Signal, Image and Video Processing
https://doi.org/10.1007/s11760-022-02387-w

ORIGINAL PAPER

Improved V-Net lung nodule segmentation method based on selective kernel

Zerong Wang[1] · **Jingru Men**[1] · **Fuchun Zhang**[1]

Received: 19 June 2022 / Revised: 7 October 2022 / Accepted: 29 October 2022

Abstract

Segmentation algorithm based on deep learning has become the main method of pulmonary nodules segmentation; nevertheless, the accuracy and lightweight of most such models are difficult to coexist. In order to accurately segment lung nodules in computed tomography images and make the model lightweight, this paper proposes a lightweight segmentation network called SKV-Net, able to achieve good performance. The overall design of the network uses the original V-Net structure and introduces a selective convolution kernel with soft attention in selective kernel networks to extract multi-scale feature information. Adopting a suitable grouped convolution can effectively reduce the number of parameters in the model while maintaining good segmentation performance. Experimental results indicate that the average segmentation accuracy of SKV-Net is 1.3% higher than that of V-Net, and the number of parameters is only 42% those of V-Net. In this paper, the Luna16 public dataset of pulmonary nodules is used to test and evaluate the performance of various improved models. The results suggest that the SKV-Net is superior to other models, achieving good segmentation performance and fast operation speed. Moreover, the SKV-Net improves the segmentation of different types of pulmonary nodules. It has the advantages of high precision and lightweight structure, which further indicate that it has significant clinical application value in the segmentation task of pulmonary nodules.

Keywords Pulmonary nodule segmentation · Deep learning · Multiscale features · Soft attention · Group convolution

1 Introduction

Lung cancer has the highest cancer-associated mortality rate worldwide, making biomarker discovery a pressing issue [1]. Lung cancer is often in the middle or advanced stage when it is discovered, and the early symptoms are mostly small pulmonary nodules with a diameter usually less than 30 mm [2]. Currently, in Chinese diagnosis and treatment standards, pulmonary nodules with a diameter between 5 and 10 mm are defined as small nodules, while those with a diameter greater than 10 mm are defined as large nodules [3]. The morphology and grayscale characteristics of pulmonary nodules are similar to those of blood vessels and tissues, making it difficult for doctors to locate them correctly.

Currently, the available studies on segmentation methods of pulmonary nodules can be mainly divided into two categories: traditional methods based on manual feature extraction [4–8] and deep learning methods based on automatic feature extraction. In recent years, with the rapid development of deep learning, many deep neural network (DNN) models have been applied to the diagnosis of pulmonary nodules in medical imaging [9]. In 2015, Long et al. [10] proposed a fully convolutional network (FCN) method, which uses a fully convolutional end-to-end network that is more suitable for image segmentation. Based on the FCN network structure, Korez et al. [11] proposed a three-dimensional (3D) FCN network structure to further improve the segmentation accuracy of spinal magnetic resonance (MR) images. Nguyen et al. [12] compared the ability of different machine learning algorithms trained with resting-state functional MRI (rfMRI) latency data to detect epilepsy. To improve the segmentation accuracy, Ronneberger et al. [13] proposed U-Net, which added a jump connection on the basis of FCNs to fuse low-level detail information with high-semantic layer information. Xu et al. [14] segmented two-dimensional (2D) lung profiles containing pulmonary

---

✉ Fuchun Zhang
yadxzfc@yau.edu.cn

[1] School of Physics and Electronic Information, Yan'an University, Yan'an 716000, China

nodules using U-Net, and the pixel accuracy of the segmentation results reached 86.7%. Zhu et al. [15] improved the structure of U-Net using the separable convolutional module of Mobile-Net in the 2D image segmentation task of U-Net. As a result, the image segmentation accuracy of 2D lung profiles including pulmonary nodules was improved to 90%. Zhong et al. [16] used Dense-Net improved U-Net to segment pulmonary nodules from small slices obtained from 2D CT images of lungs. Cicek et al. [17] proposed a 3D U-Net network structure, which performed 3D image segmentation by inputting sequential 2D slices 3D images. Hui et al. [18] designed a multi-scale model of U-Net improved based on Dense-Net, and performed pulmonary nodule segmentation on 2D CT images. The segmentation accuracy of medium and small nodules was improved by 11% and 20%, respectively. Hong et al. [19] used a region proposal network of Faster RCNN to generate candidate regions and integrated attention mechanism gates using a 3D U-Net with residual connections to detect pulmonary nodules; the free-response ROC (Froc) increased by 1.6% compared to the baseline of Res-Net. Gao et al. [20] used a residual network as the Faster RCNN of the CNN as the main stem network, and used squeeze-and-excitation (SE) and second-order fusion to replace the original res block, achieving a detection accuracy of 89%. Wang et al. [21] used YOLO combined with Dense-Net to detect and predict benign and malignant nodules in 2D lung CT images, reaching a detection recall rate of 95.6%. Singadkar et al. [22] proposed a CT image segmentation method for pulmonary nodules based on a deep deconvolution residual network (DDRN). Their method captured the complete resolution characteristics through end-to-end training, and the Jaccard index was 88.68%.

