therefore, even if multiple 2D CT slices are used, it is still difficult to make full use of the 3-D spatial contextual information of pulmonary nodules. Dou *et al.* [10] proposed a multi-level context-based 3-D CNN to detect pulmonary nodules, three 3-D CNN with different receptive fields were constructed to obtain multi-level contextual feature information, and then three scales detection results were integrated together to improve the classified results. Compared with the 2D methods, it considers the three-dimensional spatial context characteristics of medical images, and the result has been greatly improved. It has great reference value in clinical and academic field, and also provides a new idea for the follow-up researchers. For example, Huang *et al.* [26] used a filter based on the

local set model to generate candidate nodules, and classified them by 3-D CNN. Hamidian *et al.* [27] used a 3-D fully convolutional network (FCN) to generate a score map for nodule candidate identification, and employed another 3-D CNN for nodule and non-nodule discrimination. Therefore, Cao *et al.* [28] proposed a three-dimensional convolutional neural network based on a multi-branch set learning algorithm for pulmonary nodule detection, three branching structures based on VggNet, IResNet and DenseNet are constructed to correspond to three input sizes. In order to make full use of 3-D spatial context feature information, 3-D CNN was used in these branching structures to extract different deep features of pulmonary nodules, and then three classification results were integrated together to improve the detection results. Pezeshk *et al.* [29] trained a three-dimensional fully convolutional neural network for pulmonary nodule detection, and the detection sensitivity reached 80%. A comprehensive literature review of related works is seen in Table 1.

With the development of 3-D CNN technology, the application of 3-D CNN to detect lung nodules can extract the spatial context features of lung nodules. However, most of the networks (ResNet, WideResNet, DenseNet, etc.) applied in CADe schemes only improve networks performance by changing spatial dimensions of networks. Therefore, some redundant features of nodules will be extracted in the feature information transmission, and some key microscopic features will be lost. And some attention guided encoder-decoder network does not consider the extraction of spatial context features with multiple scales and different receptive fields. It may be missed or misjudged in the detection and classification. In this case, it leads to the lack of representativeness of the extracted features, which affects the sensitivity of pulmonary nodules detection.

In summary, related works have certain limitations in the diagnosis of pulmonary nodules. Considering that 2D data represents a slice of CT, while 3-D data is a combination of multiple 2D slices with continuity. Even if multiple slices are used for 2D data, there is no continuity between each slice, and the spatial context characteristics are not fully utilized. However, multiple slices of 3-D data have continuity and spatial context information. By extracting the spatial context features, the detection sensitivity can be improved. Therefore, the 3-D CNN is optimized to extract the representative features. Therefore, an efficient 3-D CAD system combined 3-D Residual U-net for context-guided attention module with multi-branch network of multi-task learning is designed and fully validated on LUNA16 dataset.

III. PROPOSED CAD SYSTEM

The appearance of pulmonary nodules on chest CT scans is often confused with blood vessels, lymph nodes, bronchus and other pathological tissues, and the surrounding texture environment is complex. The above factors lead to the huge amount of CT image information, and manual judgment may lead to the loss of important details. Hence, the overall framework of 3-D pulmonary nodules CAD system is proposed in Fig. 2, which aims to fully extract feature information and accurately detect pulmonary nodules.

The proposed 3-D pulmonary nodules CAD system is demonstrated as follows:

Datasets Generation: In order to prevent other tissues in CT images from interfering with the subsequent detection, it is necessary to pre-process the original lung CT images and its masks to segment the lung parenchyma. The pre-processing operation mainly includes normalization, morphological operation and resampling with a new resolution of $1 \times 1 \times 1$ [6], [8]. The reason for resampling is that the dataset comes from different equipment in different hospital, and the voxel spacing, resolution and initial point coordinates of each CT are different. In this case, the voxel spacing information cannot be learned by CNN, which is not conductive the training. The resampling performed Bilinear Interpolation method, and the result is smoother. It is observed that after resampling interpolation, the number of slices is increased, and the image quality is not significantly affected. (It is acceptable.) In addition, the annotation coordinates are resampled with the same resolution, and the pre-processed CT images are randomly cropped to a fixed size according to the new coordinates to construct 3-D cubic patches.

Nodule Candidate Detection: Due to heterogeneity and complex morphology of pulmonary nodules, it is difficult to fully extract three-dimensional spatial features. Hence, a 3-D Residual U-Net network is improved to detect candidate nodules. The network takes 3-D Residual U-Net as its backbone, and combines original residual, 3-D context-guided module with different dilation rates and channel attention mechanism. The nodule features extracted contains three-dimensional spatial information with different receptive fields, and more representative nodule features are extracted by suppressing the useless channel features and enhancing the useful channel features.

False-Positive Reduction: In order to reduce false positives in pulmonary nodule detection, a multi-branch classification network of multi-task learning is constructed. Here, nodule features are extracted through the shared feature network, and the detailed texture features of nodules are fine-tuned through the image reconstruction network to achieve abundant feature information. Then the multilevel nodule features with different receptive filed sizes are loaded into multi-branch classification network for feature fusion and nodule classification.

A. DETECTION NETWORK BASED ON 3-D RESIDUAL U-NET ORIENTED CONTEXT-GUIDED ATTENTION

To improve the sensitivity, a detection network based on 3-D Res-Unet oriented context-guided attention is designed.

1) NODULE CANDIDATE DETECTION FRAMEWORK

Considering that pulmonary nodule detection in volumetric CT scans is a 3-D object detection problem, the 3-D Residual U-Net detection network is improved to fully extract the 3-D spatial feature information. If the whole CT image is input to the detection network, GPU memory cannot meet

TABLE 1. Description of some related work and pros and cons of existing approaches.

Ref	Method	Gaps identified
Farzad et al. [15]	An algorithm is proposed to integrate three classifiers including multi-layer perceptron, k-nearest neighbour and SVM.	Only handcrafted extractors are explored to extracted features for limit machine learning.
Setio et al. [24]	A multi-view CNN is constructed to extract hierarchical features from nine 2D slices with different angles of view and group the high-level features for classification.	3D spatial contextual information has not been fully utilized, so limited classification performance in low-FP scans is achieved.
Zhu et al. [7]	A 3D dual path network (DPN) is proposed as a deep feature extractor and a gradient boosting machine (GBM)classifier for LIDC/IDRI.	The transfer learning concept did not explore, so training depth structure not allowed. Additionally, only one CNN architecture is explored as feature extractor and machine learning classifier.
Gong et al. [8]	A 3D automatic nodule detection system is designed by combing residual and squeeze-and-excitation (SE) module to improve the detection effect.	A large 3D size is designed to input the network, which requires a high GPU memory.
Zhai et al. [9]	A multi-task classification network is designed to reduce false positives for pulmonary nodules.	Nine 2D slices were used to reduce false positives for pulmonary nodules, but the spatial context features were not fully utilized in CT images
Qi Dou et al [10]	A multi-level context-based 3D CNN is proposed to detect pulmonary nodules, and three scales detection results were integrated together to improve the classified results.	The weights for each scale were determined manually, rather than learning from training samples.
Saleh Albahli et al [24]	Synthetic data enhancement in three deep convolutional neural network (CNN) architectures (DenseNet121, InceptionResNetV2, and ResNet152V2) is proposed to detect 14 chest related diseases.	All the images used for both training and validation are images of the chest's frontal view.
Raul Victor M et al [30]	Several CNNS (VGG16, VGG19, MobileNet) are used for transfer learning, and each set of depth features with Bayes, KNN, MLP, RF, SVM linear and SVM RBF machine learning techniques is combined.	Deep transfer learning method is proved as a relevant strategy to extract representative features from lung nodule CT images.
Dawid Poap et al [31]	A heuristic method for detection aggregated x-ray images is proposed.	The method helps doctors to concentrate on special, selected areas in x-ray images, but does not consider deep learning methods.
Giacomo Capizzi et al [32]	An evaluation model based on a composition of fuzzy system combined with a neural network is proposed.	The method greatly reduces the computational demands and increases the detection performances.
Ahmed T. Sahlol et al [33]	The author MobileNet is adopted to previously train on the ImageNet for the feature's extraction. and AEO algorithm was used for feature selection	CNN model avoided the redundancy of extraction features and improved the classification accuracy. But other CNN models are not involved in experiment comparisons.

