

RESEARCH ARTICLE

3D multi-scale deep convolutional neural networks for pulmonary nodule detection

Haixin Peng, Huacong Sun*, Yanfei Guo

College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao, Shandong, China

* 1123541829@qq.com

Abstract

With the rapid development of big data and artificial intelligence technology, computer-aided pulmonary nodule detection based on deep learning has achieved some successes. However, the sizes of pulmonary nodules vary greatly, and the pulmonary nodules have visual similarity with structures such as blood vessels and shadows around pulmonary nodules, which make the quick and accurate detection of pulmonary nodules in CT image still a challenging task. In this paper, we propose two kinds of 3D multi-scale deep convolution neural networks for nodule candidate detection and false positive reduction respectively. Among them, the nodule candidate detection network consists of two parts: 1) the backbone network part Res2SENet, which is used to extract multi-scale feature information of pulmonary nodules, it is composed of the multi-scale Res2Net modules of multiple available receptive fields at a granular level and the squeeze-and-excitation units; 2) the detection part, which uses a region proposal network structure to determine region candidates, and introduces context enhancement module and spatial attention module to improve detection performance. The false positive reduction network, also composed of the multi-scale Res2Net modules and the squeeze-and-excitation units, can further classify the nodule candidates generated by the nodule candidate detection network and screen out the ground truth positive nodules. Finally, the prediction probability generated by the nodule candidate detection network is weighted average with the prediction probability generated by the false positive reduction network to obtain the final results. The experimental results on the publicly available LUNA16 dataset showed that the proposed method has a superior ability to detect pulmonary nodules in CT images.

OPEN ACCESS

Citation: Peng H, Sun H, Guo Y (2021) 3D multi-scale deep convolutional neural networks for pulmonary nodule detection. PLoS ONE 16(1): e0244406. <https://doi.org/10.1371/journal.pone.0244406>

Editor: Haoran Xie, Lingnan University, HONG KONG

Received: July 6, 2020

Accepted: December 8, 2020

Published: January 7, 2021

Copyright: © 2021 Peng et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All CT image files are available from the publicly available LUNA16 database (<https://doi.org/10.5281/zenodo.2595812> and <https://doi.org/10.5281/zenodo.2596478>). url: <https://luna16.grand-challenge.org/Download/>.

Funding: Funded studies 1. Y.J. Peng, No. 61976126, the National Natural Science Foundation of China, <http://www.nsf.gov.cn/>, The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. 2. Y.J. Peng, No. ZR2019MF003, the Natural Science Foundation of Shandong Province,

Introduction

Lung cancer is one of the most dangerous malignancies to human health and life [1]. According to medical clinical experience, once the clinical symptoms of lung cancer show, the cure rate is very low, so the early detection of pulmonary nodules is of great significance for reducing lung cancer mortality [2]. As an important means of screening lung cancer in high-risk population, low-dose computed tomography scanning has been used in health examination on a large scale, but a large number of CT data have brought a lot of works to doctors and

<http://cloud.sdstc.gov.cn/>. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

radiologists, and high-intensity works are easy to cause misdiagnosis of doctors. With the rapid development of big data and artificial intelligence technology, computer-aided detection based on deep learning has attracted wide attention [3].

At present, deep learning has achieved excellent results in the field of pulmonary nodule detection in chest CT sequence images. Zhu et al. [4] used Faster R-CNN [5] with dual path blocks and U-Net-like encoder-decoder structure for nodule candidate detection with the FROC (average sensitivity at the false positives as 0.125, 0.25, 0.5, 1, 2, 4, 8) score on the LUNA16 dataset [6] is 0.842. Dou et al. [7] used a 3D fully convolutional network [8] for lung nodule candidate detection and achieve a FROC score of 0.839 on the LUNA16 dataset, then use a residual network for false positive reduction. The sensitivity reaches 0.905 when the average number of false positives(FPs) per scan is 1. Khosravan et al. [9] proposed a pulmonary nodule detection network called S4ND, which consists of densely connected convolution blocks and is trained in an end-to-end manner, no post-processing is required to perfect the detection result and a FROC score of 0.897 can be achieved on the LUNA16 dataset. Xie et al. [10] adjusted the structure of 2D Faster R-CNN through two region proposal networks and an deconvolutional layer to detect nodule candidates with the highest sensitivity up to 0.864 on the LUNA16 dataset, then three 2D models are used to train three types of slices with different locations to reduce the number of false positive nodules with a FROC score of 0.790. Dou et al. [11] proposed a multilevel contextual 3D convolutional neural network for false positive reduction by using CT image cubes of different sizes as input and achieve a FROC score of 0.827 on the LUNA16 dataset. Wang et al. [12] concatenated three adjacent axial slices to construct 3D RGB images for nodule detection, the highest sensitivity can reach 0.968 when the average number of candidates per scan is 60.23 on the LUNA16 dataset, then false positives are reduced by two Inception-v4 networks [13] of different receptive fields with a FROC score of 0.903. Although the above methods have achieved good results, there is still room for improvement in sensitivity and false positives. How to use the data characteristics of CT sequence images to design more efficient network structure is the key to improve the performance of computer-aided detection systems.