Milletari et al. [23] proposed V-Net, a 3D image segmentation network, which added residual connection and introduced dice loss function on the basis of U-Net to segment large targets such as prostatic glands. Subsequently, He et al. [24] improved V-Net by replacing the large-size convolution with several small-size bottleneck structure blocks to segment 3D MRI images of the brain hippocampus, which reduced the number of parameters significantly. Zhong et al. [25] segmented pulmonary nodules using a multi-scale feature network composed of multiple sub-networks based on V-Net.

In general, the above studies have produced successful results. In the segmentation of pulmonary nodules under the condition of 3D contour information, V-Net has achieved a maximum dice of 0.791 and an average dice of 0.762 [26]. Nevertheless, taking a closer look at V-Net, three certain limitations can be observed. First, the original V-Net model has been applied to the segmentation of prostatic glands, while the diameter of pulmonary nodules is much smaller. Second, the morphology and gray characteristics of pulmonary nodules are similar to those of blood vessels and tissues. Sometimes pulmonary nodules adhere to blood vessels and pleura. Moreover, the generalization ability and robustness of the model are poor, and the number of parameters is huge. Third, the model cannot exploit multi-scale information. To this end, in this work, the characteristics of the 3D image data of lung nodules are analyzed, the multi-core attention mechanism of SK-Net [27] is integrated on the basis of V-Net, and a selective kernel V-Net (SKV-Net) is proposed. The network can automatically fuse the feature information extracted from multi-core, and the network training can effectively improve the generalization ability of the model and the accuracy of lung nodule segmentation.

2 Materials and methods

2.1 Segmentation framework for pulmonary nodules

The segmentation framework for pulmonary nodules used in this paper is illustrated in Fig. 1.

For the original CT images, the Luna16 dataset not only performs HU conversion on the original CT gray images (data threshold from − 1024 to 3071), but also provides the corresponding lung masks. The parenchyma was extracted by multiplying the original data with the pulmonary mask. Since the Luna16 dataset is collected from different sources, the pixel intervals of all images need to be unified. The X-, Y-, and Z-axes were unified at 1 mm to avoid experimental errors. To facilitate model calculation, data normalization and mean removal were performed on the pixel values. In the normalized range, the CT value of the lung window was from − 1000 to 400. De-mean means that all dimensions are subtracted from the mean value of the corresponding dimension, so that all dimensions of the input data are centered to 0, and all pixel values of this data set are subtracted 0.25. Due to the limitation of computing resources, a 3D data block of $64 \times 64 \times 32$ (length × width × depth) containing at least one pulmonary nodule was intercepted. Thus, $64 \times 64 \times 32$ blocks of data were input to the model, and binary splits of the same size were exported.

2.2 SKV-Net structure

V-Net is a 3D full convolutional neural network. The left and right ends of the V-Net are encoders and decoders, respectively, which correspond to the two operations of extracting image features and restoring the resolution of feature maps. It is worth mentioning that, in order to improve the utilization degree of features, a residual connection [28] was used to connect the input of the first convolutional layer in the module with the output of the last one, while the residual