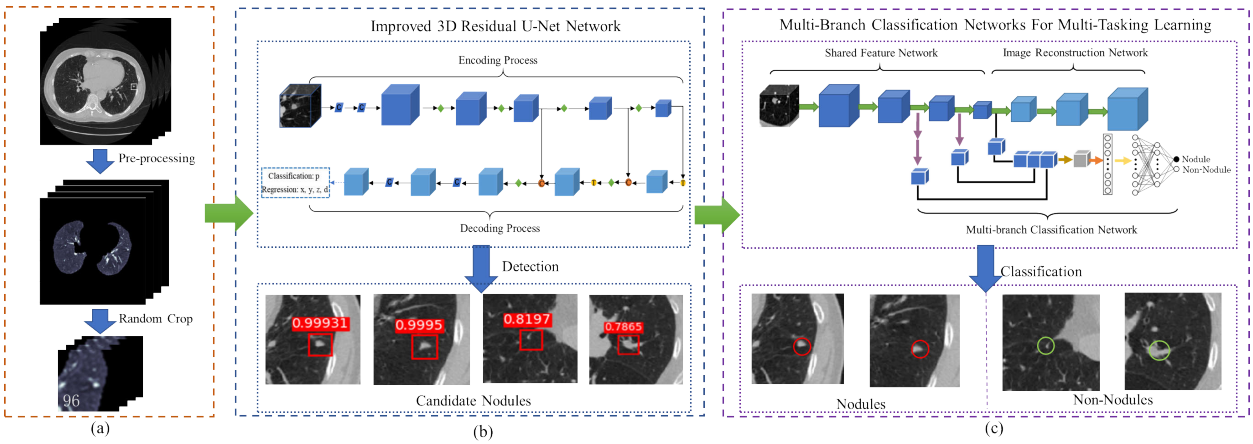


FIGURE 2. The framework of 3-D pulmonary nodules computer-aided detection system. (a) Datasets Generation. (b) Nodule Candidate Detection. (c) False-positive reduction.

such a huge data bandwidth. Hence, the pre-processed CT image is cropped into 3-D input patches with a fixed size of $96 \times 96 \times 96$, and then sent to the detection network shown in Fig. 3.

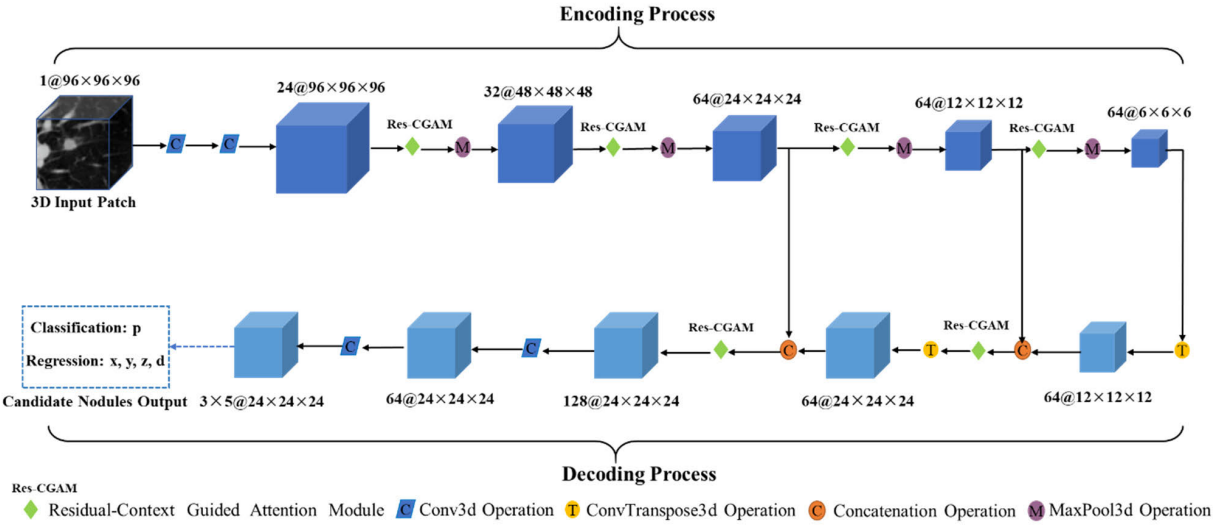


FIGURE 3. The improved 3-D Residual U-Net network structure. Each cube represents a 4D tensor, where the number outside the cube represents the number of channels and spatial size.

The nodule candidate detection network contains the encoding and decoding processes described as follows:

Encoding Process: A 3-D Input Patch is used to roughly extract three-dimensional space features and location information of pulmonary nodules through two convolutional layers (each contains 24 convolutional kernels of size $3 \times 3 \times 3$). Four 3-D Residual-Context Guided Attention Modules (3-D Res-CGAM) and four max-pooling layers make a cross combination. The former aims to fully extract three-dimensional feature information of pulmonary nodules, while the latter aims at feature reduction and network simplification. Then, the feature maps (the number of channels is 64 and the size is $6 \times 6 \times 6$) containing nodule rich abstract feature information is obtained.

Decoding Process: The feature maps obtained from the encoding path is upgraded dimension for the first time by using convtransose3d operation (64 deconvolution kernels of size $2 \times 2 \times 2$), and these feature maps are concatenated with the corresponding feature maps with the same dimension from the encoding process. Then, the concatenated feature maps extract the deep-level feature information through 3-D Res-CGAM. Furthermore, these features are upgraded dimension for the second time by convtransose3d operation, and then a same secondary concatenated operation is performed. Obviously, the concatenation operation effectively complements the feature information lost in the feature extraction process, which is beneficial to restore the detail features of pulmonary nodules.

Nodule features are further extracted by 3-D Res-CGAM, and feature maps (the number of channels is 128 and the volume size is $24 \times 24 \times 24$) containing spatial feature and location information of pulmonary nodules are obtained. Next, two convolution layers (64 convolutional kernels of size $1 \times 1 \times 1$ and 3×5 convolutional kernels of size $1 \times 1 \times 1$) are used to change the dimension of the obtained feature maps (the number of channels is 3×5 and the volume size is $24 \times 24 \times 24$), so that Region Proposal Network (RPN) operation is applied to realize the regression of boundary box coordinates and the prediction of nodular probability. In accordance with the size distribution of nodules, three anchors with the sizes of 5, 10 and 20mm are applied to the final feature mapping to generate a set of object proposals, each contains a 5×1 vector (central coordinates (x, y, z), diameter d and object score p). The coordinates and the probabilities of pulmonary nodules can be predicted by testing the trained network model. It is worth noting that a novel 3-D Res-CGAM structure is proposed, which will be described in detail.

2) 3-D RES-CGAM ARCHITECTURE

ResNet, as shown in Fig. 4(a), is a common CNN network structure, it introduces the skip connection channels to realize the feature reuse and solve the network degradation. However, there are big differences in appearance, type and size of pulmonary nodules and the complex surrounding texture environment. It is difficult to simultaneously extract the local and global features, and the model depth is too shallow to make full use of the semantic features. To address this issue, 3-D Res-CGAM structure is designed based on ResNet, as shown in Fig. 4(b). The original residual 3-D convolution is replaced by 3-D context-guided module (3-D CGM), which integrates the low-dimensional and high-dimensional local features and global features to extract the rich 3-D spatial features information of nodules. The attention channel mechanism is introduced into the model to adjust adaptively each channel feature of the feature maps, and the correlation between feature channels are modelled explicitly. Therefore, more expressive feature maps can be obtained.

It is noted that 3-D Res-CGAM structure integrates 3-D CGM structure and channel Attention mechanism module presented as follows:

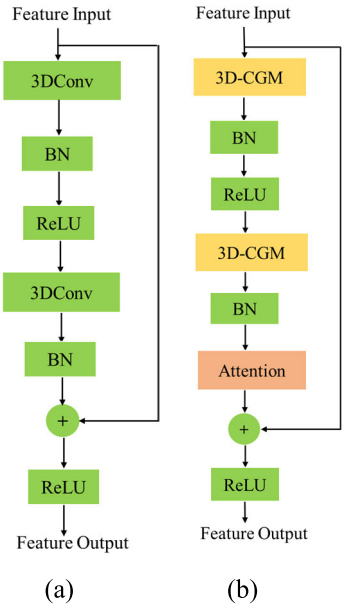


FIGURE 4. Structure diagram of improved residual block. (a) 3-D Residual module. (b) 3-D Res-CGAM.

a: 3-D CGM STRUCTURE

Pulmonary nodules often have different sizes and irregular shapes, and 3 × 3 × 3 convolution kernel is often used to extract the features information. However, it is difficult to capture the global features of pulmonary nodules with a larger diameter by the limited receptive field of 3b × 3 × 3 convolutional kernels. And the failure to make full use of the three-dimensional spatial context information of nodules may lead to the omission of pulmonary nodules detection. To address the above problems, 3-D CGM is designed by introducing the dilated convolution with different dilation rates into 3 × 3 × 3 convolution kernel, as shown in Fig. 5.

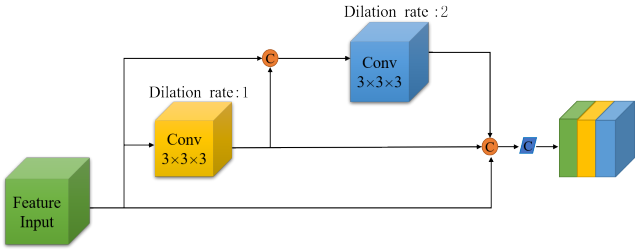


FIGURE 5. Structure of 3-D context-guide module (3-D CGM).