This paper is based on deep convolutional neural network (DCNN). In order to improve system sensitivity and reduce false positives, we first use nodule candidate detection network to detect nodule candidates, and then use false positive reduction network to further classify nodule candidates to obtain final results. The main contributions of this paper are as follows:

1. CT image is composed of continuous sequence slices and 3D CNN can better capture the spatial information of CT sequence images and extract more abundant features. As a result, we designed two 3D deep convolutional neural networks, for detecting nodule candidates and reducing false positive nodules.
2. We created a 3D multi-scale pulmonary nodule detection network by embedding the squeeze-and-excitation unit [14] into multi-scale Res2Net [15] module of multiple receptive fields at the same granularity level, and introducing context enhancement module that integrates multi-scale features and spatial attention module that makes the network pay more attention to the regions of interest to improve the detection performance [16].
3. On the basis of multi-scale Res2Net modules and the squeeze-and-excitation units, a 3D false positive reduction network was created. We weighted average the prediction probability obtained by the false positive reduction network and the prediction probability obtained by the nodule candidate detection network to obtain the final result.
4. We validated the proposed method on publicly available LUNA16 dataset. Experimental results showed that our proposed method achieved competitive performance compared to

several state-of-the-art networks. In addition, we conducted extensive ablation validation experiments to demonstrate the effectiveness of the method.

Methods

The automatic detection of pulmonary nodules can be seen as an object detection task with input as CT image I , output as pulmonary nodule location information $[x, y, z, d]$, where $[x, y, z]$ represents the central coordinate of pulmonary nodule cube bounding box, d represents the diameter of pulmonary nodule. Our purpose in this task is to construct a mapping F from I to $[x, y, z, d]$. To achieve this goal, we proposed a 3D multi-scale pulmonary nodule detection network, as shown in Fig 1. The network consists of Bottle2SEneck modules and includes two parts: nodule candidate detection network and false positive reduction network.

Bottle2SEneck

The Bottle2SEneck is the basic module of Res2SENet. Bottle2SEneck involves residual-like connections and a squeeze-and-excitation unit within a single residual block and represents multi-scale features at a granular level. The Bottle2SEneck module is composed of a Res2Net block and a squeeze-and-excitation unit, the structure is shown in Fig 2, where the $x_i, y_i (i = 1, 2, 3, 4)$ represent split feature map, and the $3 \times 3 \times 3$ represents convolution layer with a convolution kernel size of $3 \times 3 \times 3$, each convolution layer is followed by a batch normalization layer and a ReLU layer. The structure of the squeeze-and-excitation unit is shown in Fig 3.

Bottle2SEneck first extracts features from input feature map x using a filter of size $3 \times 3 \times 3$, and splits the output feature map into 4 groups on average according to the channel dimension, which is represented as x_1, x_2, x_3, x_4 in Fig 2. It is worth noting that the spatial size of x_1, x_2, x_3, x_4 are the same. Then the feature map subset x_i is added to y_{i-1} , the output of previous filter K_{i-1} , and fed into the filter K_i to obtain the output feature map y_i , which can be expressed by formula as:

$$y_i = \begin{cases} K_i x_i, & i = 1; \\ K_i (y_{i-1} + x_i), & 2 \leq i \leq 3; \\ x_i, & i = 4. \end{cases} \quad (1)$$

Next, concatenate $y_i (i = 1, 2, 3, 4)$ in the channel dimension. In the Bottle2SEneck, we omit the convolution for the fourth split, which can reuse the features; we use three small filters (convolution kernel size $3 \times 3 \times 3$, channel number $C/4$) instead of a large filter (convolution kernel size $3 \times 3 \times 3$, channel number C) to increase the receptive field of each output feature map, enable the network to fully extract global and local features and have stronger multi-scale representation capability while maintaining a computational load similar to the networks composed of modules with a large filter. The split and concatenate strategy can make convolution more efficient in processing features.

The squeeze-and-excitation unit structure is shown in Fig 3, it consists of two processes: 1) the squeeze process, which integrates global features through adaptive average pooling; 2) the excitation process, which is implemented through fully connected layer FC1-ReLU-fully connected layer FC2-Sigmoid structure, where r is reduction ratio with a value of 16 in this paper. The excitation process can fully capture the interchannel dependence according to the information gathered in the squeeze process, that is, the channel weight which contains abundant nodule information is significant, but the channel weight which does not contain nodule