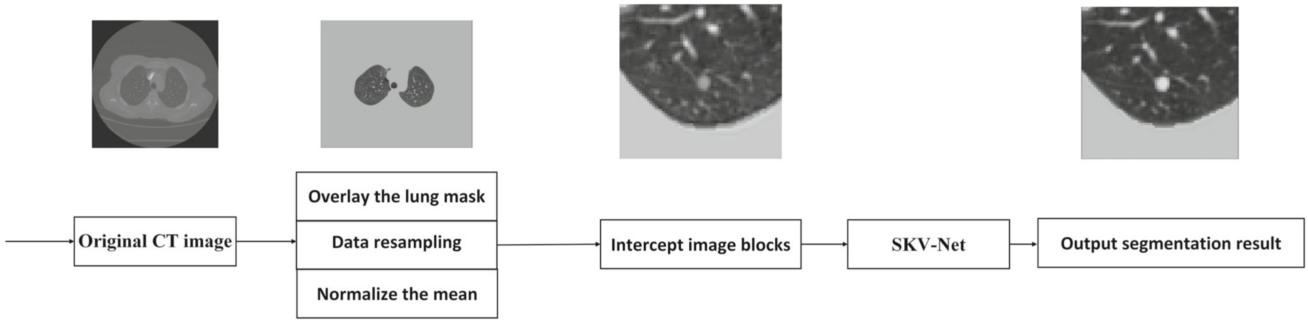

Fig. 1 Segmentation framework of pulmonary nodules

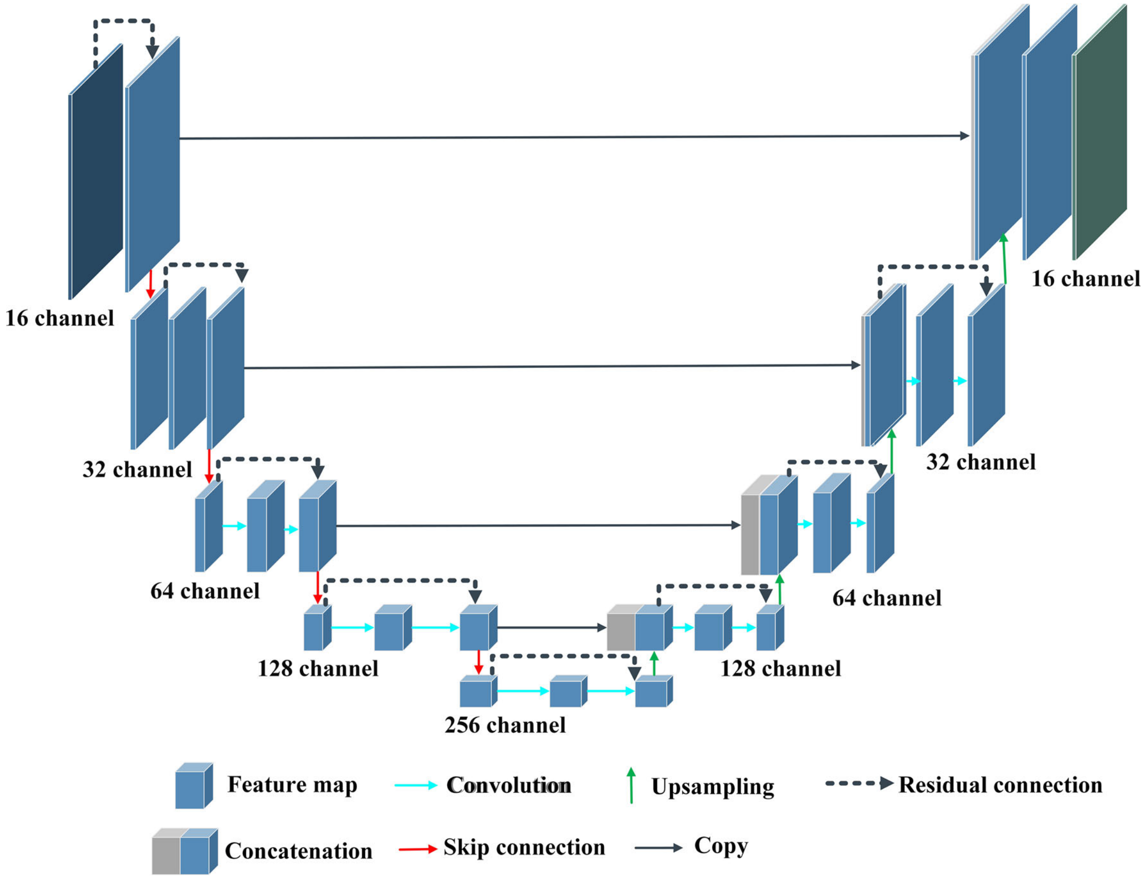

Fig. 2 V-Net network structure

connection used 1 × 1 × 1 convolution for channel alignment. This is the principal improvement of V-Net over U-Net. The structure diagram of V-Net is shown in Fig. 2.

V-Net is desirable in the overall design of feature extraction; however, since only 5 × 5 × 5 convolution is used in V-Net, the receptive field is large and single, and more detailed features are not taken into consideration. Nevertheless, pulmonary nodules are small targets and more attention needs to be paid to detailed features. For the judgment of the same point, multi-scale information can provide richer information, which is conducive to the classification of the judgment points. In view of this, the original single-size convolution in V-Net is replaced by a SK-Block that contains a multi-scale convolution kernel and soft attention. The SK-Block can extract multi-scale features from feature maps, paying attention to details and peripheral information. The soft attention mechanism enables the model to focus on more effective channels and increase the single point judgment accuracy.

