Hindawi
International Journal of Intelligent Systems
Volume 2023, Article ID 3028869, 14 pages
https://doi.org/10.1155/2023/3028869

*Research Article*

# Pulmonary Nodule Detection from 3D CT Image with a Two-Stage Network

**Miao Liao**,[1] **Zhiwei Chi**,[1] **Huizhu Wu**,[1] **Shuanhu Di**,[2] **Yonghua Hu**,[1] **and Yunyi Li**[1]

[1]*School of Computer Science and Engineering, Hunan University of Science and Technology, Xiangtan 411100, China*
[2]*College of Intelligence Science and Technology, National University of Defense Technology, Changsha 410073, China*

Correspondence should be addressed to Shuanhu Di; dish0304@163.com

Received 26 July 2023; Revised 17 October 2023; Accepted 8 December 2023; Published 31 December 2023

Academic Editor: Paolo Gastaldo



Early detection of lung nodules is an important means of reducing the lung cancer mortality rate. In this paper, we propose a three-dimensional CT image lung nodule detection method based on parallel pooling and dense blocks, which includes two parts, i.e., candidate nodule extraction and false positive suppression. First, a dense U-shaped backbone network with parallel pooling is proposed to obtain the candidate nodule probability map. The parallel pooling structure uses multiple pooling operations for downsampling to capture spatial information comprehensively and address the problem of information loss resulting from maximum and average pooling in the shallow layers. Then, a parasitic network with parallel pooling, dense blocks, and attention modules is designed to suppress false positive nodules. The parasitic network takes the multiscale feature maps of the backbone network as the input. The experimental results demonstrate that the proposed method significantly improves the accuracy of lung nodule detection, achieving a CPM score of 0.91, which outperforms many existing methods.

## 1. Introduction

Lung cancer is the most prevalent cancer worldwide, ranking first among all cancers in terms of malignancy and lethality [1]. Early detection, diagnosis, and treatment of lung cancer play a key role in improving the survival rate of patients. The early diagnosis of lung cancer relies on the detection and localization of lung nodules in medical images [2]. Compared to positron emission computed tomography (PET) and magnetic resonance imaging (MRI), computed tomography (CT) has faster imaging speed, lower cost, and higher density resolution, making it widely used in lung nodule detection [3]. Given that chest CT imaging produces a substantial number of slices, typically around 300 per patient, manual detection is exceedingly time-consuming. Moreover, lung nodules are characterized by their small size, weak boundaries, and varied locations, making manual detection challenging and leading to high rates of both missed detections and false alarms. Therefore, developing efficient and accurate automated detection methods for lung nodules in CT images is of great significance in improving the precision and efficiency of lung cancer computer-aided diagnosis and treatment.

Pulmonary nodules in CT scans are often characterized by indistinct borders, heterogeneous greyscale, and varied shapes. Malignant nodules can have a diameter as small as 3-4 mm, making automatic detection of nodules in lungs highly challenging. The existing methods for lung nodule detection are mainly divided into traditional, machine learning-based, and deep learning-based ones [4]. Traditional methods mainly use morphology, thresholding, clustering, and model optimization to identify and localize lung nodules directly from complex lung images [5, 6]. For example, Abdollahzadeh Rezaie and Ali [7] first used thresholding to obtain the region of interest of lung nodules and then used edge detection to locate the nodules. Lu et al. [8] proposed a hybrid method for lung nodule detection, which involved various traditional methods such as morphological operations, Hessian matrices, fuzzy sets, and regression trees. The traditional methods are often complicated and require human-computer interactions with

different software programs, leading to high false positive rates.

To achieve fully automated detection of lung nodules, many scholars have proposed machine learning-based methods. These methods first extract multiple artificial features from the image, such as intensity, texture, and shape, and then use a classification model to classify the extracted features to achieve the goal of recognizing target areas. For example, Aghabalaei Khordehchi et al. [9] used spectral, texture, and shape features to characterize nodules and then used a support vector machine (SVM) to identify nodules in images. Nithila and Kumar [10] extracted texture features including contrast, correlation, energy, uniformity, and moments from images and input them into a neural network to identify nodular and non-nodular areas. Machine learning relies on selecting a large number of artificial features. However, artificial features crafted manually relying on prior knowledge frequently exhibit shortcomings, such as being arbitrary, incomplete, and inefficient. Moreover, the fitting ability of most classifiers is limited and they perform poorly on samples with nonlinear features.

Deep learning can automatically learn efficient and more discriminative features from training data and enable end-to-end training and testing [11]. Current deep learning networks for lung nodule detection mainly include 2D CNNs [12] and 3D CNNs [13–15]. For example, Jiang et al. [16] utilized four identical 2D CNNs to detect four images with different resolutions and enhanced the images using Frangi filtering. As this method uses a single slice image for nodule detection, it is highly susceptible to the influence of pulmonary microvascular cross-sections, often resulting in a high false positive rate. To capture relationships between CT sequence slices, Wang et al. [17] input consecutive CT slices into a 2D CNN for multiscale feature fusion. 2D CNN-based methods are limited in acquiring three-dimensional texture and shape features and may mistakenly identify blood vessels as lung nodules, leading to a higher false positive rate. Therefore, most lung nodule detection networks are currently designed based on 3D convolutions. For example, Cao et al. [18] proposed a two-stage detection network, in which residual and dense structures were introduced into a 3D UNet for candidate nodule detection, followed by a 3D CNN-based classification network to reduce false positive rates. Liu et al. [19] developed a 3D feature pyramid network to improve the detection sensitivity of the network by using multiscale features to discriminate lung nodules. In addition, the network introduced a false positive suppression module to track the appearance changes of each candidate nodule on consecutive CT slices, further identifying true pulmonary nodules and eliminating misdiagnoses. Khosravan and Bagci [14] designed a dense connection-based segmentation network to obtain the probability of pulmonary nodule existence in CT image. The study compared the performance of different downsampling strategies, including max pooling, average pooling, and stride-2 convolution, in lung nodule detection, where max pooling achieved relatively better performance. Huang et al. [20] first used a 2D UNet network with squeeze and excitation blocks (SE blocks) to segment the candidate nodules in CT slices, followed by a 3D sequence network with SE blocks to identify the 3D pixel blocks containing candidate nodules. Currently, almost all three-dimensional lung nodule detection networks use three-dimensional maximum or average pooling for downsampling, resulting in a significant spatial compression ratio of the feature maps and a loss of structural information in the images.

To address the problems mentioned above, we propose a lung nodule detection method for 3D CT image with a two-stage network, which consists of two parts: candidate nodule extraction and false positive suppression mask generation. First, a 3D primary network based on parallel downsampling and dense blocks is used for candidate nodule extraction, which is an improvement on the UNet [21] network. Dense blocks are used to replace the convolution layers of UNet, and a parallel downsampling structure is designed to replace mean downsampling. Then, a hybrid attention-based parasitic network is proposed, which takes multiscale feature maps from the encoder of the primary network as the input to generate false positive suppression mask. The hybrid attention module is used to enhance the network's spatial awareness of the lung structure. During training, candidate nodule probabilities are introduced as spatial weights to improve the ability to detect true and false nodules in the candidate regions. Furthermore, a cross-entropy loss function with edges is designed to enhance the performance and training efficiency. The paper introduces several key innovations, including the following:

(1) A two-stage network framework is proposed, in which a primary network is used to detect candidate nodules and a parasitic one is used to suppress false positive nodules. The candidate module probability is utilized for the parasitic network training to focus the network on distinguishing true and false positives in candidate nodules.

(2) A parallel pooling downsampling structure is proposed, which incorporates multiple pooling operations to comprehensively capture spatial information during downsampling. Meanwhile, convolution and global average pooling are employed to obtain channel contributions and the pooling results are fused according to the contributions to enhance the discriminative features and suppress irrelevant ones.

(3) An edge-based cross-entropy loss function is proposed. The loss function sets a lossless region for simple samples to eliminate the loss caused by the normal background regions in the lung. In addition, this loss function balances the proportion of positive and negative sample regions in CT images and urges the network to focus on the regions that are difficult to identify.