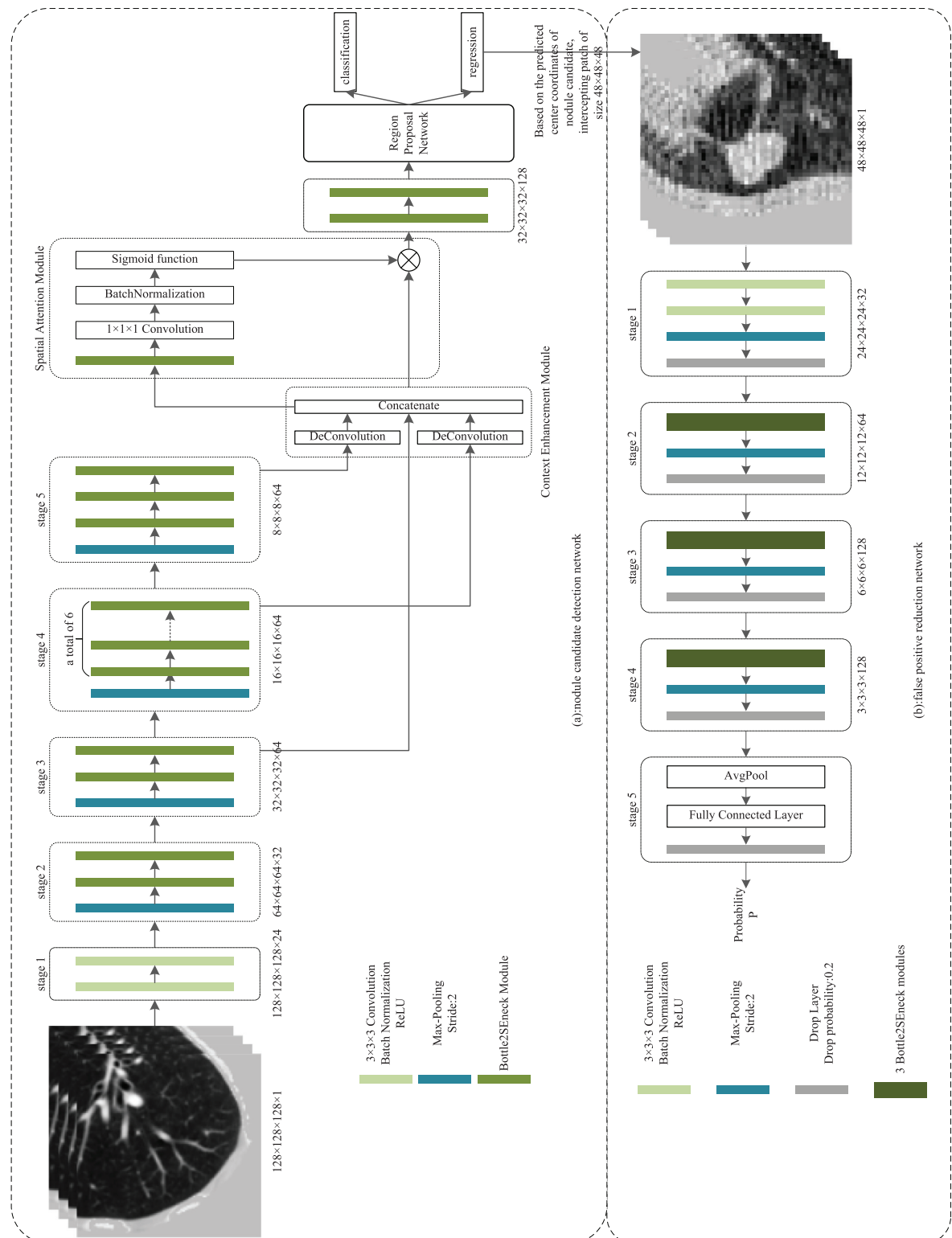


Fig 1. 3D multi-scale pulmonary nodule detection networks. The CT images in Fig 1 were accessed through the links available on this page: <https://luna16.grand-challenge.org/Download/> and had not been previously published and/or copyrighted.

<https://doi.org/10.1371/journal.pone.0244406.g001>

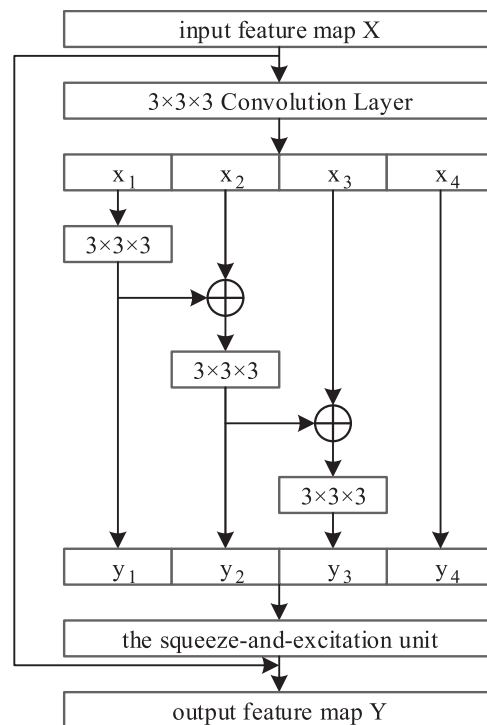


Fig 2. Bottle2SEneck module.

<https://doi.org/10.1371/journal.pone.0244406.g002>

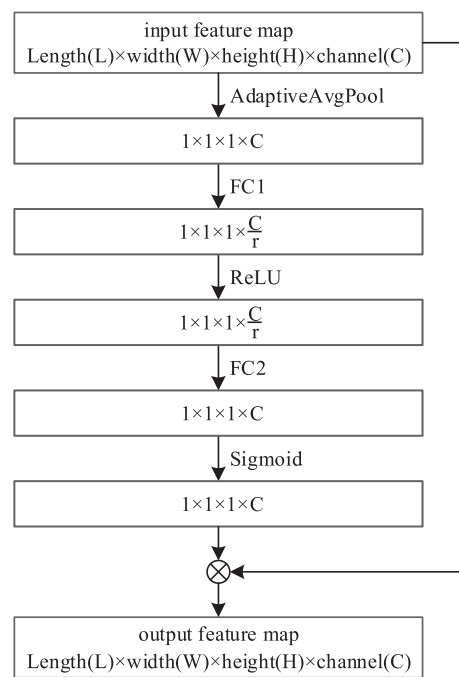


Fig 3. The squeeze-and-excitation unit.

<https://doi.org/10.1371/journal.pone.0244406.g003>

information is small. Finally, the output (the weight of each channel) generated by the excitation process is multiplied with the feature map of the corresponding channel in the initial input to emphasize the characteristics of the pulmonary nodules.

Nodule candidate detection network

Network structure. The proposed network for detecting nodule candidates in low dose CT scanning is shown in Fig 1(a). The network utilizes region proposal network structure [17–19], and according to the characteristics of this detection task, the scales of anchors in the network are set to 5,10,20. Specifically, the network consists of two parts: the backbone part Res2SENet and the detection part. Since Res2SENet is mainly made up of Res2Net modules and the squeeze-and-excitation units(SE), it is named Res2SENet. The input of the network is a cropped CT image cube with dimensions (*length* × *width* × *height* × *the number of channels*) of $128 \times 128 \times 128 \times 1$.

The backbone part Res2SENet of the nodule candidate detection network consists of five stages, the first stage includes two convolutional layers, the second to fifth stages include a max-pooling layer and several Bottle2SEneck modules, respectively, the specific number of modules is shown in Fig 1(a). Among them, the max-pooling layer is used to downsample, reducing the size of the feature map, the Bottle2SEneck module is used to change the number of channels without changing the feature map size. Here we use c_i to represent the output feature map of stage i .