In Fig. 5, the *Feature Input* represents the input feature map, which includes low-dimensional local features (texture, space and position) information. When the Feature Input is convolved by a 3 × 3 × 3 dilated convolution with a dilation rate of 1, we will get the relative input features, which are high-dimensional local features containing more semantic information of pulmonary nodules. After high-dimensional and low-dimensional local features are concatenated together, and then they are convoluted with 3 × 3 × 3 dilated convolution kernel with a dilation rate of 2 to achieve the spatial context global features. Finally, low-dimensional and high-dimensional local features and global features are concatenated together, and then they are convoluted with 1 × 1 × 1 convolution kernel to obtain the feature maps containing rich contextual features and multi-scale information.

Obviously, when the size and parameter of convolution kernel remain unchanged, the 3-D CGM structure appropriately increases the receptive field by reasonably setting the dilation rates of the dilated convolution, captures the multi-scale context information of pulmonary nodules, and extracts multi-scale feature layers including local and global features, which plays a role of context guidance.

b: CHANNEL ATTENTION MECHANISM

Channel attention mechanism shown in Fig. 6 includes Squeeze and Excitation operation for adaptive recalibration.

Squeeze Operation: The global average pooling operation is used to compress the features of the convolution feature map after residual operation, and then the global features (1 × 1 × 1 × C) on the feature channel are output. Since the channel numbers of the output feature matches that of the input feature, the global receptive field can be obtained on the feature layer close to the input. The process is descried in formula (1) [8], where x_{res} represents the features after residual operation, F_{sq} represents the squeeze function that aggregates global spatial information into channel-wise statistics by global average pooling, L × H × W is the spatial dimensions of x_{res}, and z_c is the output features after global pooling operation.

$$z_c = F_{sq}(x_{res}) = \frac{1}{L \times H \times W} \sum_{i=1}^{L} \sum_{j=1}^{H} \sum_{k=1}^{W} x_{res}(i, j, k) \quad (1)$$

Excitation Operation: The channel global features is resulted by the squeeze operation, and then they are performed excitation operation. The first layer of FC structure

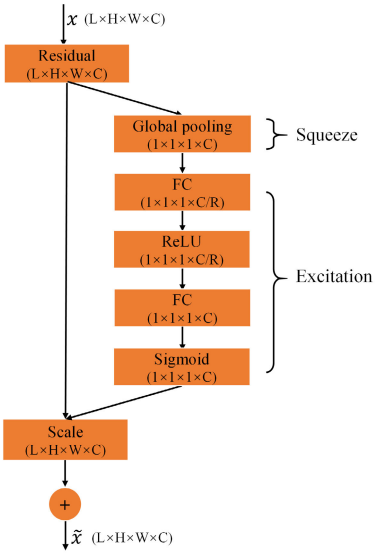


FIGURE 6. Structure of Channel attention mechanism.

is used for dimension reduction, which reduces the calculation amount by decreasing the number of channels. The hyperparameter R is set to 16, and ReLU function is used. The second layer of FC structure is designed to restore the original dimension, and sigmoid function is adopted to obtain the activation vectors ($1 \times 1 \times 1 \times C$) of each channel. Then, the activation vectors are multiplied by the original feature maps to obtain the new feature maps with more distinguishing ability for each channel feature. The process is descried in formula (2) [8], where F_{ex} is the excitation function to generate the scales of the feature channels, w_1 ($w_1 \in R^{\frac{c}{R} \times c}$) and w_2 ($w_2 \in R^{c \times \frac{c}{R}}$) represent the training weight parameters, and σ represents the sigmoid function.

$$s = F_{ex}(z, w) = \sigma(g(z, w)) = \sigma(w_2 ReLU(w_1 z)) \quad (2)$$

Benefiting from the feature recalibration strategy of two above operations, the channel attention mechanism focuses the model on the relationship between the feature channels rather than the relationship in spatial position. The trained network automatically learns the feature weights on different channel, so as to enhance the channel features with the most critical information and weaken the channel features with more interference information. By introducing channel attention mechanism into the model, the feature expression ability is improved, which is conducive to detect pulmonary nodules from the complex background environment. The details of each structure are shown in Table 2.

3) LOSS FUNCTION OF NODULE CANDIDATE DETECTION

Considering that the detection performance of RPN operation combined classification task with regression task is far superior to that only using one of these two tasks [34]. During nodule candidate detection, the loss function adopted by the whole task includes classification loss L_{cls} and regression loss L_{reg}. The loss function of the whole task is formulated as:

$$L(p, t) = \lambda L_{cls}(p, p^*) + p^* L_{reg}(t_i, t_i^*) \quad (3)$$

where λ is a hyperparameter to balances classification loss and regression loss, and it is set to 0.5. The p and p^* represent the predicted probability and the true label of anchor, respectively. If Intersection over Union (IoU) between anchor and ground truth bounding box is greater than 0.5, it is considered as a positive sample ($p^* = 1$). If IoU is less than 0.2, it is considered as a negative sample ($p^* = 0$). The $t_i(x, y, z, d)$ and $t_i^*(x^*, y^*, z^*, d^*)$ represent the coordinate vector of the predicted and the true bounding box of the nodule position, respectively.

The classification loss L_{cls} is binary cross-entropy loss function, which can be formulated as:

$$L_{cls}(p, p^*) = -w\left[p^* logp + (1 - p^*)\log(1 - p)\right] \quad (4)$$

where w represents the weight.

The regression loss L_{reg} uses the smooth L1 loss function to calculate the position information of the regression, which can be formulated as:

$$L_{reg}(t_i, t_i^*) = \sum_{i \in \{x,y,z,d\}} smooth_{L1}(t_i - t_i^*) \quad (5)$$

$$smooth_{L1}(t_i - t_i^*) = \begin{cases} 0.5(t_i - t_i^*)^2, & if \left|t_i - t_i^*\right| < 1 \\ \left|t_i - t_i^*\right| - 0.5, & otherwise \end{cases} \quad (6)$$

B. MULTI-SCALE FEATURE FUSION CLASSIFICATION NETWORK FOR MULTI-TASK LEARNING

To improve the comprehensive performance of classification network, a novel multi-scale feature fusion classification network for multi-task learning is designed.

1) NODULE CANDIDATE CLASSIFICATION FRAMEWORK

Considering that massive false-positive nodules are generated after nodule candidate detection, it is necessary to classify true positive and false-positive nodules. The classification accuracy is limited by two factors: (1) Due to pulmonary nodules vary in size and shape, and true positive and false-positive nodules have similar appearance, it is inevitable to loss some key microscopic detail features during the feature extraction. (2) The nodules feature information extracted by using single receptive field is very limited, which leads to low classification accuracy. In view of the above considerations, a multi-branch 3-D classification network of multi-task learning is designed to classify candidate nodules. The network structure is shown in Fig.7, which consists of three functional networks. Multi-branch classification (main task) and image reconstruction (auxiliary task) work together to perform multi-task learning.

Shared feature network is responsible for three-dimensional features extraction of pulmonary nodules. Its input feature map is the 3-D data (the volume size is $48 \times 48 \times 48$) cropped at the centre of the nodule candidate position. Due to the small number of positive samples, data augmentation is conducted to expand the dataset. Shared feature network consists of 15 layers: 8 convolution layers (the size of convolution kernel is $3 \times 3 \times 3$; the number of convolution kernels is 16, 16, 32, 32, 64, 64, 128, 128, respectively; each convolution layer performs convolution, BN, ReLU activation operations), 3 max-pooling layers (the size of max-pooling layers is $2 \times 2 \times 2$; the stride is 2), and 4 original residual structures.

Multi-branch classification network can fully integrate the feature information from different levels and receptive fields in the shared feature network to improve the classification accuracy. The network structure uses the feature maps of size 24 and 12 in shared feature network to obtain two feature maps of size 6 with different receptive field information through two and one max-pooling operation respectively. Then, they are concatenated with the feature maps of size 6 in the shared feature network, and the multi-scale feature information is fused through $1 \times 1 \times 1$ convolution operation to obtain the nodule feature layers with three receptive fields. Finally, global max-pooling and two-layer full connection

TABLE 2. Different structures of candidate nodule detection (SE and 3-D CGM structure are shown in Table 3 and Table 4 respectively.).

Structures	Layer	Type	Kernel Size	Number of filters
Res	1	Input	-	-
	2	Conv+BN+ReLU	3×3×3	32
	3	Conv+BN	3×3×3	32
	4	Input+Layer3	-	-
	5	ReLU	-	-
Res+3D-CGM	1	Input	-	-
	2	**3D-CGM**	3×3×3	32
	3	Conv+BN+ReLU	1×1×1	32
	4	**3D-CGM**	3×3×3	32
	5	Conv+BN	1×1×1	32
	6	Input+Layer5	-	-
	7	ReLU	-	-
Res+Attention	1	Input	-	-
	2	Conv+BN+ReLU	3×3×3	32
	3	Conv+ BN	3×3×3	32
	4	**SE**	-	32
	5	Layer3*Layer4	-	-
	6	Layer5+Input	-	-
	7	ReLU	-	-
3D Res-CGAM	1	Input	-	-
	2	**3D-CGM**	3×3×3	32
	3	Conv+BN+ReLU	1×1×1	32
	4	**3D-CGM**	3×3×3	32
	5	Conv+ BN	1×1×1	32
	6	**SE**	-	32
	7	Layer4*Layer5	-	-
	8	Layer6+Input	-	-
	9	ReLU	-	-

TABLE 3. Structure of SE.