## 2. Methods

The workflow of the proposed method for lung nodule detection from CT images is shown in Figure 1, which mainly includes two stages, namely, candidate nodule extraction and mask generation. First, a parallel pooling dense

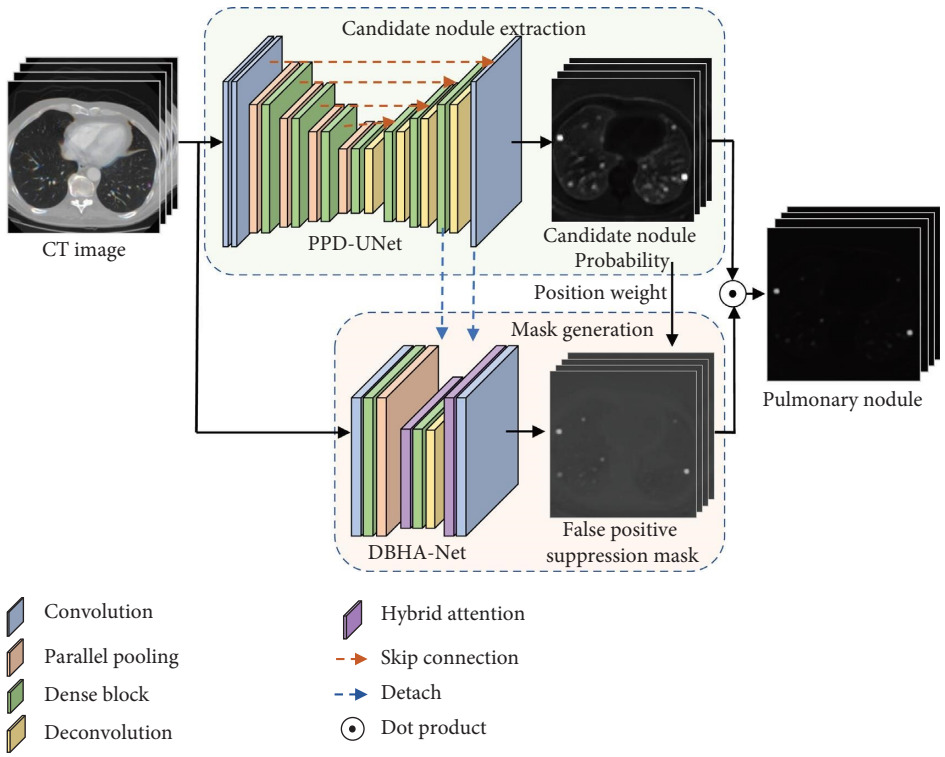


Figure 1: Workflow of the proposed method for lung nodule detection.

U-shaped network, denoted by PPD-UNet, is used to analyze image features and obtain the probability map of candidate nodules. Dense blocks are used in PPD-UNet to replace the conventional convolution layers. The dense skip connections in dense blocks can facilitate the propagation and utilization of features. Meanwhile, a parallel pooling structure is proposed in PPD-UNet, which uses convolution and global average pooling to obtain channel contributions from multiple sets of pooling results. According to the contributions, the pooling results are fused to enhance discriminative features and capture more comprehensive spatial information. In addition, considering that the nodules connected to the pulmonary wall and vessels as well as those with atypical shape and tiny size are difficult to detect, we assign higher weights to such nodules through online hard sample mining during the training process. This operation can effectively improve the network's sensitivity to lung nodule recognition and ensure that the lung nodules in the image are effectively detected as much as possible. However, it may also inevitably lead to a certain degree of false positives.

To suppress the false positive nodules, we propose a parasitic network using dense blocks and hybrid attention, denoted by DBHA-PNet. The network takes CT images as the input and shares the parsing capability of the host network by introducing the deep feature maps in PPD-UNet. Besides, a hybrid attention is introduced to enhance the network's spatial awareness of lung structure and reduce false positive nodules. To improve the discrimination ability for false positive nodules, we add the probability map of candidate nodules as position weights to the loss calculation. Affected by the position weights, regions with lower candidate nodule probabilities are given less attention. Finally, the mask obtained by DBHA-PNet is multiplied by the probability map of candidate nodules to decrease the probability of false positive nodules.

*2.1. Candidate Nodule Extraction.* In this section, a PPD-UNet is proposed to obtain a candidate nodule probability map, which uses 7 dense blocks as feature extraction units and applies parallel pooling for downsampling. The structure of PPD-UNet is shown in Figure 2. First, the input 3D lung CT image is processed by a $7 \times 7 \times 7$ convolution layer, followed by several sets of parallel pooling and a dense block in the encoder. A tensor carrying spatial coordinate information is introduced after the first dense block. The added coordinate tensor can enhance the network's perception of the spatial structure of the lungs. Skip connections are employed between the encoder and the decoder to promote the propagation and utilization of features. The probability map for candidate nodules is obtained by applying a $1 \times 1 \times 1$ convolution and Sigmoid activation function. To further obtain the size of pulmonary nodules, a threshold is used to segment the probability map. The center point of the segmentation result is considered the position of the pulmonary nodule. The diameter of the pulmonary nodule is then calculated based on the segmented volume.

Similar to H-DenseUNet [22], PPD-UNet also use dense blocks for feature extraction. However, there are still many differences between PPD-UNet and H-DenseUNet. First, the

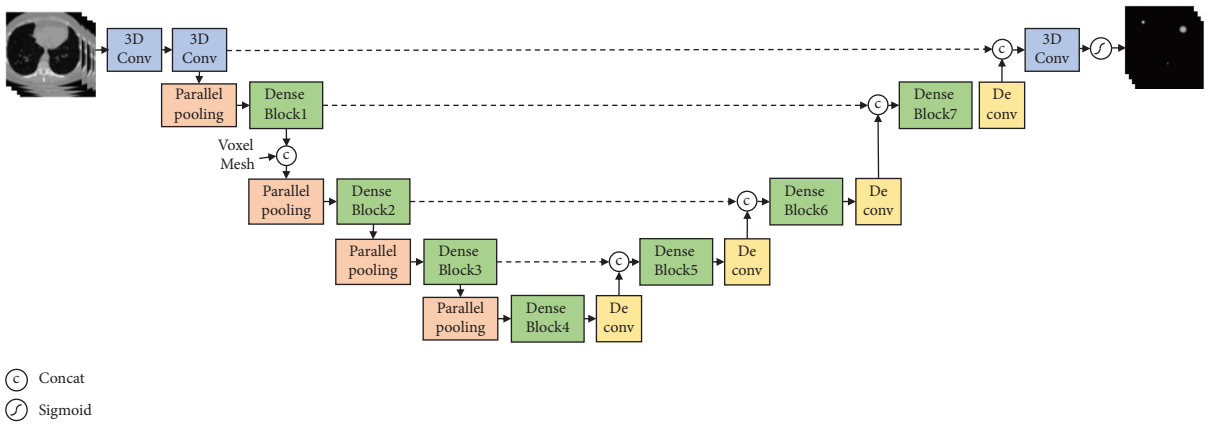


Figure 2: Structure of PPD-UNet.

structure of the dense blocks used in PPD-UNet is different from that in H-DenseUNet. The dense block used in PPD-UNet contains multiple convolution combinations, each of which is composed of two sets of the BN-Relu--Conv operation. Besides the last convolution combination, the inputs of the other ones are concatenated with their corresponding outputs by skip connections. Figure 3 shows a dense block containing four convolution combinations. The convolution combination at the end is used to control the number of channels and fuse the output of the previous feature maps. Second, the downsampling operations used in PPD-UNet and H-DenseUNet are different. The H-DenseUNet employs average pooling for downsampling, which can easily lead to the loss of structural information. To address this issue, we propose a parallel pooling downsampling, whose structure will be illustrated in Section 2.2.

The input of dense blocks passes through all convolution combinations. In the U-shaped structure, as the feature map size decreases, the number of channels increases, and the features carried become more advanced and abstract. Following this principle, we use dense blocks with more convolution combinations when dealing with smaller feature maps. The number of convolution combinations used in dense blocks is listed in Table 1.

*2.2. Parallel Pooling.* The detection of pulmonary nodules depends on accurate interpretation of their intensity and texture. Conventional downsampling operations compress image spatial information, leading to blurred texture edges. To alleviate the information loss caused by downsampling, it is necessary to convert the spatial features such as intensity and texture into a more abstract representation. The transformation from specific features to abstract features depends on the feature extraction module with a sufficiently large receptive field.