Feature pyramid network (FPN) [20] structure increases computational cost and results in enormous runtime latency due to many additional convolutions and detection branches involved. Therefore, in the detection part of the network, we introduced two effective modules, that is, the context enhancement module (CEM) and the spatial attention module (SAM). CEM can integrate multi-scale feature information and enhance feature discrimination. In the CEM, we respectively upsample c_4 , c_5 through deconvolution, and concatenate the obtained feature maps with c_3 in the channel dimension. Compared with previous FPN structure, our proposed CEM only involves two deconvolution layers and one feature maps concatenation operation, which reduces the computational cost while ensuring the network effect. SAM performs softmax operation on the feature map from upper layer to get the spatial attention map, and multiplies the spatial attention map with feature map from lower layer to make the network pay more attention to the regions of interest.

We added two Bottle2SEneck modules after SAM and set dropout layer [21] to prevent the phenomenon of overfitting. Finally, the output of the dropout layer is taken as the input of the region proposal network. The output of the region proposal network includes the predicted probability p for current anchor being a nodule, and the spatial information (coordinate $[x, y, z]$ and diameter d) of nodule candidates.

Loss function. The binary class label of each anchor box is assigned based on its intersection over union (IOU) with the target nodule. If the IoU is higher than 0.5, the anchor box is labeled as a positive sample; if the IoU is lower than 0.02, the anchor box is labeled as a negative sample; other anchors that are neither positive nor negative will be neglected during training process. Our loss function consists of classification loss and regression loss. For each labeled anchor box, the multitask loss function is defined as follows:

$$L(p_i, t_i) = \lambda L_{cls}(p_i, p_i^*) + p_i^* L_{reg}(t_i, t_i^*). \quad (2)$$

We set the weight parameter λ to 0.5, $L_{cls}(p_i, p_i^*)$ is the classification loss calculated by the binary cross-entropy loss function ($CrossEntropy_{binary}$), $L_{reg}(t_i, t_i^*)$ is the regression loss

calculated by the $smooth_{l_1}$ loss function. $L_{cls}(p_i, p_i^*)$ and $L_{reg}(t_i, t_i^*)$ are defined as follows:

$$L_{cls}(p_i, p_i^*) = p_i^* \log(p_i) + (1 - p_i^*) \log(1 - p_i). \quad (3)$$

$$L_{reg}(t_i, t_i^*) = \begin{cases} 0.5(t_i - t_i^*)^2, & \text{if } |t_i - t_i^*| < 1 \\ |t_i - t_i^*| - 0.5, & \text{otherwise.} \end{cases} \quad (4)$$

In the above formulas, p_i and p_i^* represent the prediction probability and classification label of an anchor box, respectively. If an anchor box is a positive sample, then its classification label p_i^* is 1, and if an anchor box is a negative sample, then its classification label p_i^* is 0. It is easy to see that only positive samples labeled as $p_i^* = 1$ are involved in the calculation of regression loss. t_i and t_i^* represent the predicted relative coordinates and regression label of an anchor box, respectively, they can be defined as:

$$t_i = (\frac{x - x_a}{d_a}, \frac{y - y_a}{d_a}, \frac{z - z_a}{d_a}, \log(\frac{d}{d_a})). \quad (5)$$

$$t_i^* = (\frac{x^* - x_a}{d_a}, \frac{y^* - y_a}{d_a}, \frac{z^* - z_a}{d_a}, \log(\frac{d^*}{d_a})). \quad (6)$$

where (x, y, z, d) are the coordinates and the size of the predicted bounding box, (x^*, y^*, z^*, d^*) are the coordinates and the size of the ground-truth bounding box, (x_a, y_a, z_a, d_a) are the coordinates and the size of the anchor bounding box.

False positive reduction network

A number of false positive nodules are usually produced in the nodule candidate detection stage. To accurately distinguish true nodules from a large number of nodule candidates, we designed a 3D deep convolutional neural network to further classify the nodule candidates produced in the nodule candidate detection stage to reduce false positives. The network structure is shown in Fig 1(b).

The network consists of five stages. We represent the output feature map of stage i in m_i . The size ($length \times width \times height \times the\ number\ of\ channels$) of m_i is shown below stage i in the figure. In this network, we use convolutional layers and Bottle2SEneck modules to change the number of channels, and use max-pooling layers to downsample, reduce the sizes of feature maps, and apply dropout layers to avoid over-fitting phenomenon, and introduce the binary cross-entropy loss function to optimize.

Experiments and results

Datasets

LUNA16 dataset is a subset of the publicly available pulmonary nodule dataset LIDC-IDRI [22]. LUNA16 dataset removes CT images with slice thickness greater than 2.5 mm from LIDC-IDRI dataset, leaving 888 CT images with slice thickness between 0.6 mm and 2.5 mm, spatial resolution between 0.46 mm and 0.98 mm and mean diameter of 8.3 mm. The criterion for determining nodules in the LUNA16 dataset is that at least three out of four radiologists believe that the nodule diameter is greater than 3 mm. Therefore, a total of 1186 positive nodules are annotated in the dataset.

Preprocessing

For the input CT image, we adopt four automatic preprocessing steps:

1. normalize the voxel value range of pulmonary nodules from the original $[-1200, 600]$ to $[0, 1]$, which is convenient for the neural network to extract effective image features, which is expressed as:

$$\overline{val} = \begin{cases} 0, & val < -1200 \\ \frac{val - (-1200)}{600 - (-1200)}, & -1200 \leq val \leq 600 \\ 1, & val > 600. \end{cases} \quad (7)$$

where val represents the CT value before normalization and \overline{val} represents the CT value after normalization.