In the SK-Block, the feature map passes through three routes (3 × 3 × 3 convolution, 5 × 5 × 5 convolution, and a route without any convolution operation; the 5 × 5 × 5 convolution is a stack of two 3 × 3 × 3 convolutions). Subsequently, the feature map generated after the three routes is added to obtain a multi-scale fusion feature map. Then, the feature map is globally pooled (one point of global pooling represents the average value of all points in a channel) to obtain a one-dimensional (1D) array with the same length as the number of channels. The 1D array is fully connected twice, and three groups of 1D arrays are generated during the second full connection corresponding to the multi-scale features on the three routes. Softmax operation is carried out on the three groups of 1D arrays, and adaptive selection of feature maps of different scales is carried out by using scale soft attention. Then, the new feature map is obtained by multiplying it with the feature map of the corresponding receptive field, and the three groups of feature maps are fused by concatenation. Finally, the number of channels is adjusted by 1 × 1 × 1 point convolution to obtain the final feature map.

As depicted in Fig. 3, the input X format of the SK-Block is $N \times C \times H \times W \times D$ (batch size × number of channels × length × width × depth). The number of input channels is denoted as C_{in} and the number of output channels is denoted as C_{out}. First, the multi-scale selective kernel SK-Block is performed without any operation to obtain the feature graph X_1. The feature graph X_2 is obtained by 3 × 3 × 3 convolution with a step size of 1. The feature graph X_3 is obtained by 5 × 5 × 5 convolution with a step size of 1. Afterwards, X_1, X_2, and X_3 are added together to obtain the fused feature graph U, as defined by Eq. (1). It should be noted that X_1, X_2, X_3, and U have C_{in} channels. The global average pooling F_{gp} of U is used to obtain a 1D array S, whose length is consistent with the number of channels, as described by Eq. (2). After S is processed by two fully connected layers fc_1 and fc_2, the number of channels remains unchanged, as shown in Eq. (3), where $W = \Re^{d \times C}$ is the weight of the full connection and δ is the leaky Relu activation function. For the part of the input nodule of leaky Relu activation function less than zero, the gradient can also be calculated, which can effectively avoid the vanishing gradients and the jagged gradient direction. Make full use of the information in the picture. Each full connection is followed by the activation function; however, the Batch norm is not connected, since the batch norm of the full connection layer will reduce the effect of the attention mechanism and affect model learning. Then, fc_2 generates three 1D arrays V_1, V_2, and V_3, whose parameters are shared in fc_1 and soft-max in channel dimension, as defined by Eq. (4), where V_i represents the ith 1D array. Soft-max can be regarded as the soft attention of the channel. This operation can sort the importance of the multi-scale features in the channel to obtain I_1, I_2, and I_3, and multiply them with the X_1, X_2, and X_3 feature maps as weights to generate the new feature maps Y_1, Y_2, and Y_3. Subsequently, the new feature map is concatenated, and the number of channels is $3 \times C_{\text{in}}$. Multi-scale feature fusion and channel number adjustment are conducted by point convolution. The $3 \times C_{\text{in}}$ number of channels is adjusted to output C_{out} number of channels and the output feature graph Y_{out} is obtained. The overall design of the SK-Block can take into account any number of input and output channels, and can be embedded in any network. Furthermore, its multi-scale features can effectively improve the model performance while reducing the number of parameters, which can be used as an alternative to ordinary convolution.

$$U = X1 + X2 + X3 \tag{1}$$

$$S = F_{\text{gp}}(U) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} U(i, j) \tag{2}$$

$$V_i = F_{\text{fc}}(S) = \delta(\text{WS}) \tag{3}$$

$$I = \frac{e^{(V_i)}}{\sum_{i=1}^{3} e^{(V_i)}} \tag{4}$$

$$G = \min(C_{\text{in}}, C_{\text{out}}, q) \tag{5}$$

In SK-Block, convolution can be set as group convolution. The setting of the group number G is given by Eq. (5), where q is set manually, and the group number G determines the number of the parameter. If the number of groups G is too large, the memory access cost (MAC) consumption is too large; if the number of groups G is too small, the number of parameters is too large and the efficiency is reduced. In addition, experimental groups on route selection were conducted as well. The first group included a 5 × 5 × 5 convolution route and a route without any convolution operation. In the second group, a 3 × 3 × 3 convolution route was added on the basis of the first group, and in the third group, a 3 × 3 × 3 expansion convolution was added on the basis of the second group. In the comparative experiment, the model data of different groups and different route combinations were displayed, and the model achieved the best effect when $G = 4$ and the route number $P = 3$ were weighed. According to the ablation experiment, compared to the original SK-Net block, adding a route without any convolution operation can extract the details of the feature map; thus, the model can produce better results.