To this end, a parallel pooling module is designed, which includes two stages, pooling and fusion. The structure of parallel pooling is shown in Figure 4. In the pooling phase, various pooling operations, including average pooling, maximum pooling, and stride convolution pooling, are used to capture features comprehensively. In the fusion phase, the pooling results are first concatenated to form a new feature map of size [3C, D, H, W]. Subsequently, three groups of $7 \times 7 \times 7$ convolutions are used to obtain three groups of feature maps with dimensions of [C, D, H, W], and the global average pooling is used to compress the three groups of feature maps to obtain three groups of channel vectors with a length of C. Finally, the three channel vectors are fused with the corresponding pooling results from the pooling phase by channel-wise multiplication and addition.

It is worth noting that although both of the parallel pooling and the convolutional block attention module (CBAM) [23] involve operations such as pooling and convolution, they are entirely different modules. First, the structures of parallel pooling and CBAM are different. CBAM only employs global max pooling and average pooling to get attention weights, while the proposed parallel pooling introduces local max pooling, average pooling, and stride convolution for comprehensive feature capture. Additionally, CBAM employs a serial structure, sequentially performing channel attention and spatial attention modules. In contrast, the parallel pooling contains a pooling stage and a fusion stage. In the pooling stage, three distinct downsampling operations are conducted. In the fusion stage, a series of operations, including convolution, global pooling, sigmoid activation, element-wise multiplication, and channel-wise addition, are employed to explore deep-level relationships within the feature maps, suppress low-discriminative features, and enhance high-discriminative features.

Second, the purposes of the modules are different. CBAM is an attention module designed to guide the network's focus towards essential spatial and channel information in the image. In contrast, the proposed parallel pooling module is used for image downsampling, addressing the issue of information loss caused by conventional downsampling. Concretely, it converts spatial information into higher-level features during pooling to prevent information loss. The output dimensions of the CBAM module remain unchanged compared to the input, while the parallel pooling module, after integrating various pooling features, reduces the output dimensions to half of the input.

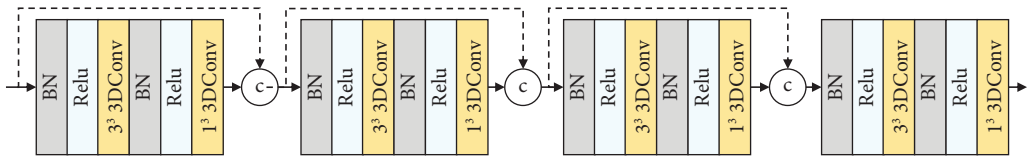


Figure 3: Structure of dense blocks.

Table 1: The number of convolution combinations used in dense blocks.

| Module | Number of convolution combinations |
|---|---|
| Dense block1 | 4 |
| Dense block2 | 6 |
| Dense block3 | 8 |
| Dense block4 | 12 |
| Dense block5 | 6 |
| Dense block6 | 4 |
| Dense block7 | 4 |

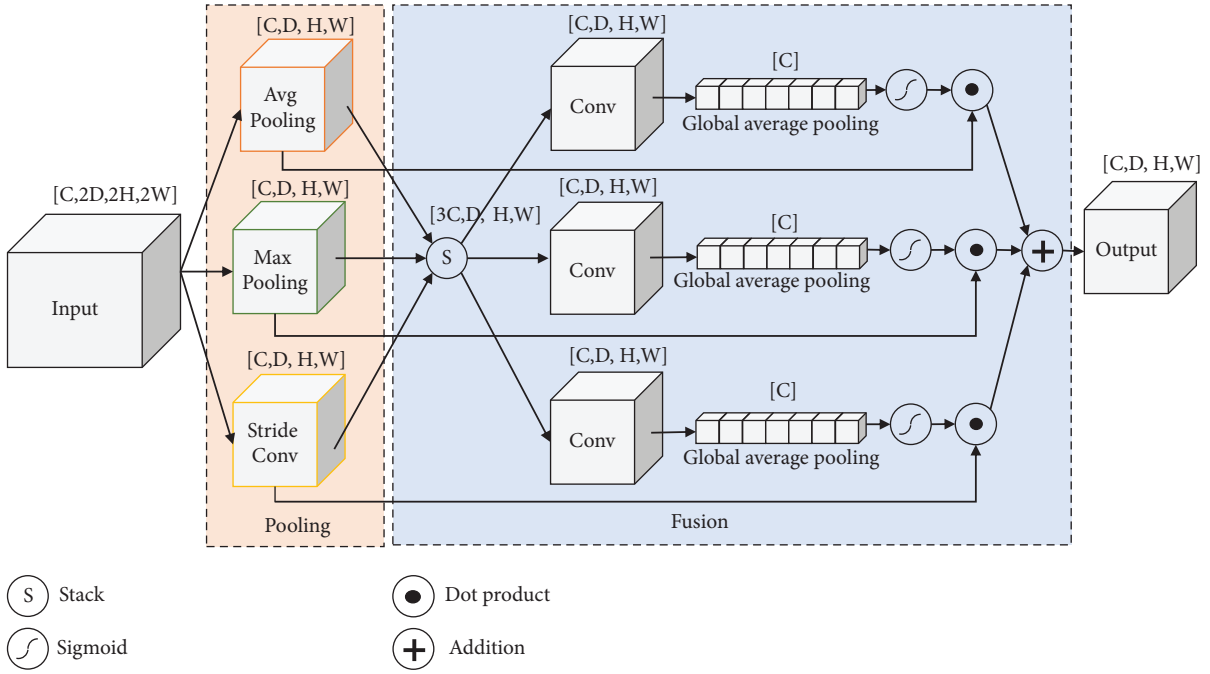


Figure 4: Structure of parallel pooling.

*2.3. False Positive Suppression Mask Generation.* To suppress the false positive nodules, we propose a parasitic network called DBHA-PNet, which is composed of parallel pooling, hybrid attention (HA), and dense block. The network takes the deep features of PPD-UNet as inputs. The hybrid attention is utilized to enhance the spatial perception of the network and alleviate the interference of non-nodules. The probability of candidate nodules is used as the position weight in the training process to guide the network to focus on discriminating true positive nodules from false ones. The structure of DBHA-PNet is shown in Figure 5. This network first takes CT images as the input and employs convolutional and dense blocks for shallow feature extraction. Meanwhile, DBHA-PNet leverages the deep features of PPD-UNet and shares its parameters, avoiding repetitive feature extraction steps. To integrate deep and shallow features, DBHA-PNet applies a HA after each feature map concatenation. The HA cascades a channel attention and a spatial attention. Finally, the false positive suppression mask is obtained by applying a convolution layer and a sigmoid activation function.

## 3. Experiment

*3.1. Dataset and Preprocessing.* The LUNA16 pulmonary nodules public dataset is applied, which contains 888 low-dose chest CT images [24]. There are 1186 lung nodules with

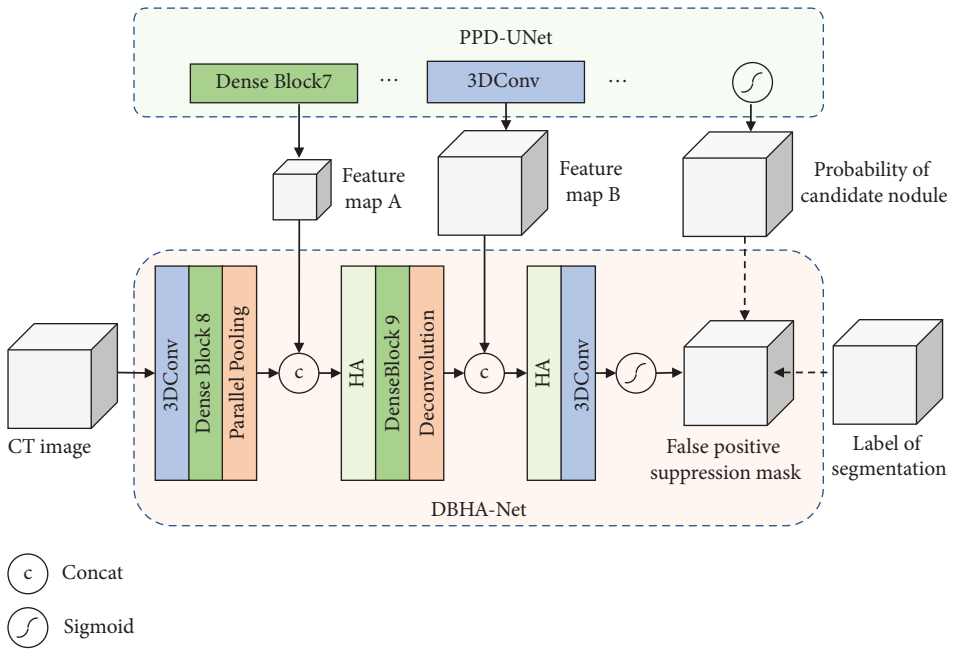


Figure 5: Structure of DBHA-PNet.

annotated positions and diameters in total. The database includes 10 subsets, and the experiment adopts 10-fold cross validation. Rotation and scaling operations are employed for data enhancement.

