LungSeek: 3D Selective Kernel residual network for pulmonary nodule diagnosis

Haowan Zhang^{1,2} · Hong Zhang^{1,2}

Accepted: 17 November 2021 / Published online: 27 January 2022

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

Early detection and diagnosis of pulmonary nodules is the most promising way to improve the survival chances of lung cancer patients. This paper proposes an automatic pulmonary cancer diagnosis system, LungSeek. LungSeek is mainly divided into two modules: (1) Nodule detection, which detects all suspicious nodules from computed tomography (CT) scan; (2) Nodule Classification, classifies nodules as benign or malignant. Specifically, a 3D Selective Kernel residual network (SK-ResNet) based on the Selective Kernel Network and 3D residual network is located. A deep 3D region proposal network with SK-ResNet is designed for detection of pulmonary nodules while a multi-scale feature fusion network is designed for the nodule classification. Both networks use the SK-Net module to obtain different receptive field information, thereby effectively learning nodule features and improving diagnostic performance. Our method has been verified on the luna16 data set, reaching 89.06, 94.53% and 97.72% when the average number of false positives is 1, 2 and 4, respectively. Meanwhile, its performance is better than the state-of-the-art method and other similar networks and experienced doctors. This method has the ability to adaptively adjust the receptive field according to multiple scales of the input information, so as to better detect nodules of various sizes. The framework of LungSeek based on 3D SK-ResNet is proposed for nodule detection and nodule classification from chest CT. Our experimental results demonstrate the effectiveness of the proposed method in the diagnosis of pulmonary nodules.

Keywords Deep learning · Medically assisted diagnosis · Pulmonary nodules · Residual network · 3D Selective Kernel network

1 Introduction

With the aggravation of global air pollution and the increase in the number of smokers among the population, the incidence and mortality of lung cancer have been on the rise in recent years. Lung cancer is one of the malignant tumors with the fastest increasing mortality and morbidity [8], and it is also one of the malignant tumors that pose the greatest threat to human health and life [35]. Lung cancer early lesions are in the form of pulmonary nodules, which are also one of the important markers in lung cancer diagnosis. Detecting pulmonary nodules in the early period is very critical for patient

care, which will increase the 5-year overall survival rate to 52% [34]. Therefore, accurate detection of lung nodules and further analysis of the extracted positive nodules will help to better discover and diagnose lesions, which is also the key to early detection of lung cancer.

The most commonly used lung health check is now in CT. As shown in Fig. 1, there are various shapes of lung nodules in CT images, Fig. 1a represents nodules in a slice of CT image while Fig. 1b indicates the nodule zoomed in it. Experts observe and diagnose CT images of the lungs to determine whether the patient has lung disease.

During traditional radiological diagnosis, doctors need to find suspicious lesions from thousands of chest CT image sequences. Repeated reading work for a long time can easily lead to visual fatigue, which may lead to misdiagnosis or missed diagnosis. Therefore, by using computer technology to assist doctors in detecting and identifying lung nodules, radiologists can reduce their burden [38]. When the computer assist radiologists in reading chest CT, some nodules that

✉ Hong Zhang
zhanghong_wust@163.com

¹ College of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430081, China

² Hubei Province Key Laboratory of Intelligent Information Processing and Real-Time Industrial System, Wuhan, China

Fig. 1 Examples of Lung Nodule in LUNA16 dataset, each diagram represents pulmonary nodule in a slice

are prone to develop into lung cancer are often difficult to distinguish because they show similar benign lesions [36]. It is possible that the detection of lung nodules will interpret non-lesions as lesions, or misunderstand benign lesions as malignant, leading to false positive results. Therefore, the classification of benign and malignant pulmonary nodules is also necessary for the lung diagnosis system. The use of CAD to obtain the diagnosis of the location, size and confidence of the lung nodules in CT scans can provide a reference for doctors when evaluating patients' CT, thereby reducing the workload of radiologists in a large number of repetitive tasks.

In last several years, many CAD methods have been introduced into the diagnosis of lung nodules on CT images. They generally include two stages: (1) candidate lung nodule detection, (2) lung nodule benign and malignant classification. The first stage usually requires the manual design of features, such as the morphological features of nodules and pixel thresholds in the traditional CAD method. Recently, deep convolutional networks such as Faster R-CNN and Mask R-CNN have been used to generate candidate bounding boxes [9] 31, 12. In the second stage, more complex features and more advanced methods are needed, such as carefully designed texture features, to remove false positive nodules. The attention mechanism is used to classify [19]. Or use a 3D dual path network (DPN) as the neural network structure to extract the features of the original CT nodules, and then use GBM as a classifier to classify [39].

In this paper, we first preprocessed the datasets, extracted the lung parenchyma and stored the preprocessed data in .npy format. Then detected the possible pulmonary nodules using the model of the nodule detection network, and select candidate lung nodules in the top-5 with a confidence level greater than 0.5 for the next step. Finally classified these detected nodules into either malignant or benign with classification subnetwork. In order to be able to automatically detect nodules and correctly diagnose multiple sizes of nodules efficiently, we proposed a 3D Selective Kernel residual network for pulmonary nodule diagnosis, named LungSeek. For the nodule detection section, we proposed a 3D region

proposal network with 3D SK-ResNet to effectively detect deep features, which uses a U-net shape structure. For the nodules classification section, the confidence fusion of the detection part combined with the classification results of the multi-scale feature fusion network model was used to make a final diagnosis.

Finally, our LungSeek system is divided into two parts: detection of pulmonary nodules from lung CT and classification of benign and malignant nodules, while two deep 3D convolution neural networks (3DCNNs) are designed for them. We designed a 3D region proposal network with 3D SK-ResNet and a U-net shape structure to detect lung nodules. Since the classification task in this article is to classify the lung nodule of benign and malignant, the proposal of the 3D regional proposal network will be directly regarded as the result of the detection part. Considering that the location of lung nodules is mostly in a non-fixed state, and the shape and size of different lung nodules are also very different, the corresponding characteristics of the nodules need to be selected from different ranges. Convolution kernels of different scales represent different ranges of receptive fields, which can obtain more comprehensive nodule feature information. So, we propose a multi-scale feature fusion network to classify the detected nodules. The multi-scale convolution operation is used for feature extraction and feature fusion of different ranges of input CT images of lung nodules, which solves the problem of incomplete feature extraction. The SK-ResNet module adaptively adjusts the size of the receptive field according to multiple scales of the input information, solves the problem of feature information loss, and efficiently obtains deeper features, then the diagnosis system outputs the classification results of benign and malignant lung nodules.