2. remove background based on the CT pulmonary segmentation images provided by the dataset;
3. resample CT images to an isotropic resolution of $1 \times 1 \times 1$ mm;
4. crop the regions of interest for pulmonary nodules.

First and second lines in Fig 4 show CT images before and after preprocessing, respectively.

Experimental setup

We experimented with eight Intel(R) Xeon(R) Silver 4210 CPUs with the master frequency of 2.20G Hz and the memory of 64 G. In this paper, all network models are built by Python 2.7 and accelerated on 2 NVIDIA GeForce RTX 2080Ti graphics cards by Pytorch parallel computing framework. The networks are all optimized by Stochastic gradient descent (SGD)

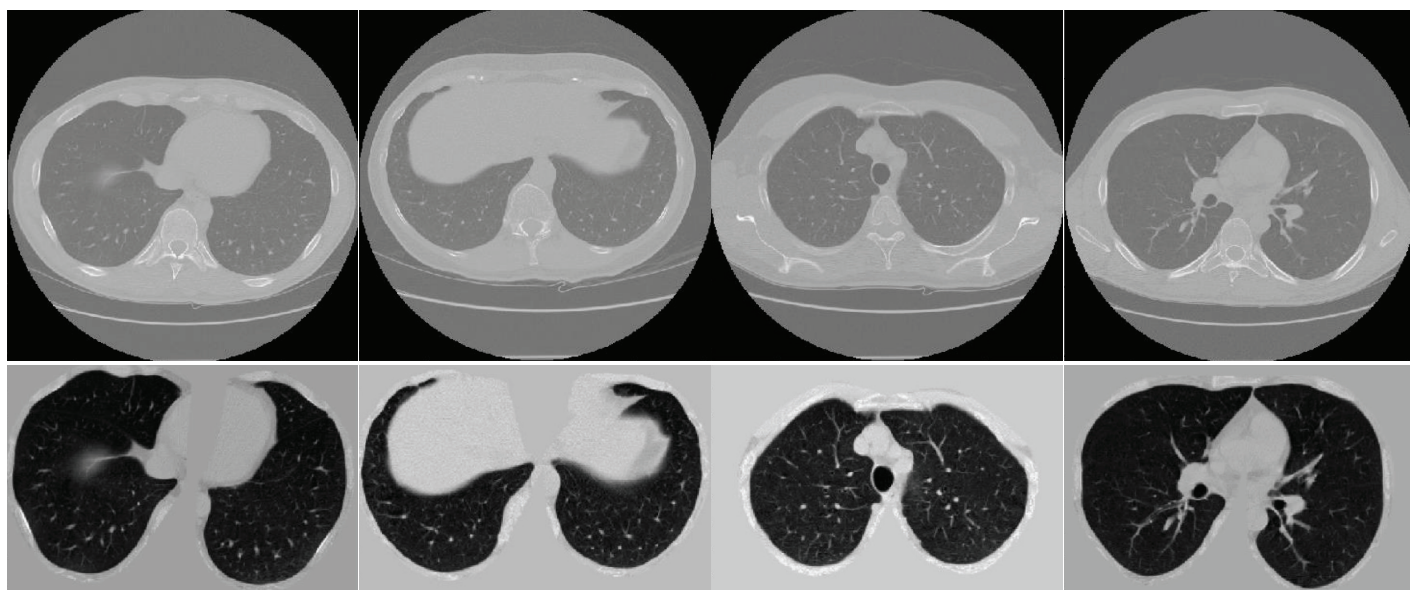


Fig 4. CT images before and after preprocessing. The CT images in Fig 4 were accessed through the links available on this page: <https://luna16.grand-challenge.org/Download/> and had not been previously published and/or copyrighted.

<https://doi.org/10.1371/journal.pone.0244406.g004>

method, where the initial learning rate is set to 0.01, the momentum parameter is set to 0.9, and the weight decay is set to 0.0001. We performed 10-fold cross validation on the dataset.

In the training stage of the nodule candidate detection network, we perform data enhancement after the preprocessing steps by randomly rotating, flipping and cropping, where the cropping scale between 0.75 and 1.25. Set the batch size to 8 and the total number of epochs to 150. After epoch 50, the learning rate is reduced to 0.001, and after epoch 100, the learning rate is reduced to 0.0001. In the testing stage of the nodule candidate detection network, we split the preprocessed CT images into small patches of size $208 \times 208 \times 208$ as inputs to the network, overlapping 32 pixels between neighbouring small patches, that is to say, the distance between the central coordinates of neighbouring small patches is 176. For each CT image, we summarize the nodule candidates obtained by all the small patches, and merge highly overlapping candidates by non-maximum suppression (NMS) [23] with IOU threshold of 0.1 to obtain the detection result.

In the training stage of the false positive reduction network, because the average number of false positive nodules per scan of the nodule candidate detection network is 22, in order to balance the number of positive and negative samples, we amplified the positive samples by 22 times, the amplification methods are consistent with the data enhancement methods in the nodule candidate detection network. We set the batch size to 8 and the total number of epochs to 40. After epoch 10, the learning rate is reduced to 0.001, and after epoch 20, the learning rate is reduced to 0.0001. In the testing stage of the false positive reduction network, the predicted probability obtained by the false positive reduction network is weighted average with the predicted probability obtained by the nodule candidate detection network to get the final classification result, and the calculation formula is as follows:

$$p = \sum_{i=1}^2 \omega_i p_i \quad (8)$$

Among them, ω_i is the weight of the prediction probability p_i of the network, we set the weight of the prediction probability of the nodule candidate detection network to 0.2, and the weight of the prediction probability of the false positive reduction network to 0.8.