In this paper, the candidate nodule probability map and false positive suppression mask are obtained by performing segmentation tasks, which requires pixel-wise segmentation labels of lung nodules for training. However, the LUNA16 dataset only provides positions and diameters of pulmonary nodules, assuming the pulmonary nodules are spherical. In fact, pulmonary nodules may have various shapes. For instance, malignant pulmonary nodules may appear as ground glass opacity or have an irregular shape, as shown in Figure 6.

Considering the imprecise edge labeling, we adopt a soft-edge segmentation labeling method, illustrated in Figure 7. The region within 0.7 times the radius of the pulmonary nodule (i.e., inside the magenta circle) is marked as positive sample and that outside 1.4 times the radius (i.e., outside the green circle) as negative sample. The region between the green and magenta circles is marked as uncertain. We can reduce the impact of imprecise label by decreasing the loss weight of the uncertain regions.

*3.2. MCE Loss Function.* The proportion of lung volume occupied by lung nodules is very small, and their diameters generally range from 3 mm to 30 mm. Although the majority of normal areas are relatively easy to distinguish, their high proportion still leads to a considerable amount of loss, which will interfere the training of key areas, such as nodules and boundaries.

To solve this problem, we propose an edge-based cross entropy loss function, denoted by MCE loss. The loss function defines lossless regions for non-nodule samples

Figure 6: Examples of pulmonary nodules with blurred boundaries and various shapes.

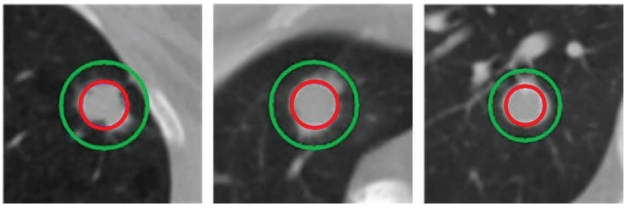

Figure 7: Illustration of soft-edge labeling.

that are easy to distinguish. This strategy is employed to eliminate the loss accumulation caused by normal samples, alleviates the problem of imbalance between difficult and easy samples, and enhances the attention to lung nodules. Moreover, the lossless region is gradually reduced in training to improve the prediction confidence of the model. The MCE loss is defined as follows:

$$\text{MCE loss}(p, g) = \frac{1}{N}\sum \begin{cases} -\alpha * g * \log\left(\frac{p}{\beta}\right), & |g - p| \geq \beta, \\ 0, & |g - p| < \beta, \end{cases} \tag{1}$$

where $p$ and $g$, respectively, represent the predicted probability and true label, $\alpha$ is used to assign weights to positive and negative samples to balance them, and $\beta$ is to define the lossless interval. Simple samples usually come from normal lung areas. The loss generated by the simple sample is relatively small. However, due to the large proportion of normal lung areas in CT images, a large amount of loss will still be generated. Considering the class imbalance in training, we do not calculate the loss for the samples with $|g - p| < \beta$ and urges the network to focus on the regions that are difficult to identify.

We compare the MCE loss with some other commonly used ones, such as cross-entropy (CE) loss and focal loss as follows [25]:

$$\begin{aligned} \text{CE Loss}(p, g) &= \frac{1}{N} \sum -g * \log(p), \\ \text{Focal Loss}(p, g) &= \frac{1}{N} \sum -\alpha * g * \log(p) * (1 - p)^{\gamma}. \end{aligned} \tag{2}$$

The CE loss is widely used for classification tasks to evaluate the model's performance. However, it cannot adapt to the problem of class imbalance. To address this, the focal loss introduces the weights $\alpha$ and $\gamma$ to the cross-entropy loss to reduce the weight of easy-to-classify samples.

The curves of CE loss, focal loss ($\alpha = 0.5$, $\gamma = 0.8$), and MCE loss ($\alpha = 0.5$, $\beta = 0.75$) functions are plotted in Figure 8. The horizontal coordinate shows the probabilities of pixels classified to ground truth class, and the vertical coordinate shows the corresponding losses. Pixels with low probabilities are considered difficult to classify, while those with high probabilities are considered easy. Compared to CE loss, MCE loss and focal loss have steeper gradients for the samples difficult to classify, which can enable the network focus on the hard-to-classify samples in the training process. Besides, to balance the weights of simple and difficult samples, MCE loss set the loss to 0 for the easy-to-classify samples whose probabilities classified to ground truth class is within (0.75, 1.0). Compared to focal Loss, the MCE loss does not involve exponential calculation, which significantly speeds up volume-rendering-based image analysis tasks. Meanwhile, the MCE loss dynamically adjusts the parameter $\beta$ during training to promote the comprehensive learning of both easy and difficult samples and improving the prediction confidence of the network. The MCE loss function was used for all segmentation training in this study. The $\alpha$ values used for positive and negative sample balance in the PPD-UNet and DBHA-PNet are set to 0.6 and 0.3, respectively.

*3.3. Training Strategy.* The experiments in this study were conducted in a Linux environment using Python 3.7 and PyTorch 1.10 framework for model training and testing. The initial learning rate for PPD-UNet training was set to 0.01. When the evaluation index of the validation set did not decrease for more than 5 epochs during the training process, the learning rate was reduced to 1/10 of the original. The training of DBHA-PNet begins after 10 epochs, and its learning rate adjustment strategy is consistent with that of PPD-UNet.

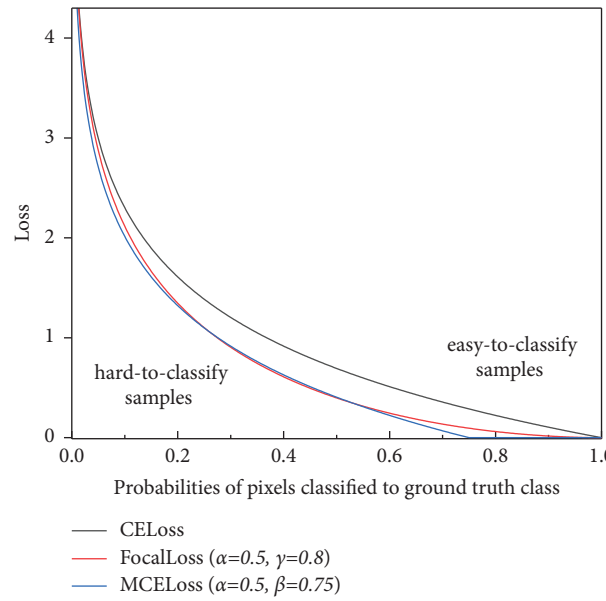


Figure 8: Curves of different objective functions.

## 4. Experimental Results and Comparison

Figure 9 presents some intermediate results by our method, where columns (a), (b), (c), (d), and (e), respectively, show the CT images, candidate nodule probability map, false positive suppression mask, nodule probability map, and ground truth. First, PPD-UNet is used to extract candidate nodules from CT images. As shown in Figure 9(b), although the lung nodule regions are effectively detected, a large number of false positive nodules are also introduced. DBHA-PNet focuses on distinguishing true and false positive nodules among the candidates to generate a false-positive suppression mask. As shown in Figure 9(c), almost all the false positive nodules are effectively suppressed. Finally, the candidate nodule probability map is multiplied by the false-positive suppression mask to obtain the final lung nodule probability map, as shown in Figure 9(d).