2.3 Number of parameters

The number of parameters M_1 of a single convolution layer can be expressed by Eq. (6), where k denotes the convolution kernel size, N is the input channel, b is the offset term, and C is the output channel. In addition, $k = 1$ and C is the number of nodes used to calculate the number of full-connection layer-parameters. According to this equation, the number of

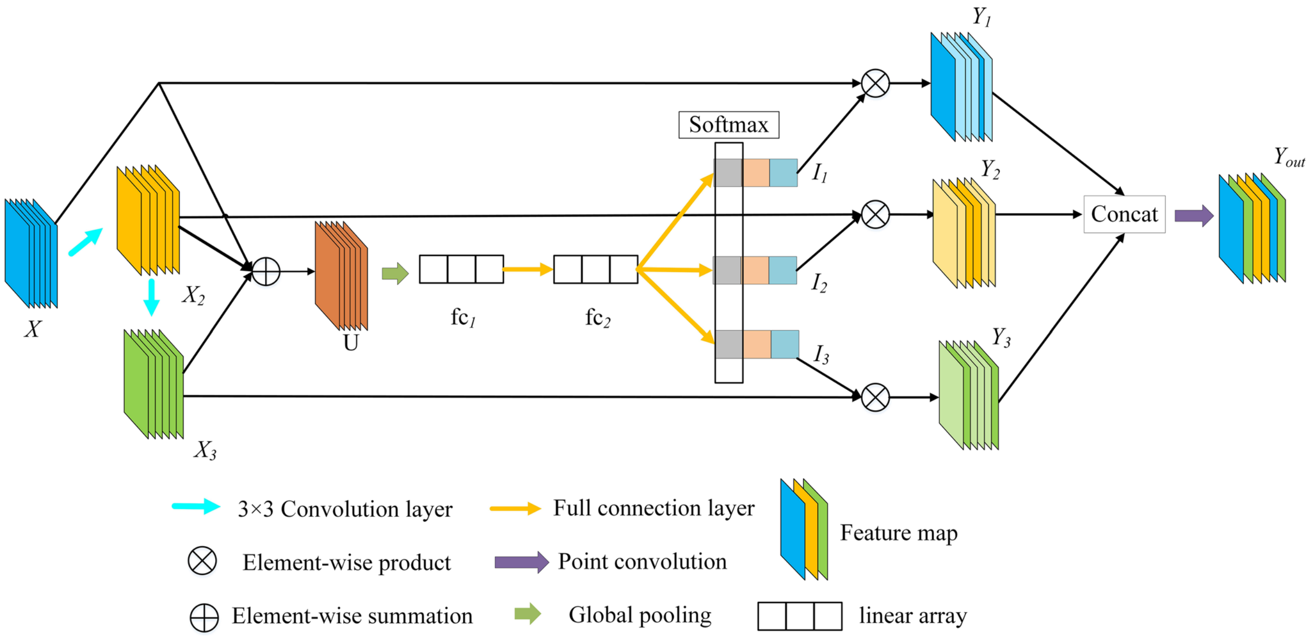

Fig. 3 Multi-scale selective nuclear structure SK-Block

parameters M_2 of the original V-Net $5 \times 5 \times 5$ convolution kernel can be calculated by Eq. (7). After the SK-Block is replaced, the convolution kernel becomes a $3 \times 3 \times 3$ common convolution and a $5 \times 5 \times 5$ common convolution, both of which use grouping. The number of parameters M_3 of the convolution with two fully connected layers and a point is determined as Eq. (8).

$$M_1 = (k \times N + b) \times C \tag{6}$$

$$M_2 = \left(5^3 \times N + 1\right) \times C \tag{7}$$

$$\begin{aligned} M_3 = {} & \left(3^3 \times N + 1\right) \times C \times \frac{1}{G} + \left(5^3 \times N + 1\right) \\ & \times C \times \frac{1}{G} + (1 \times C + 1) \times d \\ & + (1 \times d + 1) \times C + \left(1^3 \times N + 1\right) \times C \times \frac{1}{G} \end{aligned} \tag{8}$$

According to Eqs. (6)–(8), the total number of parameters in V-Net is 85.77 million, and the weight file size is 425.51 MB. On the other hand, the total number of SKV-Net parameters is 18.21 million, and the weight file size is 57.76 MB. The number of parameters between the V-Net and the method proposed in this paper was statistically compared, as well as the training time and test time when the batch size was 4; the results are given in Table 1.

As it can be seen in Table 1, the model proposed in this paper not only improved the segmentation effect of the model, but also significantly reduced the number of parameters to about 42% of the baseline. Therefore, the proposed model can be deployed on mobile devices. Moreover, the training time was 10% lower than the baseline, while the test time was higher than the baseline due to the grouping convolution.