Structure	Layer	Type	Kernel Size	Number of filters
SE	1	GlobalAvgPool	-	-
	2	Conv+ ReLU	1×1×1	32/16=2
	3	Conv	1×1×1	32
	4	Sigmoid	-	-

TABLE 4. Structure of 3-D CGM.

Structure	Layer	Type	Kernel Size	Dilation rate
3D-CGM	1	Input	-	-
	2	Conv+ BN+ReLU	3×3×3	1
	3	Concat(Input, Layer2)	-	-
	4	Conv+ BN+ReLU	3×3×3	2
	5	Concat(Layer3,Layer)	-	-

operation are performed to obtain the classification probability of pulmonary nodules.

Image reconstruction network (decoder) can recover the key microscopic feature of pulmonary nodules in the shared feature network (encoder) as much as possible, which assists multi-branch network to improve classification performance. Image reconstruction network includes 3 deconvolution layers (the size of deconvolution kernel is $2 \times 2 \times 2$), 6 convolution layers (the first five layers with $3 \times 3 \times 3$ convolution kernels; the last one layers with $1 \times 1 \times 1$ convolution kernel) and 2 original residual structures.

2) LOSS FUNCTION OF PULMONARY NODULE CLASSIFICATION

Considering that two-branch network for nodule classification task and image reconstruction task are simultaneously trained during false-positive reduction, the total loss function of multi-branch network of multi-task learning is formulated as:

$$L_{total} = L_{cls} + \alpha L_{ImRe} \tag{7}$$

Here, the classification loss and image reconstruction loss are defined as L_{cls} and L_{ImRe}, respectively. The threshold $\alpha \in (0, 1)$ represents a trade-off between the two tasks to avoid the degradation of classification ability due to excessive emphasis on image reconstruction task. After many experiments, the α is set to 0.4. Actually, $\alpha = 0$ means that there is only classification task.

In order to optimize the classification performance of pulmonary nodules, the model is evaluated by cross-entropy

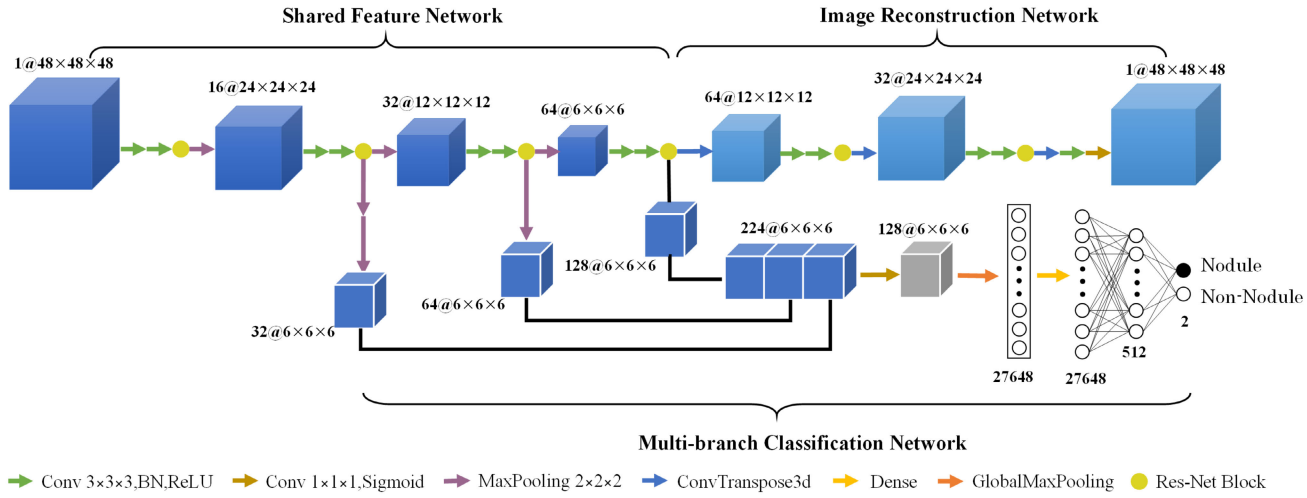


FIGURE 7. Structure of multi-branch 3-D classification network based on multi-task learning.

loss function [9] described as:

$$L_{cls} = L\left(y_n, \hat{y}_n\right) = y_n \log \hat{y}_n + (1 - y_n) \log\left(1 - \hat{y}_n\right) \quad (8)$$

where y_n is true label of the n_{th} sample, $y_n = 1$ represents the positive nodules, $y_n = 0$ represents the negative nodules, and $\hat{y}_n$ is the prediction probability of the positive nodules.

The sum of cross-entropy loss of each voxel is calculated in image reconstruction, and its average value is taken as the loss function L_{ImRe} formulated as follows:

$$L_{ImRe} = \frac{1}{N} \sum_{n=0}^{N} L\left(x_n, \hat{x}_n\right) \quad (9)$$

where L represents the cross-entropy loss function, and N represents the number of voxels of the input image with a size of 48 × 48 × 48. The x_n and $\hat{x}_n$ represent original voxel value and predicted voxel value of input image, respectively.

IV. EXPERIMENTS AND RESULTS

Numerous experiments were performed to verify the proposed 3-D computer-aided pulmonary nodule detection system, and the experimental conditions were set to Ubuntu 16.04 operating system and two NVIDIA GeForce GTX 2080Ti GPUs.

A. EXPERIMENTAL DATASET GENERATION

1) DATASET CHOSEN

Lung Nodule Analysis 2016 (LUNA16) dataset involved in this experiment is extracted from Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) database, which is primarily for large-scale algorithms evaluation competition in pulmonary nodule detection and false-positive reduction. It retains 888 chest CT scans including 1186 positive pulmonary nodules. The nodule annotations were manually marked by 4 experienced radiologists:(1) Each physician independently reviewed the CT scan and marked the lesions with non-nodules, nodules < 3mm, and nodules >= 3mm. (2) The nodule annotations from all four doctors were reviewed independently, and each radiologist decided to accept or reject each annotated information. The LUNA16 dataset included CT scans with section thickness <= 2.5mm, pulmonary nodules with diameter >= 3mm and annotated by at least 3 of 4 radiologists.

2) DATA PRE-PROCESSING

LUNA16 dataset are pre-processed to improve the detection performance of pulmonary nodules: (1) Load raw CT images and masks from LUNA16 dataset. (2) Boundary box of lung mask is obtained, and its resolution is resampled to 1 × 1 × 1. (3) Lung region is reserved with Hu value range of [−1200, 600] and normalized to range of [0,255]. (4) Lung masks are dilated to remove small holes in the lung, and & operation is applied to the new lung mask and original CT images. (5) CT images is resampled and the data within boundary box is intercepted, then the labels are converted to voxel coordinates and transformed with a new resolution of 1 × 1 × 1, and the pre-processed data and labels are stored in files with .npy extension. Table 5 includes the dataset description of before and after of pre-processing.

TABLE 5. Dataset description.

Dataset	Number of images	Volumes (G)	Nodule	Instance	Format
Original	216672	119.1	1186	888CT	Raw/mhd
Pre-processing	248412	75.2	1186	888CT	Npy

B. EXPERIMENTAL CONDITIONAL SETTING

When it comes to the selection and optimization of user-defined parameters, these works achieved excellent detection performance by setting reasonable hyperparameters [7], [8], [22]. On this basis, we adjusted the user-defined model parameters through numerous experiments presented as follows.

1) NODULE CANDIDATE DETECTION

LUNA16 competition dataset was equally split into 10 subsets. 10-fold cross-validation were performed to evaluate nodule candidate detection performance. For each fold, 90% of the dataset for training and 10% of the dataset for testing, and the average of 10-fold cross-validation results was taken as the detection results of nodule candidate. During the training, limited by memory resources of GPU, the large-size CT images were cut into 96 × 96 × 96 3-D cubes to match the data bandwidth, and the batch size was set to 8. In order to alleviate the overfitting problem caused by the model, data augmentation was performed on cubic patches of positive samples: (1) Rescaling in the range [0.75, 1.25]; (2) Left-right flipping and random rotating. In the experiment, stochastic gradient descent (SGD) was selected as the model optimizer, and the model was trained for 150 epochs. During the training experiment, the epoch was usually selected in the range of 100 to 200. When the epoch is less than 150, the loss of training and verification decreases rapidly. When the epoch exceeds 150, the loss tends to decrease gently, and excessive training time is consumed. However, ten-fold cross-validation effect at this time was equivalent to that of 150 epochs. To sum up, we make the train experiment stop at 150 epochs. The learning rate was initialized to 0.01, and it gradually decreased with the increase of training epochs. The learning rate was decreased down to 0.001 after 75 epochs and 0.0001 after 120 epochs, respectively. During the testing, a large number of nodule proposals would be detected and some of which overlapped each other. Therefore, the nodules less than 3mm detected diameter were removed, and then Non-Maximum Suppression (NMS) with IoU threshold of 0.1 was applied to remove the overlap proposals. Finally, the test results were achieved. In order to prevent network overfitting, data enhancement is carried out during the pre-processing stage. 10-fold cross-validation and dropout (P = 0.5) were used in the training process during the detection stage. The fine-tuned hyper-parameters settings in the experiment are shown in Table 6.