2 Related work

In recent years, computer-aided detection and diagnosis (CAD) system [4] has begun to be applied to the study of medical imaging. It uses digital image processing, computer vision, pattern recognition and other crossover technologies to quickly and accurately screen out the suspected lesions. For example, in order to assist in the diagnosis of the COVID-19 outbreak that has occurred in the past two years, machine and deep learning models are used for COVID-19 detection [28]. In terms of early lung cancer screening, computer-aided diagnosis system can help radiologists to quantitatively detect regional features and identify suspicious pulmonary nodules, which not only reduces the workload of doctors, but also provides reliable reference information for reading. It helps radiologists to reduce the high rate of misdiagnosis in manual diagnosis to a certain extent, improves the detection accuracy of lung nodules, and has important theoretical and practical value.

Convolutional neural networks (CNN) have been widely used in the field of image processing, image segmentation [17], medical image detection [2, 10, 11]. Since deep learning was proposed and widely used, Qi et al. [26] proposed a method that uses a 3D CNN to reduce false positives nodules detected by CT images. By using 3D images as training input to extract more feature structures, a simple and effective multi-layer background information coding strategy is proposed. And Jiang et al. [16] proposed a CT-ReCNN. To reduce the noise in the lung CT image and extract more detailed features. Their experimental results proved the effectiveness of applying multi-layer background information to a three-dimensional CNN network to automatically detect pulmonary nodules through volume CT data.

The R-CNN is often used for the segmentation and recognition of lung nodules. With R-CNN proposed by Girshick et al. [9], they use 3D CNN network to extract the region proposal and proposed the Faster R-CNN model of target detection in 2016 [31]. This model concentrates feature extraction, classification and regression problems in a network model, which is a kind of deep learning network [31], it also contain the ZF model [25], VGG16 model [5]. And Mask R-CNN [12] puts forward a new idea for the selection of candidate boxes, which is no longer the previous method of sliding window and pyramid, but the selection of Windows in the ratio of 1:1, 1:2 and 2:1, which speeds up the detection speed and can effectively solve the machine learning problem in the case of large samples. Zhu et al. [39] proposed a method combining a 3D Faster R-CNN with DPN [3] for nodule detection and classification, where DPN taken the advantage of ResNet [13] and DenseNet [15].

U-net [27] is also widely used in nodule detection. In order to evaluate the complex relationship between nodule morphology and cancer, Liao et al. [22] used an improved U-net as a public axis network and proposed a 3D deep neural network for pulmonary nodules detection and output all suspicious nodules. They also used Leaky noise-or gates to integrate the results to obtain the likelihood of lung cancer. Their proposed model also won first place in the 2017 Data Science Bowl competition.

Recently, a multi-view feature pyramid network (FPN) [23] was also applied to pulmonary nodule detection. MVP-net [19] proposes a (FPN), which extracts multi-view features from images rendered at different window widths and levels. In order to combine, this multi-view information effectively, they also proposed a location aware attention module to alleviate the problem of small data sample size.

Although CADs based on ConvNet for pulmonary nodule diagnosis have achieved good results, most networks (ResNet, ResNeXt, DPN, DenseNet, etc.) improve the performance of the network by changing the spatial dimension of the network. In 2017, the squeeze-and-excitation network [14] introduced an attention mechanism between channels

by considering the interdependence between feature channels to improve segmentation accuracy. It has been proven to be useful in target detection and image classification, especially medical images analysis has developed. Furthermore, the Selective Kernel network [20] proposed by Li et al. in 2019 uses a method of nonlinear aggregation of different scales to obtain information of different receptive fields to achieve adaptive receptive fields. Combining SK-Net with the residual network to extract lung nodule features from the dimensions of space and channel at the same time, can improve the effectiveness of our lung nodule detection. As far as we know, the effectiveness of SK-ResNet in the diagnosis of lung nodules has not been explored.

In this paper, due to the good performance of Selective Kernel network [20], and inspired by 3D faster R-CNN, we propose a 3DCNN based on 3D region network and SK-ResNet as the detection network. Then a multi-scale feature fusion network is designed to extract deep feature from different scales to efficiently classify malignant or benign nodule.

Our method has the following contributions: (1) We propose two deep 3DCNNs for nodule detection and classification. For the nodule detection part, we use 3D SK-ResNet as the backbone network to extract features from the dimensions of space and channel at the same time, which obtains a better detection effect than the residual network. And inspired by U-net, we propose a 3DCNN based on 3D region network and SK-ResNet with a U-net shape structure helps better obtain the features of small pulmonary nodules. (2) In the nodule classification part, due to the difference size of lung nodules, we propose three different sizes of convolution kernels to correspond to different ranges of feature extraction, and fuse those different splicing features, which solves the problem of incomplete feature extraction. And SK-ResNet is also used in classification network to fully utilize the channel attention mechanism to solve the problem of feature information loss effectively.

3 LungSeek framework

3.1 LungSeek system for Lung CT cancer diagnosis

The architecture of the LungSeek is shown in Fig. 2. During the nodule detection part, after preprocessing, the CT image is extracted into lung parenchyma. Then the preprocessed dataset is cut into patches of $128 \times 128 \times 128$ as the input of detection network. In the nodule classification part, after a multi-scale convolution layer and padding operation, these $32 \times 32 \times 32$ features are concatenated together in channel dimension. Next, put the multi-scale fusion features into 4 SK-ResNet blocks to obtain depth features. After AvgPooling and softmax layers, malignant and benign nodules are classified.

Fig. 2 The framework of LungSeek

Figure 2 outlines the overall structure of LungSeek. The framework of the method mainly includes followed steps. First, we obtain the original CT image and preprocess it, unify the resolution, remove the noise, cavity and other interfering factors. And we extract the lung parenchyma to reduce the search space of the image. Then as shown in the nodule detection part, a 3D region proposal network with 3D SK-Resnet and a U-net shape structure were used to extract the features of lung nodules, which contain the three-dimensional coordinates, diameter and confidence score of these detected nodules. Next, using the extracted lung nodule detection results, center crop the nodule coordinates to obtain a $32 \times 32 \times 32$ size image for each nodule, and input it into the feature extraction part. The top-5 nodules in confidence score are selected to carry out multi-scale convolution operation. After going through the first convolutional layer of three convolution kernels of different sizes ($1 \times 1 \times 1$, $2 \times 2 \times 2$, $3 \times 3 \times 3$), the padding operation is performed, respectively, so that the obtained features that are equal to the input size of the convolutional layer. Then, we use SK-ResNet blocks to extract higher-level semantic information. Lastly, global average pooling and softmax functions are used for benign and malignant classification.