*4.1. Ablation Experiment.* The main innovations in this study lie in the dense block (DB), parallel pooling (PP), and parasitic network. The dense block is used instead of conventional convolution to extract features comprehensively. Parallel pooling aims to convert the spatial features such as intensity and texture into a more abstract representation. A parasitic network DBHA-PNet is designed to improve the network's detection performance by sharing the host network's deep features.

To validate the effectiveness of each module, we conducted numerous ablation experiments. The experimental configurations are as follows:

(1) PPD-UNet w/o DB or PP: average pooling is used in place of the parallel downsampling in PPD-UNet, and conventional convolution layers were used to replace the dense blocks in PPD-UNet.

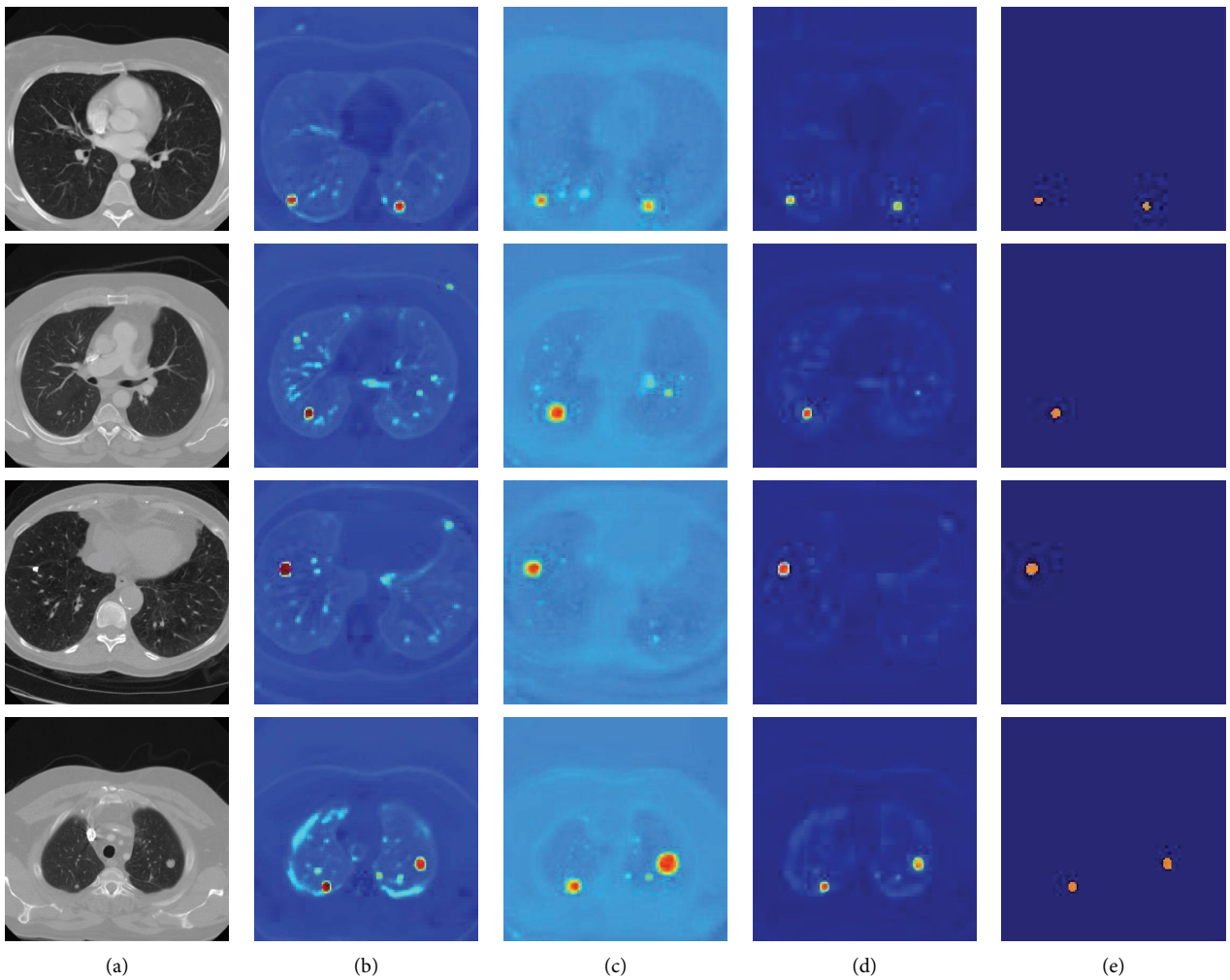


Figure 9: Some intermediate results by our method. (a) CT images, (b) candidate nodule probability map, (c) false positive suppression mask, (d) nodule probability map, and (e) ground truth.

(2) PPD-UNet w/o DB: conventional convolution layers are used to replace the dense blocks in PPD-UNet.

(3) PPD-UNet w/o PP: average pooling is used in place of the parallel downsampling in PPD-UNet.

(4) PPD-UNet: only applying the primary network PPD-UNet for lung nodule detection.

(5) PPD-UNet w/DBHA-PNet: the parasitic network DBHA-PNet is added to the primary network, i.e., the proposed method.

The FROC (free-response receiver operating characteristic) curve is used to evaluate the detection results quantitatively. The FROC curve is a commonly used evaluation indicator in object detection or localization tasks, which can evaluate the algorithm's sensitivity under different false positive conditions. The sensitivity is defined as follows:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{3}$$

where TP, FP, and FN, respectively, denote the number of true positive, false positive, and false negative nodules.

The FROC curves obtained with different configurations are shown in Figure 10. FPs/scan, applied as the horizontal axis unit, represents the average number of false positive nodules detected per scan (i.e., per patient CT sequence). Sensitivity, applied as the vertical axis unit, represents the lung nodule detection sensitivity under the false positive constraints. As can be seen, lung nodule detection sensitivity is improved as the number of false positive nodules allowed per scan increases.

Lung nodule detection is a very challenging task. Lung nodules in CT images typically exhibit small size, heterogeneous intensities, various shapes, varied locations, and weak boundaries. In lung nodule detection, it is highly susceptible to missing small nodules and misclassifying other similar tissues as lung nodules, resulting in a high false positive rate. To enhance the recognition accuracy of nodules with small size and those connected to lung walls and blood vessels, we introduce dense blocks and parallel pooling downsampling operations into the PPD-UNet network. The former can alleviate gradient vanishing, strengthening feature extraction and propagation, and the latter can

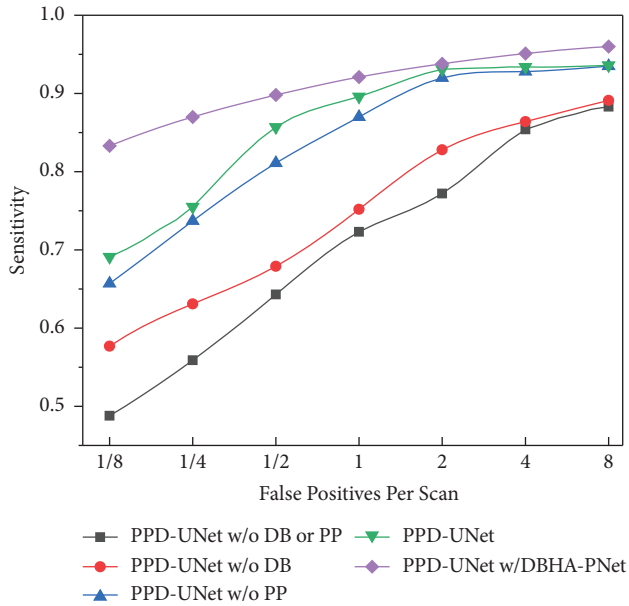


Figure 10: FROC curves obtained with different configurations.

transform spatial information into higher-level features, preventing information loss of small targets during downsampling. Furthermore, to address the issue of high false positives, we develop a parasitic network called DBHA-PNet. This parasitic network utilizes the multiscale features of the host network PPD-UNet as the input and employs hybrid attention mechanisms to enhance the network's spatial perception of lung structures and suppress false-positive nodules.