2) FALSE-POSITIVE REDUCTION

Taken the center of the candidate nodules as the coordinate, CT images were cut into 48× 48 × 48 3-D cubes and then randomly disrupted. The data proportion of training sets, testing sets and verification sets were divided into 8:1:1. Since the number of false- positive nodules was much larger than the true positive nodules in dataset, the data imbalance may lead to the model overfitting. Hence, 3-D cubic patches of including true positive nodules were rotated 90°, 180°, and 270° in the transverse plane for data augmentation. During the training, Adam stochastic optimization was applied to the backpropagation, the weights were randomly initialized from Gaussian distribution N (0, 0.012), ReLU was used as the activation function, and the batch size was set to 12. The learning rate was initialized to 0.1, and it was reduced by 10% every 3000 iterations. Batch normalization and dropout (rate = 0.5) were adopted to improve the generalization ability of model. When the model performance starts to decline on the verification set, the training process is stopped. In order to prevent network overfitting, data enhancement, early stop, dropout (P = 0.8) and L2 regularization were used in the training process at the classification stage. The fine-tuned hyper-parameters settings in the experiment are shown in Table 6. Furthermore, the classification networks' GPU Memory resource is 10907 M, and the training time is about 48h.

3) EVALUATION INDICATORS

Detection sensitivity and free receiver operating characteristic (FROC) score are used to evaluate the performance of the CAD system. Generally, nodule detection algorithms are applied for screening nodules on chest CT, and thus enable radiologists to focus only on positive cases. In this case, false-positive cases may be corrected by radiologists, and false-negative cases cannot be corrected. Therefore, in order to assist radiologists to diagnose, an excellent detection algorithm should be evaluated as lower false positive rate and higher sensitivity. In the LUNA16 competition, a candidate is regarded as a true nodule if it is located within the range of R (half the diameter of the annotated nodules) from the nodule center. The sensitivity represents the percentage of pulmonary nodules correctly identified by classification. It is evaluated by formula (10), where TP, FN and TP+FN are defined as the number of true positive nodules, false negative nodules and all detected nodules, respectively. In the LUNA16 competition, a candidate is regarded as a true nodule if it is located within the range of R (half the diameter of the annotated nodules) from the nodule center. The sensitivity represents the percentage of pulmonary nodules correctly identified by classification. It is evaluated by formula (10), where TP, FN and TP+FN are defined as the number of true positive nodules, false negative nodules and all detected nodules, respectively.

$$Sensitivity = \frac{TP}{TP + FN} \tag{10}$$

FROC curve reflects the decreasing trend of false-positive nodules, and the average sensitivity at the seven pre-defined false-positive rates (0.125, 0.25, 0.5, 1, 2, 4, 8 FPs/scan) reflects the competition performance metric (CPM). The higher the CPM score, the better the system performance.

In order to comprehensively evaluate the performance of the classification, the metrics commonly also used for classification include accuracy, precision, recall, specificity and F1-score, which can be expressed by the following formulas (11), (12), (13), (14) and (15), respectively.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{11}$$

$$Precision = \frac{TP}{TP + FP} \tag{12}$$

$$Recall = \frac{TP}{TP + FN} \tag{13}$$

TABLE 6. Setting of Hyper-parameters.

Networks	Training epoch	Optimization algorithm	Batch size	Learning rate decay strategy		
				Initial epochs	Epochs> N×epochs	Epochs>N×epochs
3D Res-CGAM	150	SGD	8	0.01	0.001 (N=1/2)	0.0001(N=4/5)
Multi-branch classification network	50	Adam	12	0.1	0.01(N=1/5)	0.001(N=2/5)

$$Specificity = \frac{TN}{TN + FP} \tag{14}$$

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{15}$$

C. EXPERIMENTAL RESULTS ANALYSIS

The CAD system is verified on the LUNA16 well, and the network performance is evaluated as follows.

1) NODULE CANDIDATE DETECTION

a: ABLATION STUDY

Under the same experimental conditions, 10-fold cross-validation comparative experiments are conducted on Res-3 × 3 × 3 structure, Res+Attention structure, Res+3D-CGM structure and 3-D Res-CGAM structure, respectively. As shown in Fig. 8, experimental results show that: (1) When 3 × 3 × 3 convolution is replaced by 3-D CGM in the original residual module, the sensitivity is improved in 10-fold cross-validation experiments. Obviously, 3-D CGM can extract more multi-scale nodule features with rich spatial context information. (2) By introducing the channel attention mechanism into the original residual module, the sensitivity curve becomes relatively flat and slightly increased in each-fold experiment.

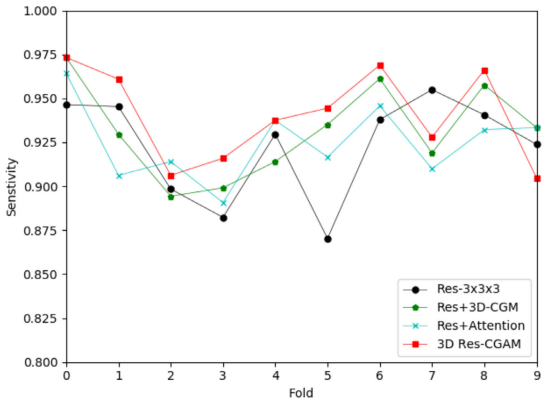


FIGURE 8. Comparison of 10-fold cross validation results from ablation study.

Obviously, the generalization ability of the whole model can be enhanced by introducing channel attention mechanism under different distributed test dataset. (3) When 3-D CGM and channel attention mechanism are both introduced into original residual module, the sensitivity curve was the flattest in each-fold experiment and the overall detection performance is optimal.

The detected results of four detection structures are compared in Table 7, which shows that: (1) Compared with Res-3 × 3 × 3, the average sensitivity of Res+3D-CGM increases from 92.2% to 93.2%. However, the average number of candidates per scan significantly increases from 31.52 to 59.60. It indicates that 3-D CGM enhances the feature extraction capability of the model, and the multi-scale features of pulmonary nodules are fully utilized, but a large number of redundant features are extracted, which reduces the detection efficiency. (2) To address the above thorn problem, the channel attention mechanism module is further introduced to Res+3D-CGM structure. The average sensitivity of 3-D Res-CGAM is up to 94.0%, and the number of candidate nodules decreases, which improves the detection performance of pulmonary nodules. It indicates that the attention mechanism dynamically adjusts the channel characteristics of the feature map. Hence, useful nodule feature information is enhanced and useless nodule feature information is suppressed. Furthermore, time complexity of structure and the standard deviation of sensitivity values calculated from cross-validation folds are shown in Table 7.

b: RESULTS VISUALIZATION

Considering that CT images belong to 3-D data, only the central slices of true positive nodules detected by 3-D Res-CGAM structure was visualized. In Fig. 9, the upper row shows the ground truth in the dataset, the true position of nodules is marked in green box, and true label 1 and diameter coordinate of the nodules are marked below each sample image. The next line shows the nodular slices detected, where the size of the red box is related to that of the detected nodules. It can be seen that the detection probability for pulmonary nodules with various shapes and different sizes is close to 1, and the diameter of the nodules is close to that of the ideal label. It indicates that the network model efficiently detects pulmonary nodules in chest CT images.

2) FALSE-POSITIVE REDUCTION

a: ABLATION STUDY

FROC curves of the multi-branch classification network performing single-task learning (image reconstruction not included) and multi-task learning (image reconstruction included) are compared in Fig. 10, which reflect the average

TABLE 7. System performance at nodule candidate detection stage.

Structures	Average sensitivity (%)	Total number of candidates	Total number of detected nodules	Average number of candidates/scan	Standard deviation of sensitivity (%)	Memory Usage (GPU)(M)	Training Time
Res-3×3×3	92.2	27998	1094	31.52	2.77	8906	About 9h/fold
Res+3D-CGM	93.2	52929	1105	59.60	2.49	17528	About 15h/fold
Res+Attention	92.5	35450	1097	39.92	2.04	9380	About 10h/fold
3D Res-CGAM	94.0	44627	1114	50.25	2.48	17998	About 21h/fold

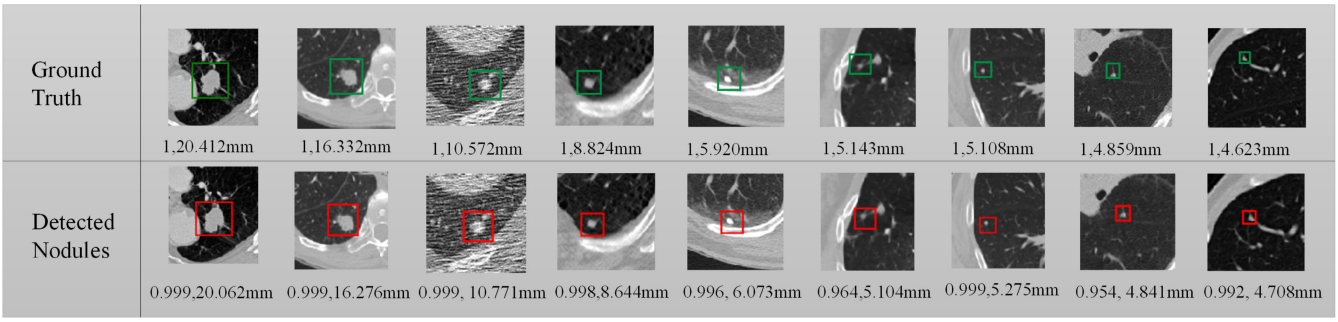


FIGURE 9. Visualization of central slices for nodule ground truths and detection results.