3.2 Nodule detection network

3.2.1 3D SK-ResNet module

3D SK-ResNet is made of 3D residual network and Selective Kernel module, it can effectively solve the eliminating gradient disappearance problem by using the quick connection of

residual learning [13] and obtain the recalibration feature by using SK block (Selective Kernel block). At the same time, the SK block can automatically acquire the importance of each feature channel through learning, so it can selectively stress useful features and inhibit less useful features.

As shown in Fig. 3, Fig. 3a is the Original 3D Residual block, Fig. 3b illustrates the Selective Kernel Residual block. Figure 3c represents the detail of Selective Kernel Convolution. The 3D residual block with bottleneck structure is composed of $1 \times 1 \times 1$, $3 \times 3 \times 3$, $1 \times 1 \times 1$ structure. The basic 3D residual network (3D ResNet) module and the 3D SK-ResNet module are demonstrated in Fig. 3a, b. In Fig. 3b, the basic structure of the SK-ResNet block is to replace the $3 \times 3 \times 3$ convolution in Fig. 3a with SK Conv (Selective Kernel convolution). The structure of SK conv is illustrated in Fig. 3c.

In the *Split* part, for the input feature map x , kernel size is, respectively, $3 \times 3 \times 3$ and $5 \times 5 \times 5$ to obtain U_1 and U_2 . In order to improve the efficiency of experiment, the traditional $5 \times 5 \times 5$ kernel convolution is replaced by the dilated convolution with a $3 \times 3 \times 3$ kernel and dilation size of 2.

In the *Fuse* part, we mainly use the gating mechanism to selectively filter the output of the previous layer, so that each branch carries a different flow of information into the next neuron. First, fuse the output of different branches, that is, add element by element:

$$U = U_1 + U_2, \quad (1)$$

then, global average pooling operations are performed on the two outputs to obtain global information on each channel.

Fig. 3 The detailed structure of deep 3D Selective Kernel residual network

Specifically, calculate the c -th element of S by shrinking U through space dimensions $L \times H \times W$:

$$S_c = F_{gp}(U_c) = \frac{1}{L \times H \times W} \sum_{i=1}^L \sum_{j=1}^H \sum_{k=1}^W U_c(i, j, k), \quad (2)$$

and a fully connect is used on the output S to find the proportion of each channel to create $z \in R^{t \times 1}$.

In the *Select* part, the soft attention between channels can select information of different sizes, which is guided by the compact feature information Z , and the softmax operation is applied in channel-wise:

$$m_c = \frac{e^{M_{c^z}}}{e^{M_{c^z}} + e^{N_{c^z}}}, n_c = \frac{e^{N_{c^z}}}{e^{M_{c^z}} + e^{N_{c^z}}}, \quad (3)$$

where $M, N \in R^{C \times t}$ and m, n denote the soft attention vector for U_1 and U_2 . Note m_c is the c -th element of M , and $M_c \in R^{1 \times t}$ is the c th row of m , so as n_c and N_c . The final feature map V is obtained through the attention weights on various kernels:

$$V_c = m_c \cdot U_{1c} + n_c \cdot U_{2c}, m_c + n_c = 1, \quad (4)$$

where $V = [V_1, V_2, \dots, V_c], V_c \in R^{L \times H \times W}$.

3.2.2 Nodule candidate detection

For the nodule detection part, a 3D region proposal network is designed as the backbone network, using the 3D SK-ResNet as part of the network. We also designed the structure with a U-net shape, which includes encoder and decoder modules. This 3DCNN is a 3D region proposal network with 3D SK-ResNet and a U-net shape structure, demonstrated in Fig. 4. The predicted z, y, x, d and P of each position are the three-dimensional coordinates, diameter and confidence of the nodule, respectively.

Due to the large size of CT images, it cannot directly be taken as the input of the model under the limitation of GPU memory. Therefore, the original data should be cut into 3D patches before input into the target detection network for training. First, we cropped the image to a cube of $128 \times 128 \times 128$ (height, length, width, channel) pixel size as input to the network.

In part of the detection network, we use the U-net shape structure, it is because the U-net network's structure can obtain multi-scale information of images in the training pro-

Fig. 4 Pulmonary nodule detection network of LungSeek

cess. As for the object of patches input in the original image, U-net is suitable for overlap due to network structure, and the surrounding overlap part can provide text information for the edge part of the segmentation area. In addition, 3D region proposal network can carry out multi-scale learning in a pixel-level way. Since the distribution of nodules is characterized by large size variation, such network integration can make the process of generating candidate nodules more accurate and effective.

For the incoming $128 \times 128 \times 128 \times 1$ cubes, the number of channels is changed to 24 by first passing through the convolutional layer with a convolution kernel of $3 \times 3 \times 3 \times 24$ (channel = 24). The down-sampling part of the U-net structure is composed of 4 groups of SK-ResNet blocks (2 SK-ResNet blocks for each group). In the part of up-sampling, we use the deconvolutional layer and SK-ResNet block to process the feature map. The up-sampling process consists of two deconvolutional layers and two sets of SK-ResNet blocks in tandem alternately. A dropout layer with a 0.5 dropout probability is used after the last SK-ResNet block in the up-sampling. Then a $32 \times 32 \times 32 \times 15$ convolution layer is used for the last layer of the network, so the output size is adjusted to $32 \times 32 \times 32 \times 3 \times 5$ (height, length, width, anchors, regressors). For the output proposal, we design three anchors: 5, 10, 30 to refer to the different

sizes of nodules. For a typed image, it corresponds to $32 \times 32 \times 32 \times 3$ anchors.

For each anchor i , the loss function has 5 parts: $(\hat{o}, \hat{h}_x, \hat{h}_y, \hat{h}_z, \hat{h}_d)$, and the first quantity uses a sigmoid function $\hat{p} = \frac{1}{1+\exp(-\hat{o})}$, $\hat{p}$ is the predicted probability of anchor i as a nodule, and $\hat{p}_i$ is the predicted probability for anchor i being a nodule currently.