Table 1 Comparison of parameters, training time and test time of the model

Model	Parameter/million	Training time/s	Test time/s
V-Net	19.61	7213	202
SKV-Net	8.24	6560	269

2.4 Loss function

In this paper, the dice loss was used as the loss function for network training. The dice coefficient of the two binary segmentation bodies can be expressed as Eq. (9). The calculation method of the dice loss can be expressed as Eq. (10). When Laplace smoothing = 1, zero division and overfitting can be avoided. Thus, after Laplace smoothing = 1 is added, the dice loss can be expressed as Eq. (11).

$$\text{DSC} = \frac{2|X \cup Y|}{|X| + |Y|} \tag{9}$$

$$\text{Dice loss} = 1 - \text{DSC} \tag{10}$$

$$\text{Dice loss} = 1 - \frac{2|X \cap Y| + 1}{|X| + |Y| + 1} \tag{11}$$

where X and Y represent predicted and true binaries, respectively, X and Y represent predicted value and true value, respectively. $X \cap Y$ denotes the sum of the values of the two points, $|X|$ said predicted sum, and $|Y|$ said real value combined. During the calculation process, dice loss calculation is performed for the feature map predicted as non-nodules and the real label predicted as non-nodules in the foreground. Subsequently, dice loss calculation is performed for the feature map predicted as nodules and the real label predicted as nodules in the foreground. The sum of the two is the actual dice loss.

3 Experimental analysis

3.1 Experimental environment and data

The experimental environment was as follows: The processor was an Intel (R) Core (TM) i7-8750 CPU @ 2.20 GHz; the memory capacity was 16 GB; the graphics card was an NVIDIA Tesla V100 (32G video memory); the operating system was Linux; the development language was Python 3.7, and the deep learning framework was PyTorch 1.6.0.

The data used in this study were derived from the lung nodule analysis 16 (Luna16) dataset, which is a chest imaging dataset. Luna16 was released in 2016 and contains 888 CT images of 1084 tumors with an ideal range of image quality and tumor sizes. It is one of the most commonly used datasets for lung tumor detection. Luna16 is derived from a larger data set, i.e., LIDC-IDRI [29], which contains a total of 1018 CT scan cases. Adding negative film will aggravate the problem of uneven positive and negative samples to a certain extent. Therefore, the data sets used in this paper are all images with nodules.

The lung nodule data provided by Luna16 are the coordinates of the lung nodule centers and their diameters; thus, spherical labels with the same size as the image data without morphological information were generated. Then, preprocessing operations were performed, such as data resampling, data normalization, and image cutting. Finally, 1186 lung CT data blocks with a size of 64 × 64 × 32 containing at least one pulmonary nodule and the corresponding pulmonary nodule labels were obtained (Fig. 4). Figure 4a depicts the pulmonary CT data block containing pulmonary nodules, and Fig. 4b shows the pulmonary nodule label corresponding to Fig. 4a. Subset0 to subset7 in Luna16 were used as the training set, subset8 as the verification set, and subset9 as the test set, at a ratio of 8:1:1.

3.2 Evaluation indicators

The dice similarity coefficient (DSC), pixel accuracy (PA), intersection-over-union (IOU), sensitivity, precision, F1 score, and Hausdorff distance (HD) were selected as the evaluation indices of pulmonary nodule segmentation. Among them, the calculation method of DSC is given by Eq. (9), that of PA is given by Eq. (12), that of IOU is given by Eq. (13), that of the sensitivity is given by Eq. (14), that of the precision is given by Eq. (15), that of the F1 score is given by Eq. (16), that of the HD is given by Eq. (17).

$$\mathrm{PA} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{12}$$

$$\mathrm{IOU} = \frac{|X \cap Y|}{|X \cup Y|} \tag{13}$$

$$\text{Sensitivity} = \frac{|X \cap Y|}{|X \cap Y| + |Y - X|} \tag{14}$$

$$\text{Precision} = \frac{|X \cap Y|}{|X \cap Y| + |X - Y|} \tag{15}$$

$$F1 = 2 \times \frac{\text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \tag{16}$$

$$\mathrm{HD}(X, Y) = \max_{x \in X} \min_{y \in Y} ||x - y|| \tag{17}$$

3.3 Parameter settings

In the training process, the epochs were set to 200 and the stochastic gradient descent (SGD) optimizer was selected. The initial learning rate was 0.001. After every 3 training sessions, the learning rate was decreased to 0.1 times the initial learning rate. The initial momentum was set at 0.99, which decreased to 0.9 in the 180th round, and the weight decayed to 1E−8. The validation set was validated every 5 epochs.

Figure 5 depicts the loss decline diagram of the model. As it can be seen in Fig. 5a, after 120 rounds of verification, the loss was basically flat, but increased slightly after 180 rounds, meaning that an over-fitting phenomenon took place. In this experiment, the early stop strategy was adopted, and the model with the minimum verification loss was selected as the optimal model. As it can be observed in Fig. 5b, the training loss entered a slow decline stage after 110 rounds and reached a stable level after 170 rounds. After testing, when batch size is set to 2, the model does not converge within 200 epochs and the training speed is very slow. When batch size is set to 8, it is fast to process the same amount of data, but too many epochs are required to achieve the same accuracy. When batch size is set to 4, the accuracy is high enough and the time is optimal.