PPD-UNet w/o DB or PP does not include dense blocks or parallel pooling. PPD-UNet w/o DB does not include dense blocks, and PPD-UNet w/o PP does not include parallel pooling. These networks have limited feature extraction capabilities. While a PPD-UNet introduces dense blocks, it does not utilize a parasitic network for false-positive suppression. From Figure 10, it can be observed that under the condition of low false-positive constraints per scan, PPD-UNet w/o DB or PP and PPD-UNet w/o DB exhibit very low detection sensitivity. This indicates that the models cannot accurately discriminate lung nodules from other similar tissues. PPD-UNet improves the network's feature extraction and downsampling modules, resulting in significantly enhanced performance compared to PPD-UNet w/o DB or PP and PPD-UNet w/o DB. However, due to the absence of a parasitic network, PPD-UNet still tends to produce a relatively high false positive rate. As shown in Figure 10, when FPs/scan is controlled within a lower range, the sensitivity of PPD-UNet significantly decreases. However, when FPs/scan exceeds 2, the sensitivity can reach above 0.93, markedly higher than that of PPD-UNet w/o DB or PP and PPD-UNet w/o DB and comparable to that of our proposed method, i.e., PPD-UNet w/DBHA-PNet.

By comparing the different FROC curves, it is evident that the introduction of dense blocks, parallel pooling, and parasitic network all contributed to the improvement of the network's detection performance. The parasitic network focuses on distinguishing true positive and false positive nodules among the candidates. The mask generated by the parasitic network exerts a suppression effect on the false positive nodules. As shown in Figure 10, by combining the main network and the parasitic network, the proposed method can significantly improve the detection sensitivity under the condition of lower false positive constraints per scan, such as when FPs/scan is less than 1/2.

Table 2 presents the sensitivity and CPM score for pulmonary nodule detection under 1/8 FPs/scan, 1/4 FPs/scan, 1/2 FPs/scan, 1 FPs/scan, 2 FPs/scan, 4 FPs/scan, and 8 FPs/scan. The CPM score is the average sensitivity across the seven conditions. Compared to PPD-UNet w/o DB or PP and PPD-UNet w/o PP, PPD-UNet w/o DB and PPD-UNet, respectively, resulted in 6.1% and 2.9% improvements in the CPM score by introducing the parallel-pooling structure. Compared to PPD-UNet, PPD-UNet + DBHA-PNet results in a 5.7% improvement in the CPM score. In addition, there was a significant improvement in the detection sensitivity at lower false-positive rates in the PPD-UNet + DBHA-PNet model. For instance, the detection sensitivity has a significant improvement of 18.8% under 1/8 FPs/scan. This means that the model is able to accurately detect pulmonary nodules even at a lower number of allowed false positives per scan. This is an important factor as it reduces the potential for false alarms or unnecessary treatments.

To validate the effectiveness of the proposed MCE loss, we compared it with the focal loss. Figure 11 displays training losses obtained with different loss functions. Considering the class imbalance in training, MCE loss sets lossless regions for non-nodule samples that are easy to distinguish, urging the network to focus on the regions that are difficult to identify. As depicted in Figure 11, when MCE loss is employed as the objective function, the network demonstrates a notably accelerated convergence rate and attains reduced loss values. Table 3 shows the detection sensitivity and the CPM score of the models trained with MCE loss and focal loss. The model trained with MCE loss demonstrates a clear advantage over that trained with focal loss.

*4.2. Comparison with Other Methods.* The proposed approach utilizes 3D chest CT images as input and employs a U-shaped network to obtain a probability map of nodules. Then, a feature fusion network based on parallel pooling and attention is used to obtain a false positive suppression mask. Finally, a point-by-point multiplication is applied to combine the probability map of the nodules with the false positive suppression mask, resulting in better detection performance. The approach focuses more on the 3D spatial structural information and makes full use of the spatial information of the lung to improve the network's ability to distinguish false positives.

We compared our method with many existing ones, including convolution-based models such as NoduleNet [26], S4ND [14], DBResNet [13], V-Net [15], and H-DenseUNet [22], as well as transformer-based models such as Swin-UNetr [27] and UNetr [28]. In all the

Table 2: Detection sensitivity and the CPM score achieved by different configurations.

| Model | 1/8 FPs/scan | 1/4 FPs/scan | 1/2 FPs/scan | 1 FPs/scan | 2 FPs/scan | 4 FPs/scan | 8 FPs/scan | CPM |
|---|---|---|---|---|---|---|---|---|
| PPD-UNet w/o DB or PP | 0.488 | 0.559 | 0.643 | 0.723 | 0.772 | 0.854 | 0.883 | 0.703 |
| PPD-UNet w/o DB | 0.577 | 0.631 | 0.679 | 0.752 | 0.828 | 0.864 | 0.891 | 0.746 |
| PPD-UNet w/o PP | 0.657 | 0.737 | 0.811 | 0.870 | 0.920 | 0.928 | 0.935 | 0.837 |
| PPD-UNet | 0.701 | 0.775 | 0.857 | 0.896 | 0.931 | 0.934 | 0.936 | 0.861 |
| PPD-UNet + DBHA-PNet | 0.833 | 0.87 | 0.898 | 0.921 | 0.938 | 0.951 | 0.960 | 0.910 |

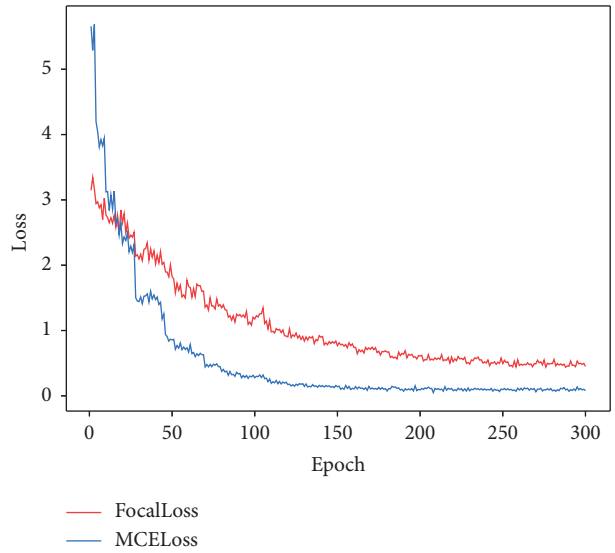


Figure 11: Training losses obtained with different loss functions.

Table 3: Detection sensitivity and the CPM score achieved with different loss functions.

| Loss functions | 1/8 FPs/scan | 1/4 FPs/scan | 1/2 FPs/scan | 1 FPs/scan | 2 FPs/scan | 4 FPs/scan | 8 FPs/scan | CPM |
|---|---|---|---|---|---|---|---|---|
| Focal loss | 0.784 | 0.837 | 0.863 | 0.899 | 0.916 | 0.942 | 0.954 | 0.885 |
| MCE loss | 0.833 | 0.870 | 0.898 | 0.921 | 0.938 | 0.951 | 0.960 | 0.910 |

comparative experiments, 3D images are used as input for all networks. The position and radius of lung nodules are calculated using the predicted probability maps of the nodules.

Figure 12 shows some pulmonary nodule detection results by different methods. Column (a) displays the original CT image of the pulmonary nodule, where the golden circle represents the true location of the nodule, and R represents the radius of the nodule. Columns (b), (c), (d), (e), (f), (g), (h), and (i), respectively, show the detection results using the proposed method, NoduleNet, S4ND, DBResNet, V-Net, H-DenseUNet, Swin-UNetr, and UNetr, where P represents the probability of the region being predicted as a nodule and R represents the predicted radius of the nodule. Alternating rows show the amplified versions of ROIs marked by white rectangles in the upper row. The proposed method introduces hybrid attention mechanism in the parasitic network, which can effectively suppress false positive nodules. Swin-UNetr adopts a shifted window-based attention mechanism, which is capable of capturing long-range dependencies effectively, aiding in handling of complex tasks. Compared to the conventional transformer-based UNetr, Swin-UNetr demonstrates superior performance in pulmonary nodule detection. Although Swin-UNetr can identify the pulmonary nodules in CT images accurately, our proposed method still exhibits advantages in predicting the confidence and radius of pulmonary nodules.

The FROC curves obtained by different methods on the LUNA16 dataset are shown in Figure 13. All methods achieved a detection sensitivity of over 0.9 under the condition of allowing 8 FPs/scan. However, under the condition of allowing relatively low false positive, such as 1/8 FPs/scan and 1/4 FPs/scan, and the detection sensitivities obtained by our method are significantly higher than those of other ones.