Denote the predicted nodule coordinates and diameter in the original space by (A_x, A_y, A_z, A_d) , the bounding box of anchor i by (T_x, T_y, T_z, T_d) , and the ground truth bounding box of a target nodule by $(\hat{T}_x, \hat{T}_y, \hat{T}_z, \hat{T}_d)$. Intersection over Union (IoU) is used to determine the label of each positioning box. If the IoU of the anchor i and the ground truth bounding box overlaps higher than 0.5, i is treated as a positive anchor point ($p_i = 1$). Otherwise, if anchor i has no IoU with ground truth bounding boxes higher than 0.02, we consider it to be negative ($p_i = 0$). Other anchors that are neither positive nor negative will be ignored during training. In order to prove the superiority of the method in this paper, besides set the IoU threshold of the positive sample to 0.5 for comparison experiments, we also added two comparative experiments in Sect. 4.3 where the threshold of IoU is set of 0.7 and 0.9. The experimental results prove the advantages of our network under different IoU threshold settings.

We define the loss function of each anchor i as:

$$L(p_i, g_i) = L_{cls}(\hat{p}_i, p_i) + p_i L_{reg}(g_i, \hat{g}_i), \quad (5)$$

where L_{cls} defined the classification loss, and we used binary cross entropy loss function for it:

$$L_{cls}(\hat{p}_i, p_i) = p_i \log(\hat{p}_i) + (1 - p_i) \log(1 - \hat{p}_i), \quad (6)$$

and total regression loss is defined by L_{reg} , we use smooth l_1 regression loss function for it:

$$L_{reg}(g_i, \hat{g}_i) = \begin{cases} (g_i - \hat{g}_i)^2, & \text{if } |g_i - \hat{g}_i| < 1, \\ |g_i - \hat{g}_i|, & \text{else.} \end{cases}, \quad (7)$$

g_i is the bounding box regression labels, which is defined by:

$$g_i = \left(\frac{T_x - A_x}{A_x}, \frac{T_y - A_y}{A_y}, \frac{T_z - A_z}{A_z}, \log\left(\frac{T_d}{A_d}\right) \right), \quad (8)$$

and we define (G_x, G_y, G_z, G_d) to be the ground truth bounding box of a target nodule, and for the position of ground truth nodule, $\hat{g}_i$ is defined as:

$$\hat{g}_i = \left(\frac{G_x - A_x}{A_x}, \frac{G_y - A_y}{A_y}, \frac{G_z - A_z}{A_z}, \log\left(\frac{G_d}{A_d}\right) \right), \quad (9)$$

3.3 Nodule classification network

Due to the wide range of morphological changes of pulmonary nodules, in the classification of nodules, a method that can more effectively extract 3D volume characteristics is needed. We use a multi-scale feature fusion network to perform feature extraction and feature fusion splicing in different ranges on the input CT images of lung nodules to solve the problem of incomplete feature extraction. For each CT image, top-5 proposals are picked as the input to the classification network according to the proposals' confidence score. The SK-ResNet is used for deep feature extraction to classifies the detected lung nodule.

According to the detected nodules' coordinate center, $32 \times 32 \times 32$ cubes were clipped as the input of the classification network. Then these cubes are convolved in the first convolutional layer with three different convolution kernels ($1 \times 1 \times 1$, $3 \times 3 \times 3$, $5 \times 5 \times 5$, stride = 2) to extract characteristic information of pulmonary nodules in different scales. Use padding operation to keep the output feature map size ($32 \times 32 \times 32$) consistent with the input. After group a variety of different scale's feature, matching the feature in the channel dimension, the fusion feature was input to the three consecutive SK-ResNet modules to extract a higher-level of semantic information. Finally, after global average pooling and full connection operation, the softmax function was used to output benign and malignant classification results.

4 Experiment

4.1 Datasets and experiment settings

The proposed method was evaluated on the LUNA16 dataset[30]. LUNA16, fully known as Lung Nodule Analysis 16, is a lung nodule detection data set introduced in 2016. And it is also a subset of the publicly available Lung tubercle data set LIDC-LDRI [24]. The LIDC-LDRI includes multiple doctors' annotation of pulmonary nodules' size, edge information, location of nodules, texture features diagnostic results, etc.[3, 29]. LUNA16 data set selects nodules marked by at least three experts as the final area to be detected. The dataset consisted of 888 CT images corresponding to 888 patient samples, in which a total of 1186 nodules were identified [1].

In addition, LUNA16 removes the CT images with a thickness of more than 3 mm, inconsistent layer spacing or missing layer from the LIDC-LDRI, and only the detection annotation is retained. Therefore, the section thickness of LUNA16 is < 3 mm, and only nodules with a diameter of > 3 mm are taken as samples, while nodules with a diameter of < 3 mm and non-nodules are not included. Most nodules are 4 to 8 mm in diameter, with an average diameter of 8.3 mm. The diameter distribution of labeled nodules is demonstrated in Fig. 5.

Through the experimental demonstration of the LungSeek system in this paper, we now randomly and equally split LUNA16 datasets into ten subsets, of which nine subsets are taken as the training set (800CT) and one subset (88 CT) as the test set, to carry out tenfold cross-validation of the test model. There are 36,378 nodules were marked in the 888 CT images, in which 1186 nodules were selected as the last

Fig. 5 The diameter distribution of pulmonary nodules of Lung Nodule Analysis 16 datasets

nodule area to be detected, and the rest 35,192 nodules were ignored as irrelevant findings.

During the training of detection network, for the verification of each fold, we also performed some augmentations on the data. In order to solve the problem of too few positive samples in the experimental, and efficiently expand the number of positive samples, the data are enhanced by randomly flipping the image and scaling the image in proportion of 0.75 to 1.25. The imbalance of positive and negative samples can be alleviated while effectively increasing the training set of data.

At the same time, we also collected 40 CT scans of patients from hospital in order to further verify the performance of our detection network. This collected data set is prepared contains 40 CT images, with 78 nodules identified in it, which will be used as a test set to evaluate the performance of the lung nodule detection network. Medical images are usually stored in the DICOM file format, but for ease of reading and use, they are often converted into one .mhd file and a .raw file for each patient. The paired .mhd and .raw format files have the same name, the .mhd file stands for meta header data, while .raw file storage pixel information according to header information. We first convert the incoming collected data from CT format to .mhd and .raw format before preprocessing, then use the same preprocessing process as the luna16 data set on it. While LUNA16 datasets are stored in .mhd and .raw format, this step can make the collected CT images have the same data format as LUNA16.