Evaluation metrics

Here, the average sensitivity of FROC curve under 7 different false positives (0.125, 0.25, 0.5, 1, 2, 4, 8) is taken as the evaluation result of algorithm performance. The sensitivity formula is as follows:

$$\text{sensitivity} = \frac{TP}{TP + FN} \quad (9)$$

Among them, the TP represents the number of true positive nodules, and the criterion for determining the predicted nodule as true positive nodule is that the center coordinates of predicted nodule are within the ground truth positive nodule. The FN represents the number of false negative nodules, which are the ground truth positive nodules that are not detected.

Results

The performance of the nodule candidate detection network and the whole 3D multi-scale pulmonary nodule detection network are evaluated by FROC curves, average sensitivity, highest sensitivity, and average number of false positives per scan. FROC curves of networks are shown in Fig 5, where the curves are obtained by interpolating true prediction. In Table 1, we

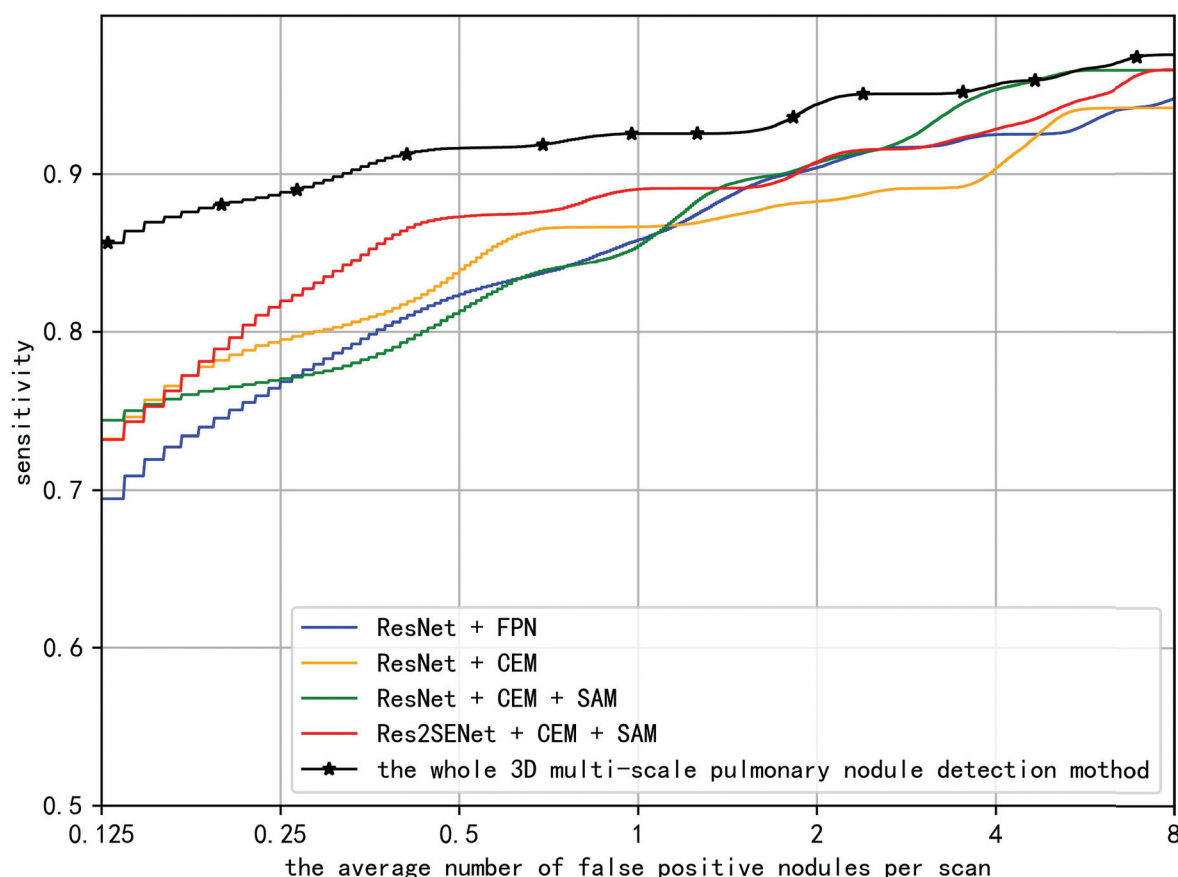


Fig 5. FROC curves.

<https://doi.org/10.1371/journal.pone.0244406.g005>

compared different 3D nodule candidate detection networks by average sensitivity, highest sensitivity, and average number of false positives per scan. In Table 2, we compared the whole pulmonary nodule detection network with the experimental results of others in terms of average sensitivity.

To demonstrate the effectiveness of CEM, SAM and Res2Net backbone network in our proposed network structure, we conducted ablation validation experiments on the dataset. The experimental results are shown in Table 1.

To demonstrate the effectiveness of CEM, we compared the combination of FPN and Residual Network(ResNet) [24] with the combination of CEM and ResNet, the experimental results show that the combination of CEM and ResNet has lower highest sensitivity but the average sensitivity is higher, which proves that the CEM with simple structure has comparable performance with FPN. To demonstrate the effectiveness of SAM, we added SAM after CEM,

Table 1. Performance comparison of different nodule candidate detection networks.

networks	Average sensitivity	Highest sensitivity	FPS per scan	run time cost per epoch (minutes)
<i>ResNet</i> [24] + <i>FPN</i> [20]	0.846	0.975	18	31
<i>ResNet</i> [24] + <i>CEM</i>	0.849	0.958	31	28
<i>ResNet</i> [24] + <i>CEM</i> + <i>SAM</i>	0.856	0.983	26	29
<i>Res2SENet</i> + <i>CEM</i> + <i>SAM(ours)</i>	0.872	0.983	22	34