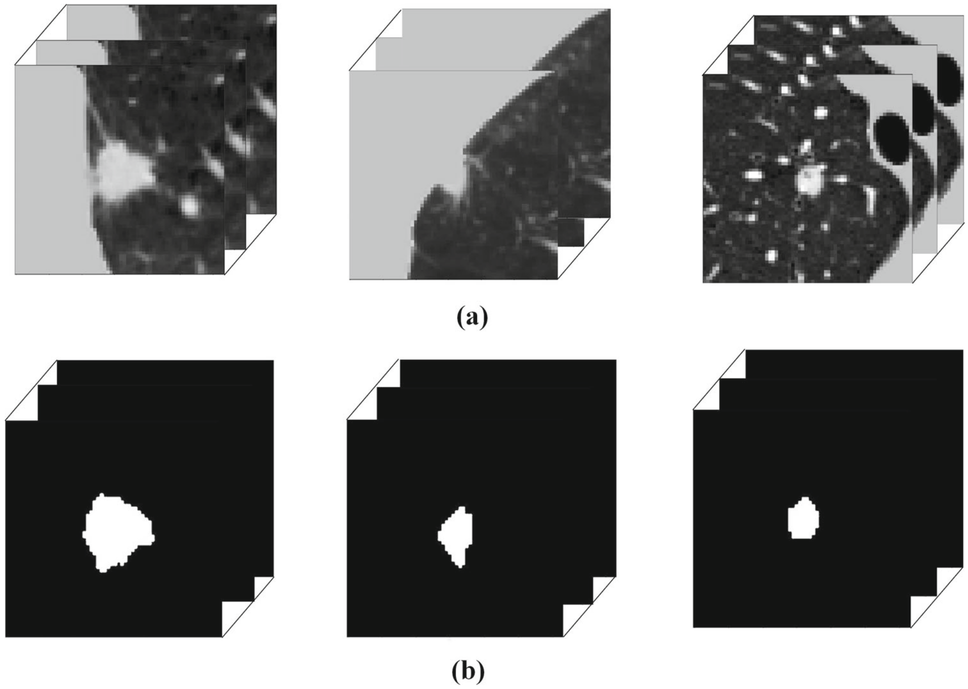

Fig. 4 Experimental data, **a**, Lung CT data block. **b**, Pulmonary nodule label

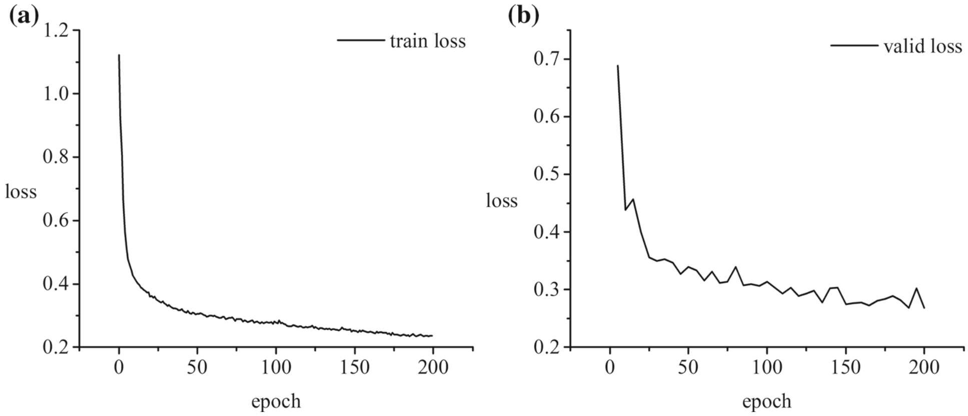

Fig. 5 Network training and verification curve, **a**, train loss. **b**, valid loss

Table 2 Comparison between the number of SKV-Net ablation participants and dice

Dice	$G = 2$	$G = 4$	$G = 16$	Parameter quantity	$G = 2$	$G = 4$	$G = 16$
Path = 2	0.784	0.786	0.778	Path = 2	57.25	29.32	8.37
Path = 3	0.795	0.796	0.781	Path = 3	73.4	39.85	14.21
Path = 4	0.772	0.749	0.739	Path = 4	127.91	68.84	21.03

3.4 Ablation experiments

In Sect. 2.2, it has been mentioned that the grouping number G and route number P of the grouping convolution affect the number of parameters and the accuracy of the SKV-Net model. The following study regarding their specific impact was conducted. The number of grouping convolution was set as $G = 2$, $G = 4$, and $G = 16$, and the grouping was calculated for the convolution on all routes. The number of routes was set to $P = 2$, $P = 3$, and $P = 4$. The routes were set as no-operation route $p1$, $3 \times 3 \times 3$ convolution route $p2$, $5 \times 5 \times 5$ convolution route $p3$, and $3 \times 3 \times 3$ dilatative convolution route $p4$, while the number of expansion steps was 1. When $P = 2$, routes $p1$ and $p2$ were taken; when $P = 3$, routes $P1$, $P2$, and $P3$ were taken; when $P = 4$, all routes were taken (Table 2).