In the classification of lung nodules, for the nodules in the LUNA16 data set, we extracted the corresponding nodule labels from the LIDC-IDRI data set. The degree of malignancy of all nodules is judged by radiologists between grades 1 to 5. The higher the evaluation grade, the higher the degree of malignancy and the greater the risk of cancer. We averaged the doctors' scores. For nodules with a score greater than 3, we mark them as malignant, otherwise they are marked as benign. Because of the uncertainty of the unknown grade nodules with a score of 3, nodules with an average score of 3 were deleted. After treatment, a total of 1004 nodules remained, of which 450 were malignant. We use the same cross-validation split as LUNA16 for classification.

The experiments involved in this paper use Windows operating system and NVIDIA GeForce GTX 1080 GPU, while the computer equipped with an AMD Ryzen 5 3500X CPU and 16 GB of RAM, and the framework of our experiments are implement with Python's Pytorch deep learning library. Under the limit by GPU memory, the batch size parameter set to 4. In addition, we adopt the learning rate attenuation strategy. And during the training, we use the stochastic gradient descent optimizer in order to optimize the model. The momentum of stochastic gradient descent was 0.9, and we set the weight attenuation coefficient as 0.0001. For each model, 250 rounds of training were set up. The initial learning rate

is set as 0.01, when running to 100 batches, the learning rate becomes 0.001; 0.0005 after batch 150; after 200 batches, it is 0.0001.

4.2 Preprocessing

As demonstrated in Fig. 6. represents the preprocessing of the normal lung image, and 2. represents the slice at the top of the CT sequence. Figure 6a transforms the image into HU, Fig. 6b thresholding and binarization of the image, Fig. 6c standardization of gray scale, Fig. 6d erosion and expansion of the lung to remove the small cavity, Fig. 6e convex enveloped the image, Fig. 6f mask was applied to the image.

The pretreatment is mainly as the following steps:

4.2.1 Step1: HU Transformation

The value of the original CT image is first converted into HU value (Hounsfield Unit) as Fig. 6a, which is the standard quantity to describe the radiation intensity and reflects the degree of X-ray absorption of various tissues. Each organization has a specific HU range, which is the same for different people.

4.2.2 Step2: Mask extraction

For each slice in the image, the gaussian filter (standard deviation is 1) is first used to filter, and we use -600 is as the threshold to binarize the image, as demonstrated in Fig. 6b. Then find the boundary of the mask, which is the edge of the non-zero part, to get a box. As the resolution of the original CT is often inconsistent, it is further resampled and the new resolution is applied to unify the resolution of the image data.

4.2.3 Step3: Grayscale standardization

As the HU in the lungs is around -500 , the original HU in the image data is retained (from air to bone) between $[-1200, 600]$, and the area beyond this range can be considered as irrelevant. The data are then threshold and normalized to a linear transformation to $[0, 255]$. The normalized HU value of water was 170, and the normalized HU value of bone tissue was 210. As Fig. 6c shows, since bone tissue outside the mask can easily be mistaken for calcified nodules, we fill the area outside the mask to 170.

4.2.4 Step4: Erosion and expansion

Since there are small voids in the lungs and shadows of other tissues in the mask, we apply a corrosion procedure to the mask to remove the voids in the lungs and then inflate the mask to bring it back to its original size as Fig. 6d.

Fig. 6 The example of preprocessing of different slices

4.2.5 Step5: Convex hull

Since some nodules are distributed around the edges of the lungs, it is necessary to make the mask contain nodules at the edges of the lungs, so we performed a convex wrap operation on the mask and then expanded outward by 5 pixels as Fig. 6e. Then as illustrated in Fig. 6f, multiply the image with the mask and apply the new mask to the original data. And fill the parts outside the mask with 170, which is the brightness of normal tissue.

4.2.6 Step6: Coordinate transformation

Read the label, since the coordinates in the given label are world coordinates, it needs to be converted to voxel coordinates, and then the new resolution is applied to it. Since the data are inside the box, the coordinates are relative to the box. The background area unrelated to the lung is then deleted, leaving the lung parenchyma, and the preprocessed data are stored in npy format.

4.3 Nodule detection

We trained and evaluated the detection network on the LUNA16 data set. tenfold cross-validation was performed on 1,186 nodules in 888 cases. In the nodule detection part, we proposed a 3D Selective Kernel residual network, and we use a 3D residual network with faster R-CNN[13] and 3D dual path network[39] as the comparison. In order to verify the influence of ResNet[23], DPN[6] and SK-ResNet on the recognition performance, the 3D SK-ResNet block is replaced with the 3D Res18 and 3D DPN network.

The U-net shape structure network of Fig. 5 was used, while the SK-ResNet block is replaced with the Res18 and DPN network for training, the results obtained were compared with the model in this paper.

Fig. 7 The val_loss curve while training the detection network with 3D SK-ResNet, 3D Res18 and 3D DPN modules

Compared with the U-Net shaped regional proposal network with DPN or ResNet, the SK-ResNet used in this paper reduces loss more rapidly during training and has better convergence speed as Fig. 7. Draw the loss of the verification set of the network using ResNet, DPN and SK-ResNet, respectively. It is found that compared with the other two types of loss, loss of SK-ResNet has the fastest convergence speed. Compared with the other two similar networks, the method proposed in this paper shows that loss decreases and converges faster.

The evaluation indicator free-response area under the Receiver Operating Characteristic Curve (FROC) represents the average recall rate at each scan at 0.125, 0.25, 0.5, 1, 2, 4 and 8 at the average number of false positives, respectively. Then perform a non-maximum suppression (NMS) [9] operation to exclude the overlapping proposals and set the IoU threshold to 0.1.

Fig. 8 The FROC curve of the 3D SK-ResNet, 3D Res18 and 3D DPN network for detection under IoU threshold of 0.5, 0.7 and 0.9 (0.5 is the threshold we use in this paper)

We implemented 3D Res18 with faster R-CNN[22] and 3D DPN network [6] as our baseline, while the 3D Res18 with faster R-CNN is the state-of-the-art method. FROC performance of LUNA16 is shown in Fig. 8. The solid line is interpolated FROC based on true prediction result. The FROC of 3D SK-ResNet, 3D Res18 and 3D DPN network. It represents the sensitivity rate with respect to false positives per scan, the average recall rate are 0.125, 0.25, 0.5, 1, 2, 4, 8.