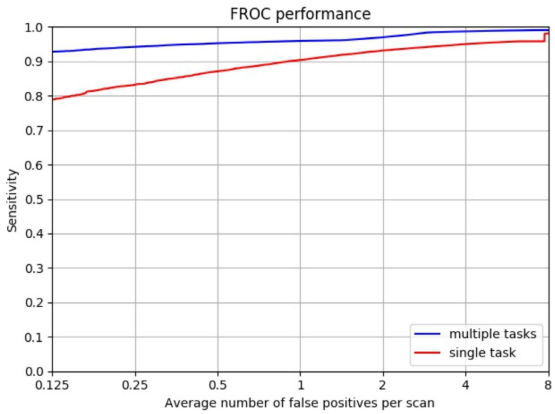


FIGURE 10. Comparison of FROC curves for ablation study.

sensitivity at different false positive rates. On the FROC curve, the abscissa represents the false positives rates of 1/8, 1/4, 1/2, 1, 2, 4, and 8 per scan, and the ordinate represents the sensitivity. It can be seen, the average sensitivity of the two tasks increases steadily with the increase of the average false positive rate of per scan, but the sensitivity of multiple tasks learning is higher than that of single task learning.

The average sensitivity at 7 predefined false positive rates and CPM performance are shown in Table 8. The multiple tasks learning reaches higher average sensitivity at 7 false positive rates per scan than single task learning, and CPM score is up to 0.959.

The confusion matrices of single task and multiple tasks are illustrated in Table 9. The designed model can correctly classify most nodules. It is obviously noted that the number of nodules correctly classified by multi-task model was more than that by single-task model.

Furthermore, more metrics is list in Table 10 to evaluate the network classification performance. Obviously, Multiple tasks overall outperforms than single task in accuracy, precision, recall, specificity and F1-score, which means that it achieves an excellent classification performance of an acceptable false negative rate (Precision) and low false positive rate (Recall).

Overall, the classification performance of multi-task learning is better than that of single-task learning. Therefore, the image reconstruction auxiliary network of the multi-task learning plays a role in adjusting the CNN hierarchy of the shared feature network. And the detail information at pixel level which is beneficial to nodule classification is preserved to improve classification accuracy as much as possible.

b: LOSS WEIGHT SELECTION

Image reconstruction can recover some detail features lost during the feature extraction, and the model performance is improved by making full use of the detail feature information. Hence, the image reconstruction task is introduced into the classification task to improve the classification effect. However, in order to train the prediction model with superior classification performance, it is necessary to ensure that the two task losses are properly weighted in the combined loss function. So, the threshold is set as $\alpha \in (0, 1)$ in formula (7), where α represents the weight ratio of nodule classification loss and image reconstruction loss. As shown in Fig. 11, 6 threshold values are selected as α (0, 0.1, 0.3, 0.4, 0.5, 1.0) between [0,1] to carry out a comparative experiment, and the optimal weight of α is found for image reconstruction loss.

The experimental results are analyzed as follows: (1) When $\alpha = 0$, it means that the model cannot obtain any feature information from the image reconstruction task, which

TABLE 8. System performance at the false positive reduction stage.

Task type	FPs/scan							CPM
	0.125	0.25	0.5	1	2	4	8	
Single task	0.789	0.832	0.870	0.903	0.931	0.950	0.980	0.891
Multiple task	0.928	0.942	0.952	0.959	0.983	0.986	0.990	**0.959**

TABLE 9. Confusion matrix of the classification results.

Confusion Matrix		Predict results			
		0		1	
Task type		Single task	Multiple tasks	Single task	Multiple tasks
Real label	0	106195	107457	3869	2607
	1	1566	527	31007	32046

TABLE 10. Network Classification Performance.

Model	Accuracy	Precision	Recall	Specificity	F1-score
Single task	0.961	0.889	0.951	0.964	0.919
Multiple tasks	0.978	0.924	0.983	0.976	0.953

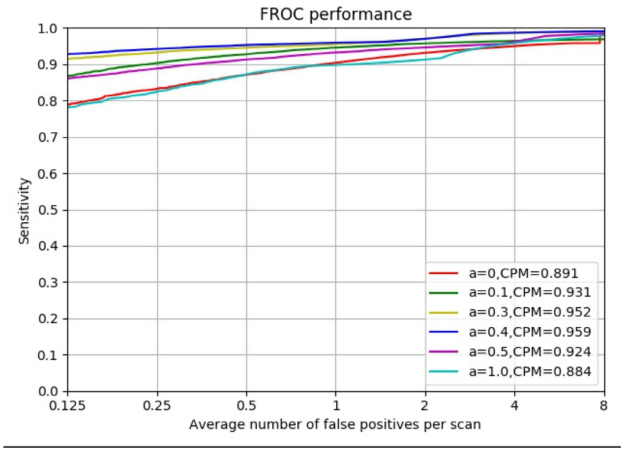


FIGURE 11. FROC curve comparisons with different image reconstruction loss weight in LUNA16.

is equivalent to single task. (2) When $\alpha \in (0, 0.4)$, the classification performance is improved with the increase of threshold α. When $\alpha = 0.4$, the CPM score is the highest and nodules classification effect is the best. When $\alpha \in (0.4, 1)$, the classification performance is decreased with increasing α. However, the classification effect of multiple tasks is slightly better than that of single task, which shows that the image reconstruction task can improve the classification performance. (3) When $\alpha = 1$, it means that classification task and image reconstruction task are not divided into the main task and the auxiliary task, and they are equally important. At this time, the classification model learns more redundant and useless feature information from the image reconstruction. Therefore, the classification effect is even slightly inferior to that of single task.

When a loss threshold α is appropriately selected for image reconstruction, the multi-task model can learn more useful microscopic feature at pixel level to improve the classification performance. So multi-task learning is superior to single-task learning.

V. DISCUSSIONS

To be fair, the performance comparison is conducted on different nodule candidate detection systems under LUNA16 dataset in Table 11. Machine learning algorithm was applied to detect pulmonary nodules in methods [17], [35]–[37], which largely depended on the manual feature extraction of pulmonary nodules, and it was difficult to achieve automatic detection, resulting in poor generalization ability of the model. Traditionally, thresholding, morphological operation, and region growing are often used for lung segmentation and nodule candidate detection, and SVM and its variants are used as classifiers [17], [35]–[37].

Deep learning algorithm was applied to detect pulmonary nodules in methods [7], [25]–[27], [38]–[40], which could learn automatically features from data without relying on manual extraction of features. It can be found that deep learning methods show high sensitivity in the detection of pulmonary nodules. In the methods [25], [38], multiple 2D cross-sectional slices of pulmonary nodules were selected as input, but the spatial context information of 3-D pulmonary nodule features could not be extracted, so the detection sensitivity of 2D network model was lower than that of 3-D network model. However, 3-D network [7], [26], [27], [39], [40] involved more parameters, which increased the storage space requirements, but it could extract rich target spatial information, which was more conducive to detect pulmonary nodules from CT images.

For considering the above-mentioned, a 3-D CAD system is designed to effectively improve the model performance in pulmonary nodule detection and false-positive reduction. We constructed 3-D Res-CGAM architecture based on 3-D Residual U-Net with lower network complexity and bigger model depth, which enhanced the3-D feature extraction ability and model generalization. Obviously, the detection sensitivity is high and the average number of candidates per scan is relatively low.

To comprehensively evaluate false-positive reduction performance of automatic nodules detection system, CPM scores are compared on the LUNA16 dataset in Table 12. Method [10] designed the 3-D input feature maps of three sizes, and the fusion weight was manually determined to classify pulmonary nodules, but its detection rates needed further improvement. Method [28] integrated with three 3-D CNN subnetworks to classify pulmonary nodules at the expense of a high network complexity. Method [11] designed

TABLE 11. Performance comparison during nodule candidate detection.