Table 4 provides the detection sensitivity and the CPM score for different methods under some false positive rate conditions. As expected, the proposed method shows obvious advantage in lung nodule detection, especially for the case that the false positive rate is set to a low value.

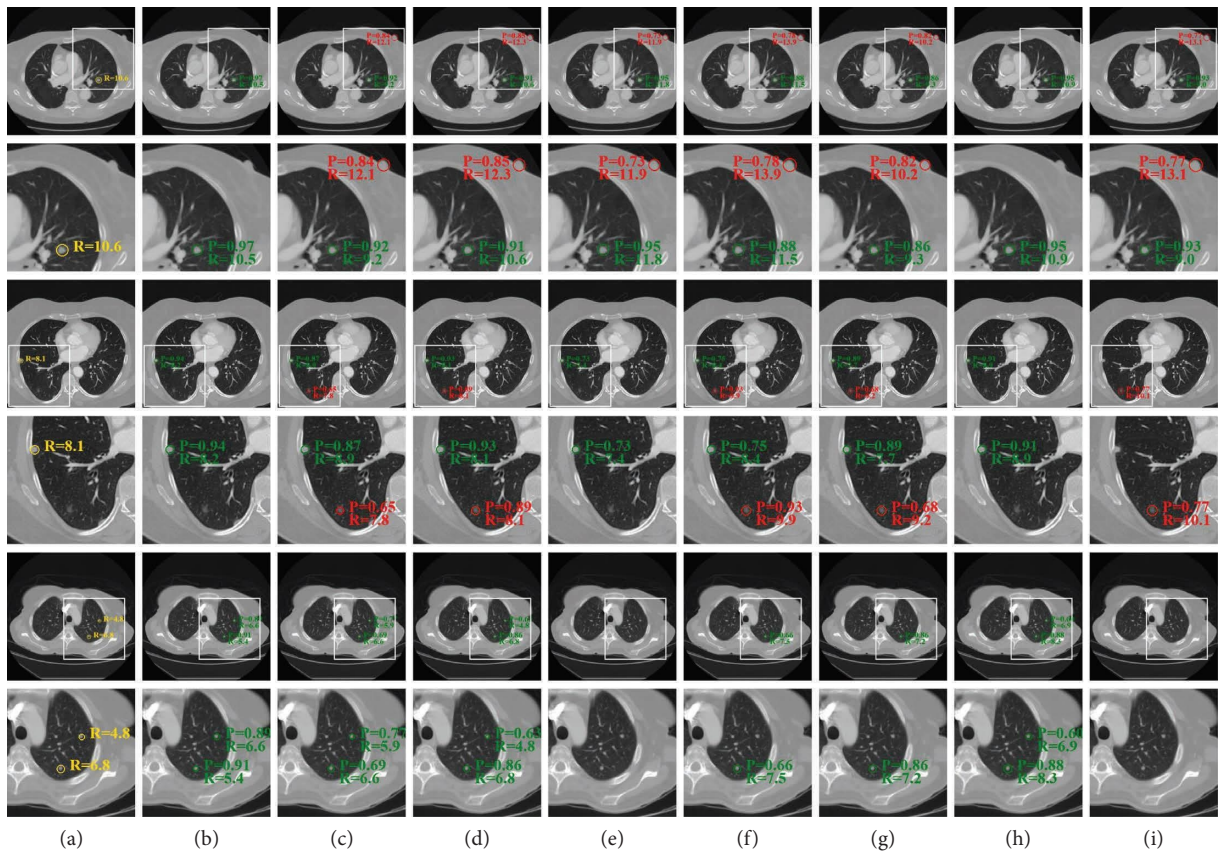


FIGURE 12: Some pulmonary nodule detection results by different methods. (a) Ground truth, (b) the proposed mothed, (c) NoduleNet, (d) S4ND, (e) DBResNet, (f) V-Net, (g) H-DenseUNet, (h) Swin-UNetr, and (i) UNetr.

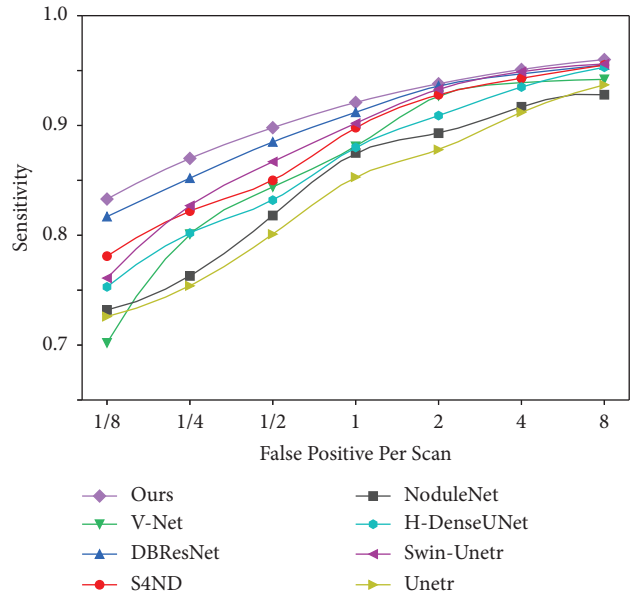


FIGURE 13: FROC curves obtained by different methods on the LUNA16 dataset.

*4.3. Implementation Platform and Running Time.* Table 5 shows the parameter numbers, FLOPs, and the average inference time per case of different methods. All the experiments are executed on the same device (NVIDIA GeFroce RIX 3090 GPU with 24 GB memory). UNetr and Swin-UNetr utilize Transformer [29] and Swin Transformer [30], respectively, as their encoders. As a result, when processing high-dimensional data, they require substantial

Table 4: Detection sensitivity and the CPM score achieved by different methods.

| Methods | 1/8 FPs/scan | 1/4 FPs/scan | 1/2 FPs/scan | 1 FPs/scan | 2 FPs/scan | 4 FPs/scan | 8 FPs/scan | CPM |
|---|---|---|---|---|---|---|---|---|
| NoduleNet (2020) [26] | 0.732 | 0.763 | 0.818 | 0.875 | 0.893 | 0.917 | 0.928 | 0.846 |
| S4ND (2019) [14] | 0.781 | 0.822 | 0.85 | 0.898 | 0.928 | 0.943 | 0.955 | 0.882 |
| DBResNet (2020) [13] | 0.817 | 0.852 | 0.885 | 0.912 | 0.936 | 0.947 | 0.955 | 0.900 |
| V-Net (2020) [15] | 0.702 | 0.801 | 0.844 | 0.881 | 0.927 | 0.939 | 0.942 | 0.862 |
| H-DenseUNet (2018) [22] | 0.753 | 0.802 | 0.832 | 0.88 | 0.909 | 0.935 | 0.953 | 0.866 |
| Swin-UNetr (2021) [27] | 0.761 | 0.827 | 0.867 | 0.902 | 0.933 | 0.949 | 0.956 | 0.885 |
| UNetr (2022) [28] | 0.726 | 0.754 | 0.801 | 0.853 | 0.878 | 0.912 | 0.937 | 0.837 |
| Ours | **0.833** | **0.87** | **0.898** | **0.921** | **0.938** | **0.951** | **0.96** | **0.910** |

The best results are highlighted in bold.

Table 5: Comparison of different methods on parameter numbers, FLOPs, and inference.

| Methods | #Parameters (M) | FLOPs (T) | Inference time (s) |
|---|---|---|---|
| NoduleNet (2020) [26] | 50.19 | 0.28 | 0.93 |
| S4ND (2019) [14] | 4.58 | 0.17 | 0.84 |
| DBResNet (2020) [13] | 7.4 | 0.17 | 0.85 |
| V-Net (2020) [15] | 65.95 | 4.31 | 1.73 |
| H-DenseUNet (2018) [22] | 40 | 1.18 | 1.27 |
| Swin-UNetr (2021) [27] | 61.98 | 0.68 | 1.02 |
| UNetr (2022) [28] | 102.2 | 4.48 | 1.74 |
| Ours | 93.32 | 1.09 | 1.27 |

computational resources. DBResNet and S4ND only focus on the design of the encoding path, resulting in relatively simple network structures and fewer parameters. NoduleNet, V-Net, and H-DenseUNet have complete encoding and decoding paths, and the parameter numbers involved in them are significantly more than those in DBResNet and S4ND. The proposed method is a two-stage model, comprising a candidate nodule extraction network PPD-UNet and a false-positive suppression parasitic network DBHA-PNet. The model is relatively complex, with a parameter number of 93.32 M. The average testing time for a case of our method is 1.27 s, slightly higher than the methods [13, 14, 26, 27] but still acceptable.