Figure 8 shows the FROC curves of the proposed method and the two comparison networks under different IoU thresholds. As illustrated in Fig. 8a, when the standard of the positive sample is $\text{IoU} > 0.5$, the 3D region proposal network with a 3D Res18 block and a U-net shape structure got a FROC score of 85.36%, and the network uses 3D DPN block got a FROC score of 84.20%. The 3D region proposal network with 3D deep SK-ResNet and a U-net shape structure

achieves 89.48%, which is a better detection result than baseline. As shown in Fig. 8b,c, when the IoU threshold is 0.7 and 0.9, our network reaches 83.87% and 79.63% in FROC score, which is better than the other two baseline networks. With the improvement of IoU, the performance of the model decreases, but the proposed method is at the leading level in different IoU settings. Compared with 3D Res18 and 3D DPN, 3D SK-ResNet can adaptively adjust the size of the receptive field by introducing multiple convolution kernels and extract features more accurately at different recalls.

During the experiment, Sen(Sensitivity), FP(average number of False Positive), FROC score and CPM(Challenge Performance Metric) were used to evaluate the performance of the pulmonary nodules detective algorithm.

The sensitivity formula is as follows:

$$\text{Sen} = \frac{TP}{TP + FN}, \quad (10)$$

Table 1 Nodule detection comparisons on LUNA16 dataset

Model	Sen	FP	FROC score	CPM
3D Res18 [13]	95.53%	21.56	82.32%	85.36%
3D DPN [6]	94.32%	30.82	84.14%	84.20%
3D SK-ResNet	95.78%	25.51	89.48%	89.13%

Bold text indicates the best results obtained in the comparative experiments

where Sen is sensitivity, TP is the number of all nodules detected correctly, FN is the number of non-pulmonary nodules divided into pulmonary nodules, $TP + FN$ represents all detected number of nodules. The FROC score represents a comprehensive indicator of the detection recall rate and the number of tolerable false positives. The higher the FROC score, the better the performance of the model. The number of FPs refers to the average number of false positive nodules in the test results. The smaller the value, the less the average number of false positives generated by the model, which means the model performs better.

In addition, take the average recall rate of point 0.125, 0.25, 0.5, 1, 2, 4 and 8 on the horizontal axis of FROC curve as the evaluation index CPM, and the CPM index formula is as follows:

$$CPM = \sum_{i=1}^n Rc_i/n, \quad (11)$$

where $n = 7$, $Rc_i \in \{0.125, 0.25, 0.5, 1, 2, 4, 8\}$, and Rc_i refers to the average number of false positive pulmonary nodules in CT images recall rate.

The pairs of results on nodule detection comparisons using LUNA16 dataset are illustrated in Table 1. We compared the differences between the other two networks and the network proposed in this paper on LUNA16 data set. It can be concluded from Table 1 that in some test sets, when the basic residual structure is replaced with SK-ResNet, the sensitivity is similar to that of the basic residual structure, which is nearly 1.5% higher than DPN. Compared with the other two networks, AP has increased significantly, the CPM indicator increased by nearly 5 percentage points.

SK-ResNet introduces a selection kernel mechanism in the network to focus on extracting the spatial location information of lung nodules in CT images. Our network obtained 95.78% average FROC score on LUNA16 datasets, which is higher than the score of DPN[6].

At the same time, Table 2 summarized the performance of those methods on nodule detection comparisons on the collected test dataset of CT images of patients, the proposed 3D SK-ResNet achieved the highest average FROC score of 85.89%, which is 0.81% and 2.17% higher than Res18 [13]

Table 2 Nodule detection comparisons on collected test dataset

Model	Sen	FP	FROC score	CPM
3D Res18 [13]	91.70%	32.67	85.17%	82.21%
3D DPN [6]	90.53%	45.33	83.72%	87.10%
3D SK-ResNet	91.39%	31.96	85.89%	87.42%

Bold text indicates the best results obtained in the comparative experiments

and DPN [6], this demonstrates the superior suitability of the proposed 3D SK-ResNet for detection.

In terms of calculation amount and complexity, when the depth separable convolution operation is used alone, the size of the generated model is about the same as the basic residual structure.

The results show that: compared with the basic residual structure, the use of the depth separable convolution operation can ensure the accuracy of lung nodule detection, while greatly reducing the computational complexity and complexity of the model, and demonstrated that our method can greatly outperform the state-of-the-art nodule detection methods.

4.4 Nodule classification

We validate the nodule classification performance of the proposed LungSeek system on the LUNA16 dataset with tenfold cross-validation. In the selection of the training scheme, we rearranged the training set and the verification set, using a quarter of the original training set as the new verification set, and combining the rest with the original verification set to form the new training set.

As Fig. 9 shows, when the test result was the same as the ground truth, it is called TP (true positive), and when the ground truth do not exist detected nodules, it is FP (false positive) as the right two figures.

There are 1004 nodules of which 450 are positive. The total epoch number is 800. The initial learning rate is 0.01, which decreases to 0.001 after the 400 epochs, and finally decreases to 0.0001 after the 600 epochs. Due to the limitation of training time and resources, we use the fold 1, 2, 3, 4 and 5 for testing. The final performance is the average performance of five test folds.

The proposed network uses multiple convolution kernels of different sizes to convolve the input CT images of lung nodules in a single convolutional layer. And extract the feature information of lung nodules in different ranges at different scales, then aggregate these characteristic information. These features are spliced in the channel dimension, and the fusion features are input into the SK-ResNet module to extract higher-level semantic information. After this operation, the features of the receptive field obtain performance

Fig. 9 Examples of True positive and False Positive

gains. Finally, the features undergo global average pooling and fully connected operations, after the softmax function outputs benign and malignant classification results.

We first filled the nodules in the size of $32 \times 32 \times 32$ into $36 \times 36 \times 36$, and the filled parts were enhanced by horizontal, vertical and Z-axis roll-over data. Then, $32 \times 32 \times 32$ was randomly clipped from the filled data and normalize the data using the average value and standard deviation obtained from the training data.