<https://doi.org/10.1371/journal.pone.0244406.t001>

Table 2. Performance comparison of different methods for pulmonary nodules detection.

methods	Average sensitivity
<i>Khosravan et al. [9]</i>	0.897
<i>Xie et al. [10]</i>	0.790
<i>Wang et al. [12]</i>	0.903
<i>Ding et al. [25]</i>	0.891
<i>Pezeshk et al. [26]</i>	0.832±0.011
<i>Li et al. [27]</i>	0.912
<i>Proposed</i>	0.923

<https://doi.org/10.1371/journal.pone.0244406.t002>

the experimental results show that the average sensitivity of the nodule candidate detection network using SAM increases by 0.7%, the highest sensitivity increases by 2.5%, and the average number of false positives per scan decreases by about 5, indicating that SAM could effectively improve the performance of nodule detection. To demonstrate the effectiveness of Res2SENet backbone network, we combined Res2SENet alternative ResNet, with CEM and SAM. The experimental results show that the average sensitivity of nodule candidate detection network using Res2SENet backbone network increases by 1.6% compared with nodule candidate detection network using ResNet backbone network, and the average number of false positives per scan decreases by about 4, which prove the effectiveness of Res2SENet backbone network.

To demonstrate the effectiveness of the whole 3D pulmonary nodule detection network proposed in this paper, we compared our method with existing state-of-art methods in terms of average sensitivity. The results are shown in Table 2.

For the dataset, the average sensitivity of the proposed method is 0.923, which is higher than the existing state-of-art methods, which shows the superiority of our proposed method. Fig 6 shows the detection results, but because of the 3D nature of CT images, we can only display the slice where the detection center is located. In attention, because pulmonary nodules

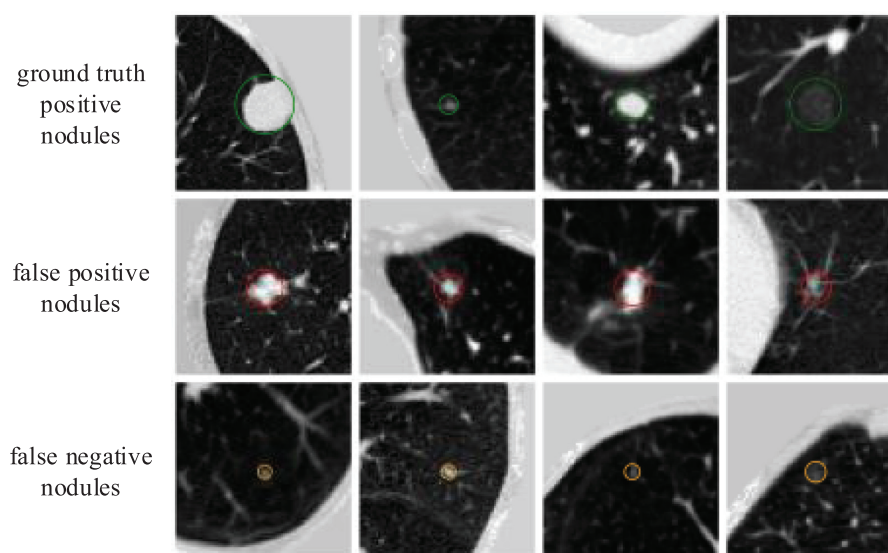


Fig 6. Experimental results. The CT images in Fig 6 were accessed through the links available on this page: <https://luna16.grand-challenge.org/Download/> and had not been previously published and/or copyrighted.

<https://doi.org/10.1371/journal.pone.0244406.g006>

are relatively small in the slices, we only crop the square area with the detection center as the center and the side length of 64 to visualize. Among them, the first line shows the detected true positive nodules, circled with green circles, and the second line shows the detected false positive nodules, which have very similar characteristics to true positive nodules, circled with red circles, and the third line shows the undetected ground truth positive nodules, i.e., false negative nodules. It is not difficult to see that most of them are extremely small in size, and it is possible to improve the detection performance by special data enhancement on these extremely small nodules. Compared with the traditional methods, our proposed method is not only good for the detection of solid nodules, but also good for the detection of ground glass opacity nodules.

Conclusion

This paper proposed a 3D multi-scale pulmonary nodule detection method based on deep convolutional neural network. This method consists of two stages: nodule candidate detection stage and false positive reduction stage. In order to fully extract multi-scale features of pulmonary nodules, we combine the Res2Net module with the squeeze-and-excitation unit to build nodule candidate detection network and false positive reduction network. In addition, in the nodule candidate detection network, in order to integrate high-level semantic information with low-level position information, we proposed a context enhancement module with simple structure but excellent performance; To make the network pay more attention to regions of interest, we introduced the spatial attention module after the context enhancement module. Compared with the existing state-of-art pulmonary nodule detection methods, our proposed method has higher average sensitivity and less false positive nodules, and has practical value in the field of pulmonary nodule detection in chest CT sequence image.

Because the 3D multi-scale pulmonary nodule detection method proposed in this paper still has a few extremely small nodules missed diagnosis, it needs to be further optimized in the future to improve the detection performance of this system, for example special data enhancement for these extremely small nodules. In addition, the system can only output the location information of nodules, but in the actual lung cancer screening, the growth site, edge morphology and internal structure of nodules are of great significance for clinical diagnosis. In the future, the sizes, types and characteristics of nodules can be analyzed to provide suggestions for subsequent works.

Acknowledgments

Thanks to Yanjun Peng for providing hardware support in the laboratory for experiments and thanks to the reviewers and editorial teachers for their busy schedule to review this paper.

Author Contributions

Formal analysis: Huacong Sun, Yanfei Guo.