## 5. Conclusion

Accurate detection of lung nodules in CT images is a prerequisite for early diagnosis and treatment of lung cancer. To improve the accuracy of lung nodule detection, we propose a two-stage lung nodule detection method. First, a primary network based on parallel pooling and dense blocks is designed to obtain the candidate nodule probability map. Then, a parasitic network is designed to extract false positive suppression masks. By sharing the parameters of the primary network, the parasitic network focuses on distinguishing true and false positives in candidate nodules. In addition, considering the imbalanced positive and negative samples, an edge-based cross-entropy loss function is proposed. By setting a lossless region for healthy regions, the network pays more attention on difficult-to-classify samples. The proposed method was extensively evaluated on the publicly available LUNA16 dataset and compared with various existing methods. The detection sensitivity of different methods under different false positive limits was discussed by applying the FROC curve. The results show that the proposed method effectively reduces the false positive rate by introducing the parasitic network and achieves higher detection sensitivity when the false positive rate is controlled low.

## Data Availability

The data used to support the findings of this study are publicly available as described in the main text.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Authors' Contributions

M. L., Z. C., and S. D. conceived the initial ideas. M. L., Z. C., and H. W. designed the model and experiments. M. L., S. D., and Y. H. performed the data analysis. M. L., Z. C., and H. W. wrote the manuscript with feedback from all the other authors. S. D., Y. H., and Y. L. provided helpful discussions and supervised the project. All the authors have reviewed the manuscript.

## Acknowledgments

This work was supported by the National Key R&D Program of China (grant 2022ZD0119003) and the National Natural Science Foundation of China (grant 62272161).

## References


[1] H. Sung, J. Ferlay, R. L. Siegel et al., "Global cancer statistics 2020: globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *Ca-a Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209–249, 2021.

[2] R. Nooreldeen and H. Bach, "Current and future development in lung cancer diagnosis," *International Journal of Molecular Sciences*, vol. 22, no. 16, p. 8661, 2021.

[3] J. Yanase and E. Triantaphyllou, "A systematic survey of computer-aided diagnosis in medicine: past and present developments," *Expert Systems with Applications*, vol. 138, 2019.

[4] A. Halder, D. Dey, and A. K. Sadhu, "Lung nodule detection from feature engineering to deep learning in thoracic ct images: a comprehensive review," *Journal of Digital Imaging*, vol. 33, no. 3, pp. 655–677, 2020.

[5] B. Wang, X. Tian, Q. Wang et al., "Pulmonary nodule detection in ct images based on shape constraint cv model," *Medical Physics*, vol. 42, no. 3, pp. 1241–1254, 2015.

[6] S. A. El-Regaily, "Lung nodule segmentation and detection in computed tomography," in *Proceedings of the 2017 Eighth International Conference on Intelligent Computing and Information Systems (ICICIS)*, pp. 72–78, IEEE, Cairo, Egypt, December 2017.

[7] A. Abdollahzadeh Rezaie and H. Ali, "Detection of lung nodules on medical images by the use of fractal segmentation," 2017, https://reunir.unir.net/bitstream/handle/123456789/11782/ijimai20174_5_2_pdf_65326.pdf?sequence=1&isAllowed=y.

[8] L. Lu, Y. Tan, L. H. Schwartz, and B. Zhao, "Hybrid detection of lung nodules on ct scan images," *Medical Physics*, vol. 42, no. 9, pp. 5042–5054, 2015.

[9] E. Aghabalaei Khordehchi, A. Ayatollahi, and M. R. Daliri, "Automatic lung nodule detection based on statistical region merging and support vector machines," *Image Analysis and Stereology*, vol. 36, no. 2, pp. 65–78, 2017.

[10] E. E. Nithila and S. S. Kumar, "Automatic detection of solitary pulmonary nodules using swarm intelligence optimized neural networks on ct images," *Engineering science and technology, an international journal*, vol. 20, no. 3, pp. 1192–1202, 2017.

[11] H. Ali, F. Mohsen, and Z. Shah, "Improving diagnosis and prognosis of lung cancer using vision transformers: a scoping review," *BMC Medical Imaging*, vol. 23, no. 1, p. 129, 2023.

[12] G. Pezzano, V. Ribas Ripoll, and P. Radeva, "Cole-cnn: context-learning convolutional neural network with adaptive loss function for lung nodule segmentation," *Computer Methods and Programs in Biomedicine*, vol. 198, 2021.

[13] H. Cao, H. Liu, E. Song et al., "Dual-branch residual network for lung nodule segmentation," *Applied Soft Computing*, vol. 86, 2020.

[14] N. Khosravan and U. Bagci, "S4nd: single-shot single-scale lung nodule detection," in *Proceedings of the Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference*, pp. 794–802, Springer, Granada, Spain, September 2018.

[15] S. Kumar and S. Raman, "Lung nodule segmentation using 3-dimensional convolutional neural networks," in *Soft Computing for Problem Solving: SocProS 2018*,vol. 1, pp. 585–596, Springer, 2020.

[16] 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 journal of biomedical and health informatics*, vol. 22, no. 4, pp. 1227–1237, 2018.

[17] S. Wang, M. Zhou, Z. Liu et al., "Central focused convolutional neural networks: developing a data-driven model for lung nodule segmentation," *Medical Image Analysis*, vol. 40, pp. 172–183, 2017.

[18] H. Cao, H. Liu, E. Song et al., "A two-stage convolutional neural networks for lung nodule detection," *IEEE journal of biomedical and health informatics*, vol. 24, no. 7, pp. 2006–2015, 2020.

[19] J. Liu, L. Cao, O. Akin, and Y. Tian, "3dfpn-hs 2: 3d feature pyramid network based high sensitivity and specificity pulmonary nodule detection," in *Proceedings of the Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference*, pp. 513–521, Springer, Shenzhen, China, October 2019.

[20] W. Q. Huang, Y. T. Zhang, H. Dong, and Y. M. Wang, "Dsenet: double three-dimensional squeeze-and-excitation network for pulmonary nodule detection," in *Proceedings of the 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, IEEE, Chongqing, China, March 2021.

[21] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference*, pp. 234–241, Springer, Munich, Germany, October 2015.

[22] X. Li, H. Chen, X. Qi, Q. Dou, C. W. Fu, and P. A. Heng, "H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes," *IEEE Transactions on Medical Imaging*, vol. 37, no. 12, pp. 2663–2674, 2018.

[23] S. Woo, J. Park, J.-Y. Lee, and K. So, "Cbam: convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, Munich, Germany, September 2018.

[24] A. A. A. Setio, A. Traverso, T. De Bel et al., "Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the luna16 challenge," *Medical Image Analysis*, vol. 42, pp. 1–13, 2017.

[25] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 318–327, 2020.

[26] H. Tang, C. Zhang, and X. Xie, "Nodulenet: decoupled false positive reduction for pulmonary nodule detection and segmentation," in *Proceedings of the Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference*, pp. 266–274, Springer, Shenzhen, China, October 2019.

[27] H. Ali, V. Nath, and Y. Tang, "Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images," in *Proceedings of the International MICCAI Brainlesion Workshop*, pp. 272–284, Springer, Strasbourg, France, September 2021.

[28] H. Ali, Y. Tang, and V. Nath, "Unetr: transformers for 3d medical image segmentation," in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pp. 574–584, Waikoloa, HI, USA, January 2022.

[29] Alexey Dosovitskiy, L. Beyer, and K. Alexander, "An image is worth 16x16 words: transformers for image recognition at scale," in *Proceedings of the International Conference on Learning Representations (ICLR)*, Vienna, Austria, May 2021.

[30] Z. Liu, Y. Lin, and Y. Cao, "Swin transformer: hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF international conference on computer vision*, Montreal, BC, Canada, October 2021.