To evaluate the performance of our network, we compared the state-of-the-art methods and the proposed method on LUNA16 datasets. The nodule classification performance is summarized in Table 3, our method achieves better performance than those of Vanilla 3D CNN [37], Multi-crop CNN [33] and 3D DPN [39], because of the advantages of 3D structure and Selective Kernel networks. The proposed model has stronger expressive power in accuracy and is better than other classification networks in terms of model multi-convolution design and feature fusion strategy. Multi-scale feature fusion network with SK-ResNet achieves better performance than DenseNet [7], SE-ResNeXt [21] and Hierarchical semantic CNN [32] and obtain 92.75% accuracy, it exceeds DenseNet by 1.01%. which shows the effectiveness of the proposed network.

In addition, we compared the predictions of our method with three experienced doctors on the LUNA16 dataset and concluded that our method has higher accuracy than those doctors' average accuracy. As demonstrated in Table 4, by comparing the doctor's diagnosis conclusion with the method proposed in this article, we found that our accuracy rate is

Table 3 Nodule diagnosis accuracy between other methods and the classification network in our method on Luna16 dataset

Diagnostician	Year	Accuracy
Vanilla 3D CNN [37]	2016	87.40%
Multi-crop CNN [33]	2017	87.14%
3D DPN [39]	2017	88.76%
DenseNet [7]	2018	91.54%
SE-ResNeXt [21]	2018	91.31%
Hierarchical semantic CNN [32]	2019	90.82%
IDSNet [18]	2020	91.10%
Our method	2021	92.75%

Bold text indicates the best results obtained in the comparative experiments

Table 4 Nodule diagnosis accuracy between experienced doctors nodule and the classification network in our method on LUNA16 data set

Diagnostician	Accuracy
Dr.1	90.38%
Dr.2	91.27%
Dr.3	89.82%
Average	90.64%
Our method	91.75%

Bold text indicates the best results obtained in the comparative experiments

also higher than the accuracy rate of doctor's diagnosis data, which means that our method can reach the level of experienced doctors and can be very useful for assisting doctors in making accurate and efficient diagnoses.

By comparing the labeled information of doctors with the classified results of our network, it can be seen that the diagnosis results of the network are higher than the average diagnosis level of doctors. Our LungSeek system can be used to provide medical assistance to doctors.

5 Discussion

This paper presents a LungSeek system based on deep learning for automatic lung cancer CT diagnosis. The Selective Kernel network is designed for the detection of nodules, while a multi-scale feature fusion Network is used to assess the detected nodules and predict the cancer probability, where both of them achieve excellent results using the SK-ResNet module. We used a 3D Region proposal Network with 3D deep SK-ResNet to detect the candidate nodules and obtain coordinates, diameter and confidence of the nodules. Next, the detected deep features are sent to the nodule classification network, which uses the multi-scale feature fusion network to extract the multi-scale feature and classify the results as benign and malignant. Extensive experimental results on publicly available LUNA16 data sets certificate the superior performance of the proposed method.

Declarations

Conflict of interest We declare that we have no conflict of interest.

Availability of data and material (data transparency): The datasets used or analyzed during the current study are available from the corresponding author on reasonable request.

References

- Adiyoso Setio, A. A., Traverso, A. et al.: Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge. <https://arxiv.org/pdf/1612.08012.pdf> [cs.CV] 15 July 2017
- Alghamdi, A., Hammad, M., Ugail, H., et al.: Detection of myocardial infarction based on novel deep transfer learning methods for urban healthcare in smart cities [J]. *Multimedia Tools Appl* **2020**:1–22. <https://doi.org/10.1007/s11042-020-08769-x>
- Armato, S.G., et al.: The lung image database consortium (lidc) and image database resource initiative (idri): a completed reference database of lung nodules on ct scans. *Med. Phys.* **38**(2), 915–931 (2011). <https://doi.org/10.1118/1.3528204>
- Betz. Architecture and CAD for Deep-Submicron FPGAs[M]. Springer US, 1999. <https://doi.org/10.1007/978-1-4615-5145-4>
- Burt, P.J., Adelson, E.H.: The Laplacian pyramid as a compact image code. *IEEE Trans. Commun.* **31**(4), 532–540 (1983). <https://doi.org/10.1109/TCOM.1983.1095851>
- Chen, Y.P., Li, J.N., Xiao, H.X., Jin, X.J., Yan, S.C., Feng, J.S.: Dual path networks. In: NIPS (2017)
- Dey, R., Lu, Z., Hong, Y.: Diagnostic classification of lung nodules using 3D neural networks. In: The 15th International Symposium on Biomedical Imaging. Washington D C: IEEE, 2018, pp. 774–778. <https://doi.org/10.1109/ISBI.2018.8363687>
- Ferlay, J., et al.: Cancer incidence and mortality worldwide: sources, methods and major patterns in globocan 2012. *Int. J. Cancer* **136**(5), E359–E386 (2015). <https://doi.org/10.1002/ijc.29210>
- Girshick, R., Donahue, J., Darrell, T., et al.: Rich feature hierarchies for accurate object detection and semantic segmentation[J]. 2014. <https://doi.org/10.1109/CVPR.2014.81>
- Hammad M A, Iliyasu A M, Subasi A, et al. A Multi-tier Deep Learning Model for Arrhythmia Detection[J]. *IEEE Transactions on Instrumentation and Measurement*, 2020, PP(99).
- Hammad M A, Alkinani M H, Gupta B B, et al. Myocardial Infarction Detection Based on Deep Neural Network on Imbalanced Data[J]. *Multimedia Systems*, 2021(2).
- He, K., Gkioxari, G., Dollár, P. et al.: Mask R-CNN[J]. *IEEE transactions on pattern analysis & machine intelligence* (2017). <https://doi.org/10.1109/ICCV.2017.322>
- He, K., Zhang, X., Ren, S., et al.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE (2016). <https://doi.org/10.1109/CVPR.2016.90>
- Hu, J., Shen, L., Albanie, S., Sun, G., Wu, E.H.: Squeeze-and-excitation networks. In: IEEE CVPR 2018, pp 7132–7141. (2018) <https://doi.org/10.1109/TPAMI.2019.2913372>
- Huang, G., Liu, Z., Laurens, V., et al.: Densely connected convolutional networks[C]// IEEE Computer Society. IEEE Computer Society (2016). <https://doi.org/10.1109/CVPR.2017.243>
- Jiang, X., Jin, Y., Yao, Y.: Low-dose CT lung images denoising based on multiscale parallel convolution neural network. *Visual Comput.* (2020). <https://doi.org/10.1007/s00371-020-01996-1>
- Lei, B., Feng, D., Kim, J.: Dual-Path Adversarial Learning for Fully Convolutional Network (FCN)-Based Medical Image Segmentation[J]. *Visual Computer* **34**(6–8), 1–10 (2018)
- Li, X., Shen, X., Zhou, Y., et al.: Classification of breast cancer histopathological images using interleaved DenseNet with SENet (IDSNet). *PLoS ONE* **15**(5), e0232127 (2020)
- Li, Z., Zhang, S., Zhang, J., et al.: MVP-Net: multi-view FPN with position-aware attention for deep universal lesion detection (2019). https://doi.org/10.1007/978-3-030-32226-7_2
- Li, X., Wang, W., Hu, X. et al.: Selective Kernel networks (2019). <https://doi.org/10.1109/CVPR.2019.00060>
- Li, Y., Fan, Y.: DeepSEED: 3D squeeze-and-excitation encoder-decoder convolutional neural networks for pulmonary nodule detection. (2019). <https://doi.org/10.1109/ISBI45749.2020.9098317>
- Liao, F., Liang, M., Li, Z., et al.: Evaluate the Malignancy of pulmonary nodules using the 3D deep leaky noisy-or network. *IEEE Trans. Neural Netw. Learn. Syst.* (2017). <https://doi.org/10.1109/TNNLS.2019.2892409>
- Lin TY, Dollár P, Ross G, et al. Feature pyramid networks for object detection. In: IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE (2017), pp. 936–944. <https://doi.org/10.1109/CVPR.2017.106>
- Messay, T., Hardie, R.C., Tuinstra, T.R.: Segmentation of pulmonary nodules in computed tomography using a regression neural network approach and its application to the lung image database consortium and image database resource initiative dataset. *Med. Image Anal.* **22**(1), 48–62 (2015). <https://doi.org/10.1016/j.media.2015.02.002>
- Patroutpas, K.: Multi-scale window specification over streaming trajectories. *J. Spatial Inf. Sci.* **7**, 45–75 (2013). <https://doi.org/10.5311/JOSIS.2013.7.132>
- Qi, D., Hao, C., Yu, L., et al.: Multilevel contextual 3-D CNNs for false positive reduction in pulmonary nodule detection[J]. *IEEE*