Algorithm	Methods	Types	#Scans	#Nodules	Sensitivity (%)	FPs/Scans
Machine Learning	Javaid et al. [17]	2D+3D	1308	Unknown	91.65	3.19
	Lu et al. [35]	2D	294	631	85.2	3.31
	Murphy et al. [36]	2D	888	1186	85.6	335.9
	Torres et al. [37]	2D	888	1186	76.8	22.2
Deep Learning	Zhu et al. [7]	3D	888	1186	91.7	Unknown
	Setio et al. [25]	2D	888	1186	90.1	85.4
	Huang et al. [26]	3D	99	Unknown	90.0	Unknown
	Hamidian et al. [27]	3D	534	Unknown	80	15.28
	ZNET. [38]	2D	888	1186	88.4	79.3
	Ypsilantis et al.[39]	3D	1080	Unknown	90.5	4.5
	Dou et al. [40].	3D	888	1186	90.6	86.5
	3D Res-CGAM	3D	888	1186	**94.0**	**50.25**

TABLE 12. Performance comparison during false-positive reduction.

CAD system	False positives per scan							CPM	P-Value
	0.125	0.25	0.5	1	2	4	8		
Zhang et al. [6]	0.890	0.931	0.944	0.949	0.965	0.972	0.976	0.947	0.285
Dou et al. [10]	0.677	0.737	0.815	0.848	0.879	0.907	0.922	0.827	0.002
Zuo et al. [11]	0.630	0.753	0.819	0.869	0.903	0.915	0.920	0.830	0.007
Cao et al. [28]	0.695	0.801	0.851	0.914	0.939	0.949	0.961	0.873	0.035
Ye et al. [41]	0.621	0.733	0.849	0.899	0.918	0.924	0.934	0.839	0.019
Li et al. [42]	0.739	0.803	0.858	0.888	0.907	0.916	0.920	0.861	0.003
Wang et al. [43]	0.788	0.847	0.895	0.934	0.952	0.959	0.963	0.905	0.052
Zheng et al. [44]	0.876	0.899	0.912	0.927	0.942	0.948	0.953	0.922	0.013
H Cao et al. [45]	0.848	0.899	0.925	0.936	0.949	0.957	0.960	0.925	0.051
Setio et al. [46]	0.859	0.937	**0.958**	**0.969**	0.976	0.982	0.982	0.952	0.571
Our method	**0.928**	**0.942**	0.952	0.959	**0.983**	**0.986**	**0.990**	**0.959**	**1**

a 3-D input feature map and multi-scale feature layers were used for feature fusion to improve the feature extraction for high classification accuracy, but the microscopic information of nodules is not fully utilized. Method [43] adopted deep learning model with adaptive feature map size for pulmonary nodule detection, it scores higher than the previous methods. Method [44] input maximum intensity projection images (MIP) of different thickness to improve the detection effect, and used feature fusion of different thickness to reduce false positives. Method [6] used multi-scale LOG Gaussian filter to detect candidate nodules and achieved good classification results, but it contained more false-positive nodules in the nodules candidate detection. Method [46] designed a CAD system that focused on some special types of nodules (proximal pleural nodules, proximal vascular nodules, etc.), and achieved good nodule detection performance, but it lacks universality.

Anova test is performed to supplement the P-Values, which records the significant difference between errors of the proposed and existing method in Table 12. Take P-Value = 0.05 as the evaluation standard of method difference, it is obviously noted that our woks are significantly different from the works in [10], [11], [41], [42] and [44] but equivalent to the rest works with high CPM.

In view of the above works, the proposed 3-D automatic nodules detection system is summarized as follows: In the detection stage, 3-D Res-CGAM structure is designed to obtain rich and representative context feature information of multi-scale pulmonary nodules, which improves the detection sensitivity and model generalization. In the classification stage, the classification task is combined with the image reconstruction task well, and the multi-level features of different receptive fields in the feature extraction process are fully utilized to improve nodules classification accuracy and reduce false positives. Overall, the proposed 3-D detection system applied to detect multi-scale pulmonary nodules with different shapes obviously outperforms than the previous detection systems, and CPM score is up to 0.959. It is worth noting that network model has good generality. Table 13 is supplemented to host all the frequently used abbreviations with their descriptions to improve the readability.

TABLE 13. Abbreviation of different words used in the article.

Abbreviation	Description
CAD	Computer-aided detection
CT	Computerized tomography
CNN	Convolutional neural network
RPN	Region Proposal Network
SE	Squeeze-and-excitation
3D-CGM	3D context-guided module
3D Res-CGAM	3D Residual-context guided attention module
Res+3D-CGM	Residual+3D context-guided module
FROC	Free receiver operating characteristic
CPM	Competition performance metric

When it comes to the limitations of our methodology and the threats-to-validity of the experimental results, unstructured CT data of patients was used to improve the network

model, and achieved good detection sensitivity and classification accuracy. Actually, structured data including clinical baseline data, disease history and laboratory examination are usually used to make an accurate judgment of the patient's condition in clinical practice. Therefore, the future works focus on how to combine the structured data with unstructured data to integrate multi-modal data for perfect features construction is vital to nodule screening.

VI. CONCLUSION

The mainstream feature extraction network adopted in CT image processing is easy to extract redundant macro-features and lose key micro-features during feature information transmission, so the extracted features are lack of representativeness, which resulting in low detection sensitivity and high false positive. Hence, an efficient three-dimensional automatic pulmonary nodule detection system was developed to solve the above thorny problems. (1) A 3-DCGAM module was designed to improve the network feature extraction capability. By extracting representative nodular features, the detection performance of candidate nodules was prior to that of mainstream 3-D CNN, and the sensitivity was up to 94% under the condition of low false positive. (2) A multi-task learning model was designed by combining CT image reconstruction task and nodule classification task, which can fully learn the microscopic features of image spatial context to improve nodule classification performance. The single-channel input mode greatly reduces network parameters, and the receptive field features from different branches can effectively identify pulmonary nodules with different sizes and shapes. The CMP of LUNA16 challenge evaluation was increased to 0.959. This research work is beneficial to promote the research on chest CT image automatic detection equipment under the background of complex medical engineering. It aims to achieve benign and malignant diagnosis of pulmonary nodules and early screening of lung cancer through clinical medical imaging process. Our future work will focus on combing structured and unstructured data to construct excellent pulmonary nodule features and incorporating clinical records into the nodule detection process.

REFERENCES

[1] M. Stimpfel and I. Virant-Klun, "Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012," *J. Cancer Stem Cell Res.*, vol. 4, no. 3, pp. E359–E386, 2016.

[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. Informat.*, vol. 4, no. 1, pp. 27–40, Feb. 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] M. Firmino, A. H. Morais, R. M. Mendoça, M. R. Dantas, H. R. Hekis, and R. Valentim, "Computer-aided detection system for lung cancer in computed tomography scans: Review and future prospects," *Biomed. Eng. OnLine*, vol. 13, no. 1, p. 41, Apr. 2014.

[5] J. Zhang, Y. Xia, H. Cui, and Y. Zhang, "Pulmonary nodule detection in medical images: A survey," *Biomed. Signal. Process.*, vol. 43, pp. 47–138, May 2018.

[6] J. Zhang, Y. Xia, H. Zeng, and Y. Zhang, "NODULe: Combining constrained multi-scale LoG filters with densely dilated 3D deep convolutional neural network for pulmonary nodule detection," *Neurocomputing*, vol. 317, pp. 159–167, Nov. 2018.

[7] W. Zhu, C. Liu, W. Fan, and X. Xie, "DeepLung: Deep 3D dual path nets for automated pulmonary nodule detection and classification," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2018, pp. 673–681.

[8] L. Gong, S. Jiang, Z. Yang, G. Zhang, and L. Wang, "Automated pulmonary nodule detection in CT images using 3D deep squeeze-and-excitation networks," *Int. J. Comput. Assist. Radiol. Surgery*, vol. 14, no. 11, pp. 1969–1979, Nov. 2019.

[9] P. Zhai, Y. Tao, H. Chen, T. Cai, and J. Li, "Multi-task learning for lung nodule classification on chest CT," *IEEE Access*, vol. 8, pp. 180317–180327, 2020.

[10] 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.

[11] W. Zuo, F. Zhou, and Y. He, "An embedded multi-branch 3D convolution neural network for false positive reduction in lung nodule detection," *J. Digit. Imag.*, vol. 33, no. 4, pp. 846–857, Aug. 2020.

[12] B. C. Lassen, C. Jacobs, J.-M. Kuhnigk, B. van Ginneken, and E. M. van Rikxoort, "Robust semi-automatic segmentation of pulmonary subsolid nodules in chest computed tomography scans," *Phys. Med. Biol.*, vol. 60, no. 3, pp. 1307–1323, Jan. 2015.

[13] Z. Jing, L. Bin, and T. Lianfang, "Lung nodule classification combining rule-based and SVM," in *Proc. IEEE 5th Int. Conf. Bio-Inspired Comput., Theories Appl. (BIC-TA)*, Sep. 2010, pp. 1033–1036.

[14] J. Gong, J.-Y. Liu, L.-J. Wang, X.-W. Sun, and S.-D. Nie, "Automatic detection of pulmonary nodules in CT images by incorporating 3D tensor filtering with local image feature analysis," *Phys. Med.*, vol. 46, pp. 124–133, Feb. 2018.