Methodology: Haixin Peng, Huacong Sun.

Software: Haixin Peng.

Supervision: Haixin Peng, Yanfei Guo.

Validation: Haixin Peng.

Writing – original draft: Huacong Sun.

Writing – review & editing: Yanfei Guo.

References

1. Bray FI, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*. 2018; 68(6):394–424.
2. Wood D, Kazerooni EA, Baum SL, Eapen GA, Ettinger DS, Hou L, et al. Lung cancer screening, version 3.2018. *Journal of The National Comprehensive Cancer Network*. 2018; 16(4):412–441. <https://doi.org/10.6004/jnccn.2018.0020> PMID: 29632061
3. Murphy A, Skalski MR, Gaillard F. The utilisation of convolutional neural networks in detecting pulmonary nodules: a review. *British Journal of Radiology*. 2018; 91 (1090). <https://doi.org/10.1259/bjr.20180028> PMID: 29869919
4. Zhu W, Liu C, Fan W, Xie X. DeepLung: Deep 3D Dual Path Nets for Automated Pulmonary Nodule Detection and Classification. *Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. 2018:673–681.
5. Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2017; 39(6):1137–1149. <https://doi.org/10.1109/TPAMI.2016.2577031>
6. Setio AAA, Traverso A, Bel TD, Berens MSN, Bogaard CVD, Cerello P, et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge. *Medical Image Analysis*. 2017:1–13. <https://doi.org/10.1016/j.media.2017.06.015> PMID: 28732268
7. Dou Q, Chen H, Jin Y, Lin H, Qin J, Heng P. Automated Pulmonary Nodule Detection via 3D ConvNets with Online Sample Filtering and Hybrid-Loss Residual Learning. *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2017:630–638.
8. Shelhamer E, Long J, Darrell T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2017; 39(4):640–651. <https://doi.org/10.1109/TPAMI.2016.2572683>
9. Khosravan N, Bagci U. S4ND: Single-Shot Single-Scale Lung Nodule Detection. *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2018:794–802.
10. Xie H, Yang D, Sun N, Chen Z, Zhang Y. Automated pulmonary nodule detection in CT images using deep convolutional neural networks. *Pattern Recognition*. 2019:109–119.
11. Dou Q, Chen H, Yu L, Qin J, Heng P. Multilevel Contextual 3-D CNNs for False Positive Reduction in Pulmonary Nodule Detection. *IEEE Transactions on Biomedical Engineering*. 2017; 64(7):1558–1567. <https://doi.org/10.1109/TBME.2016.2613502>
12. Wang J, Wang J, Wen Y, Lu H, Niu T, Pan J, et al. Pulmonary Nodule Detection in Volumetric Chest CT Scans Using CNNs-Based Nodule-Size-Adaptive Detection and Classification. *IEEE Access*. 2019:46033–46044. <https://doi.org/10.1109/ACCESS.2019.2908195>
13. Szegedy C, Ioffe S, Vanhoucke V, Alemi AA. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. *Proceedings of the Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*. 2016:4278–4284.
14. Hu J, Shen L, Albanie S, Sun G, Wu E. Squeeze-and-Excitation Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2019:1–1.
15. Gao S, Cheng M, Zhao K, Zhang X, Yang M, Torr PHS. Res2Net: A New Multi-scale Backbone Architecture. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2019:1–1.
16. Qin Z, Li Z, Zhang Z, Bao Y, Yu G, Peng Y, et al. ThunderNet: Towards Real-Time Generic Object Detection on Mobile Devices. *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision*. 2019:6718–6727.
17. Yuan J, Xue B, Zhang W, Xu L, Sun H, Zhou J. RPN-FCN based Rust detection on power equipment. *Procedia Computer Science*. 2019:349–353.
18. Tong G, Chen H, Li Y, Du X, Zhang Q. Object detection for panoramic images based on MS-RPN structure in traffic road scenes. *Iet Computer Vision*. 2019; 13(5):500–506. <https://doi.org/10.1049/iet-cvi.2018.5304>
19. Yang D, Zou Y, Zhang J, Li G. C-RPNs: Promoting object detection in real world via a cascade structure of Region Proposal Networks. *Neurocomputing*. 2019:20–30.
20. Lin T, Dollar P, Girshick R, He K, Hariharan B, Belongie S. Feature Pyramid Networks for Object Detection. *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*. 2017:936–944.

21. Srivastava N, Hinton GE, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*. 2014; 15(1):1929–1958.
22. Armato SG, McLennan G, Bidaut L, McNittgray MF, Meyer CR, Reeves AP, et al. The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A Completed Reference Database of Lung Nodules on CT Scans. *Medical Physics*. 2011; 38(2):915–931. <https://doi.org/10.1118/1.3528204> PMID: 21452728
23. Rothe R, Guillaumin M, Gool LV. Non-Maximum Suppression for Object Detection by Passing Messages between Windows. *Proceedings of the Asian Conference on Computer Vision*. 2014:290–306.
24. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*. 2016:770–778.
25. Ding J, Li A, Hu Z, Wang L. Accurate Pulmonary Nodule Detection in Computed Tomography Images Using Deep Convolutional Neural Networks. *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2017:559–567.
26. Pezeshk A, Hamidian S, Petrick N, Sahiner B. 3-D Convolutional Neural Networks for Automatic Detection of Pulmonary Nodules in Chest CT. *IEEE Journal of Biomedical and Health Informatics*. 2019; 23(5):2080–2090. <https://doi.org/10.1109/JBHI.2018.2879449>
27. Li F, Huang H, Wu Y, Cai C, Huang Y, Ding X. Lung Nodule Detection with a 3D ConvNet via IoU Self-normalization and Maxout Unit. *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*. 2019:1214–1218.