- Trans. Biomed. Eng. **64**(7), 1558–1567 (2017). <https://doi.org/10.1109/TBME.2016.2613502>
27. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer International Publishing (2015). https://doi.org/10.1007/978-3-319-24574-4_28
 28. Sedik, A., Ilyasu, A., El-Rahiem, B., Samea, M., Abdel-Raheem, A., Hammad, M., Peng, J., Abd El-Samie, F., Abd, E.-L.: Deploying machine and deep learning models for efficient data-augmented detection of COVID-19 infections. *Viruses* **12**, 769 (2020). <https://doi.org/10.3390/v12070769>
 29. Setio, A.A.A., et al.: Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the luna16 challenge. *Med. Image Anal.* (2017). <https://doi.org/10.1016/j.media.2017.06.015>
 30. Setio, A., Ciompi, F., Litjens, G., et al.: Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks. *IEEE Trans. Med. Imaging* **35**(5), 1160–1169 (2016). <https://doi.org/10.1109/TMI.2016.2536809>
 31. Shaoqing, R., Kaiming, H., Girshick, R., et al.: Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(6), 1137–1149 (2017). <https://doi.org/10.1109/TPAMI.2016.2577031>
 32. Shen, S., Han, S.X., Aberle, D.R., Bui, A.A., Hsu, W.: An interpretable deep hierarchical semantic convolutional neural network for lung nodule malignancy classification. *Expert Syst. Appl.* **128**, 84–95 (2019). <https://doi.org/10.1016/j.eswa.2019.01.048>
 33. Shen, W., Zhou, M., Yang, F., Yu, D., Dong, D., Yang, C., Zang, Y., Tian, J.: Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification. *Pattern Recogn.* (2017). <https://doi.org/10.1016/j.patcog.2016.05.029>
 34. Siegel, R.L., Miller, K.D.: Jemal A (2018) Cancer statistics. *CA Cancer J Clin* **68**(1), 7–30 (2018). <https://doi.org/10.3322/caac.21442>
 35. Torre L A, Siegel R L, Jemal A. Lung Cancer Statistics.[J]. *Advances in Experimental Medicine & Biology*, 2016, 893:1–19. https://doi.org/10.1007/978-3-319-24223-1_1
 36. Wood DE, Kazerooni EA, Baum SL, et al. Lung Cancer Screening, Version 3.2018, NCCN Clinical Practice Guidelines in Oncology[J]. *Journal of the National Comprehensive Cancer Network Jccn*, 2018, 16(4):412.
 37. Yan, X., Pang, J., Qi, H., Zhu, Y., Bai, C., Geng, X., Liu, M., Terzopoulos, D., Ding, X.: Classification of lung nodule malignancy risk on computed tomography images using convolutional neural network: a comparison between 2d and 3d strategies. In *ACCV* (2016). https://doi.org/10.1007/978-3-319-54526-4_7
 38. Zhang, G., Jiang, S., Yang, Z., Gong, L., Ma, X., Zhou, Z., Bao, C., Liu, Q.: Automatic nodule detection for lung cancer in CT images: a review. *Comput Biol Med* **103**, 287–300 (2018). <https://doi.org/10.1016/j.combiomed.2018.10.033>
 39. Zhu, W., Liu, C., Fan, W., et al.: DeepLung: 3D deep convolutional nets for automated pulmonary nodule detection and classification (2017) <https://doi.org/10.1101/189928>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Haowan Zhang is currently a master student in the college of software engineering, Wuhan University of Science and Technology, China. She received her BS degree from Computer Science and Technology College of Wuhan University of Science and Technology, China, in 2019. Her research interests include deep learning and Medical Imaging.

Hong Zhang received the BS degree from Wuhan University of Technology, China, in 2001, the MS degree from Wuhan University of Technology, China, in 2004, and Ph.D. degree from Zhejiang University, China, in 2007. She is currently a professor in the college of computer science and technology, Wuhan University of Science and Technology, China. Her research interests include content-based multimedia analysis, machine learning and cross-media retrieval.