[15] F. V. Farahani, A. Ahmadi, and M. H. F. Zarandi, "Lung nodule diagnosis from CT images based on ensemble learning," in *Proc. IEEE Conf. Comput. Intell. Bioinf. Comput. Biol. (CIBCB)*, Aug. 2015, pp. 1–7.

[16] R. Arulmurugan and H. Anandakumar, "Early detection of lung cancer using wavelet feature descriptor and feed forward back propagation neural networks classifier," *Comput. Vis. Bio Inspired Comput.*, vol. 28, pp. 103–110, Feb. 2018.

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

[18] M. Poongodi, M. Hamdi, M. Malviya, A. Sharma, G. Dhiman, and S. Vimal, "Diagnosis and combating COVID-19 using wearable Oura smart ring with deep learning methods," *Pers. Ubiquitous Comput.*, 2021, doi: 10.1007/s00779-021-01541-4.

[19] H. R. Roth, L. Lu, J. Liu, J. Yao, A. Seff, K. Cherry, L. Kim, and R. M. Summers, "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.

[20] J. Ding, A. Li, Z. Hu, and L. Wang, "Accurate pulmonary nodule detection in computed tomography images using deep convolutional neural networks," in *Proc. MICCAI*, vol. 10435, 2017, pp. 559–567.

[21] T. Meraj *et al.*, "Lung nodules detection using semantic segmentation and classification with optimal features," *Neural Comput. Appl.*, vol. 33, pp. 10737–10750, 2020.

[22] X. Hongtao, D. Yang, N. Sun, Z. Chen, and Y. Zhang, "Automated pulmonary nodule detection in CT images using deep convolutional neural networks," *Pattern Recognit.*, vol. 85, pp. 109–119, Jan. 2019.

[23] S. Albahli, H. T. Rauf, M. Arif, M. T. Nafis, and A. Algosaibi, "Identification of thoracic diseases by exploiting deep neural networks," *Comput., Mater. Continua*, vol. 66, no. 3, pp. 3139–3149, 2021.

[24] S. Albahli, "AI-driven deep CNN approach for multi-label pathology classification using chest X-rays," *PeerJ Comput. Sci.*, vol. 7, p. e495, Apr. 2021.

[25] A. A. A. Setio, F. Ciompi, G. Litjens, P. Gerke, C. Jacobs, S. J. Van Riel, M. M. W. Wille, M. Naqibullah, C. I. Sánchez, and B. Van Ginneken, "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.

[26] 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. (ISBI)*, Apr. 2017, pp. 379–383.

[27] S. Hamidian, B. Sahiner, N. Petrick, and A. Pezeshk, "3D convolutional neural network for automatic detection of lung nodules in chest CT," in *Proc. SPIE*, Mar. 2017, Art. no. 1013409.

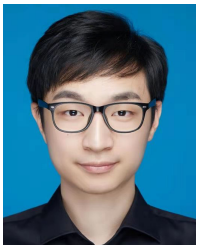
[28] H. Cao, H. Liu, E. Song, G. Ma, X. Xu, R. Jin, T. Liu, and C.-C. Hung, "Multi-branch ensemble learning architecture based on 3D CNN for false positive reduction in lung nodule detection," *IEEE Access*, vol. 7, pp. 67380–67391, 2019.
[29] A. Pezeshk, S. Hamidian, N. Petrick, and B. Sahiner, "3-D convolutional neural networks for automatic detection of pulmonary nodules in chest CT," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 5, pp. 2080–2090, Sep. 2019.
[30] R. V. M. da Nóbrega, P. P. R. Filho, M. B. Rodrigues, S. P. P. da Silva, C. M. J. M. Dourado, Jr., and V. H. C. de Albuquerque, "Lung nodule malignancy classification in chest computed tomography images using transfer learning and convolutional neural networks," *Neural Comput. Appl.*, vol. 32, no. 15, pp. 11065–11082, Aug. 2020.
[31] D. Poap, M. Wozniak, R. Damasevicius, and W. Wei, "Chest radiographs segmentation by the use of nature-inspired algorithm for lung disease detection," in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, Nov. 2018, pp. 2298–2303.
[32] G. Capizzi, G. L. Sciuto, C. Napoli, D. Polap, and M. Wozniak, "Small lung nodules detection based on fuzzy-logic and probabilistic neural network with bioinspired reinforcement learning," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 6, pp. 1178–1189, Jun. 2020.
[33] A. T. Sahlol, M. A. Elaziz, A. T. Jamal, R. Damaševičius, and O. F. Hassan, "A novel method for detection of tuberculosis in chest radiographs using artificial ecosystem-based optimisation of deep neural network features," *Symmetry*, vol. 12, no. 7, p. 1146, Jul. 2020.
[34] 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.
[35] 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.
[36] K. Murphy, B. van Ginneken, A. M. R. Schilham, B. J. de Hoop, H. A. Gietema, and M. Prokop, "A large-scale evaluation of automatic pulmonary nodule detection in chest CT using local image features and k-nearest-neighbour classification," *Med. Image Anal.*, vol. 13, no. 5, pp. 757–770, Oct. 2009.
[37] E. L. Torres, E. Fiorina, F. Pennazio, C. Peroni, M. Saletta, N. Camarlinghi, M. E. Fantacci, and P. Cerello, "Large scale validation of the M5L lung CAD on heterogeneous CT datasets," *Med. Phys.*, vol. 42, no. 4, pp. 1477–1489, Apr. 2015.
[38] A. A. A. Setio, A. Traverso, T. De Bel, M. S. N. Berens, C. Van Den Bogaard, P. Cerello, H. Chen, Q. Dou, M. E. Fantacci, B. Geurts, and R. van der Gugten, "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.
[39] P.-P. Ypsilantis and G. Montana, "Recurrent convolutional networks for pulmonary nodule detection in CT imaging," 2016, *arXiv:1609.09143*.
[40] Q. Dou, H. Chen, Y. Jin, H. Lin, J. Qin, and P. A. Heng, "Automated pulmonary nodule detection via 3D convnets with online sample filtering and hybrid-loss residual learning," in *Proc. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, 2017, pp. 630–638.
[41] Y. Ye, M. Tian, Q. Liu, and H.-M. Tai, "Pulmonary nodule detection using V-Net and high-level descriptor based SVM classifier," *IEEE Access*, vol. 8, pp. 176033–176041, 2020.
[42] Y. Li and Y. Fan, "DeepSEED: 3D squeeze-and-excitation encoder-decoder convolutional neural networks for pulmonary nodule detection," in *Proc. IEEE 17th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2020, pp. 1866–1869.
[43] J. Wang, J. Wang, Y. Wen, H. Lu, T. Niu, J. Pan, and D. Qian, "Pulmonary nodule detection in volumetric chest CT scans using CNNs-based nodule-size-adaptive detection and classification," *IEEE Access*, vol. 7, pp. 46033–46044, 2019.
[44] S. Y. Zheng, "Automatic pulmonary nodule detection in CT scans using convolutional neural networks based on maximum intensity projection," *IEEE Trans. Med. Image*, vol. 39, pp. 795–805, 2020.
[45] H. Cao, H. Liu, E. Song, G. Ma, X. Xu, R. Jin, T. Liu, and C.-C. Hung, "A two-stage convolutional neural networks for lung nodule detection," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 7, pp. 2006–2015, Jul. 2020.
[46] A. A. A. Setio, A. Traverso, T. D. Bel, M. S. N. Berens, C. Van Den Bogaard, P. Cerello, H. Chen, Q. Dou, M. E. Fantacci, B. Geurts, and R. van der Gugten, "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.

HAIYING YUAN received the B.S., M.S., and Ph.D. degrees from the University of Electronic Science and Technology of China, in 1999, 2004, and 2006, respectively. As a Visiting Scholar, she had been involved in scientific research for a long time with the University of California at Santa Barbara, from 2014 to 2015, the University of Tsinghua, from 2016 to 2017, the IMEC Inter-University Microelectronics Center, and Ku Leuven University (Katholieke University Leuven), in 2017. She is in charge of a number of national, provincial, and ministerial research projects. As an Associate Professor at the Beijing University of Technology, her research interests include image processing, pattern recognition, medical imaging diagnosis and data analysis, AI algorithm and chip design and verification, and energy-efficient computing systems. She is currently a Senior Member of the CCF Professional Committee.

YANRUI WU is currently pursuing the master's degree in electronic science and technology from the Faculty of Information Technology, Beijing University of Technology. His research interests include computer vision and medical image analysis based on deep learning.

JUNPENG CHENG is currently pursuing the master's degree with the Beijing University of Technology. His research interests include deep learning algorithm and hardware implementation in computer vision.

ZHONGWEI FAN is currently pursuing the master's degree in electronic science and technology with the Faculty of Information Technology, Beijing University of Technology. His research interests include deep learning and medical imaging diagnosis and its application.



ZHIYONG ZENG is currently pursuing the master's degree with the Faculty of Information Technology, Beijing University of Technology. His research interests include the neural network accelerator architecture design and implementation.