Table 2 compares the number of parameters and dice values of all ablation structures. It can be observed that the

Table 3 Performance comparison of different models

Model	Dataset	PA	Dice	Jaccard	Sensitive	F1	HD/mm	3D
U-Net	LIDC	0.867	–	–	–	–	–	No
MobileUNet	LIDC	0.90	–	–	0.899	0.898	–	No
SquExUNet	LIDC	–	0.80	–	0.90	–	–	No
InceptionUNet	LIDC	–	–	0.766	0.90	0.861	–	No
V-Net	LIDC	0.994	0.791	0.665	0.754	0.798	3.2 ± 1.6	Yes
MSDS-Unet	TCIA	–	0.675	–	0.746	0.682	–	Yes
SK-Net	LIDC	0.985	0.787	0.661	0.751	0.794	3.5 ± 1.8	Yes
SKV-Net (This work)	LIDC	**0.996**	**0.796**	**0.665**	**0.789**	**0.800**	2.8 ± 1.3	Yes

Table 4 Performance comparison of different size

Dice	Small nodules	Medium nodules	Large nodules
V-Net	0.789	0.766	0.751
SKV-Net	0.795	0.769	0.778

increase in the number of routes introduces more scale features. Through experimental verification, it can be concluded that $p1$, $p2$, and $p3$ improve the segmentation effect. Nevertheless, when $p4$ is added, the segmentation effect worsens; thus, $p4$ is the scale that needs to be abandoned. $P1$ is a new route proposed on the basis of the SK-Net, which can effectively improve the segmentation accuracy of the model. In addition, a too low number of groups will result in too many parameters and a large model. Too many packets will cause too much MAC consumption, slow down the running speed of the program, and reduce the segmentation accuracy. The effect of $G = 2$ was similar to that of $G = 4$, but the effect became worse when $G = 16$. For comprehensive comparison, $G = 4$ and $P = 3$ were used in the final model configuration.

3.5 Experimental result

Based on the setting of the network parameters in Sect. 2.4, the training was conducted in the V-Net and SKV-Net models, and a comparison with other indicators related to this work was performed. The comparative analysis of the test results is exhibited in Table 3.

Compared with the traditional U-Net model, the MobileUNet index was improved and the number of parameters was reduced. This is because the MobileUNet uses residual connection and the depthwise separable convolution of MobileNetV3 [30]. In the MobileUNet block, the point convolution uses a lightweight channel attention mechanism to reduce the number of parameters and improve the channel efficiency. The precision and recall of MobileUNet are effectively improved. SquExUNet [31] adopts a simple channel attention mechanism to emphasize the detection ability of nodules, which can availably improve the training indicators while maintaining the model parameters basically unchanged. The dice value of SquExUNet achieves the second-best result. InceptionUNet [32] refers to the block structure of InceptionV4 [33], and its asymmetric convolution and high-dimensional decomposition methods can achieve a balance between the reduction in the number of parameters and the improvement of the indicators. InceptionUNet uses a bottleneck structure to improve the recognition ability of positive samples. In the 3D CT image segmentation task, the MSDS-Unet [34] only adds residual connections in the 3D U-Net in terms of model improvement, which is similar to the V-Net, and index improvement is mainly accomplished through depth supervision. Furthermore, due to the large input scale of the MSDS-UNet model, the overall index is lower than that of other models. The SK-Net contains two kernels of different sizes and adaptively processes the output through three operators: split, fuse and select.

Compared with SK-Net, SKV-Net can effectively utilize multi-scale information and improve the model accuracy under the condition of reducing parameter quantity. In all 3D models, SKV-Net outperforms other methods in six evaluation indexes such as pa, dice and Jaccard, which indicates that the proposed method based on SKV-Net can achieve lung nodule segmentation accurately, and the segmentation effect is comparable to the current advanced segmentation algorithms. Therefore, it can serve as a lightweight alternative to V-Net. In the SK-Block, the route setting determines the source of multi-scale information extraction. When the number of routes is 1, the model degenerates into V-Net. Compared with the two routes of the block in SK-Net, SKV-Net adds $p1$ route without any processing is very important, since it facilitates the precise extraction of model features to a large extent. Moreover, the contribution to the semantic accuracy originates mainly from convolution routes such as $p2$ and $p3$. The HD value decreased to 2.8 ± 1.3 mm.

According to the size of nodules, lung nodules with diameters less than 5 mm are classified as small nodules, those with diameters ranging from 5 to 10 mm are classified as medium nodules, and those with diameters greater than 10 mm are

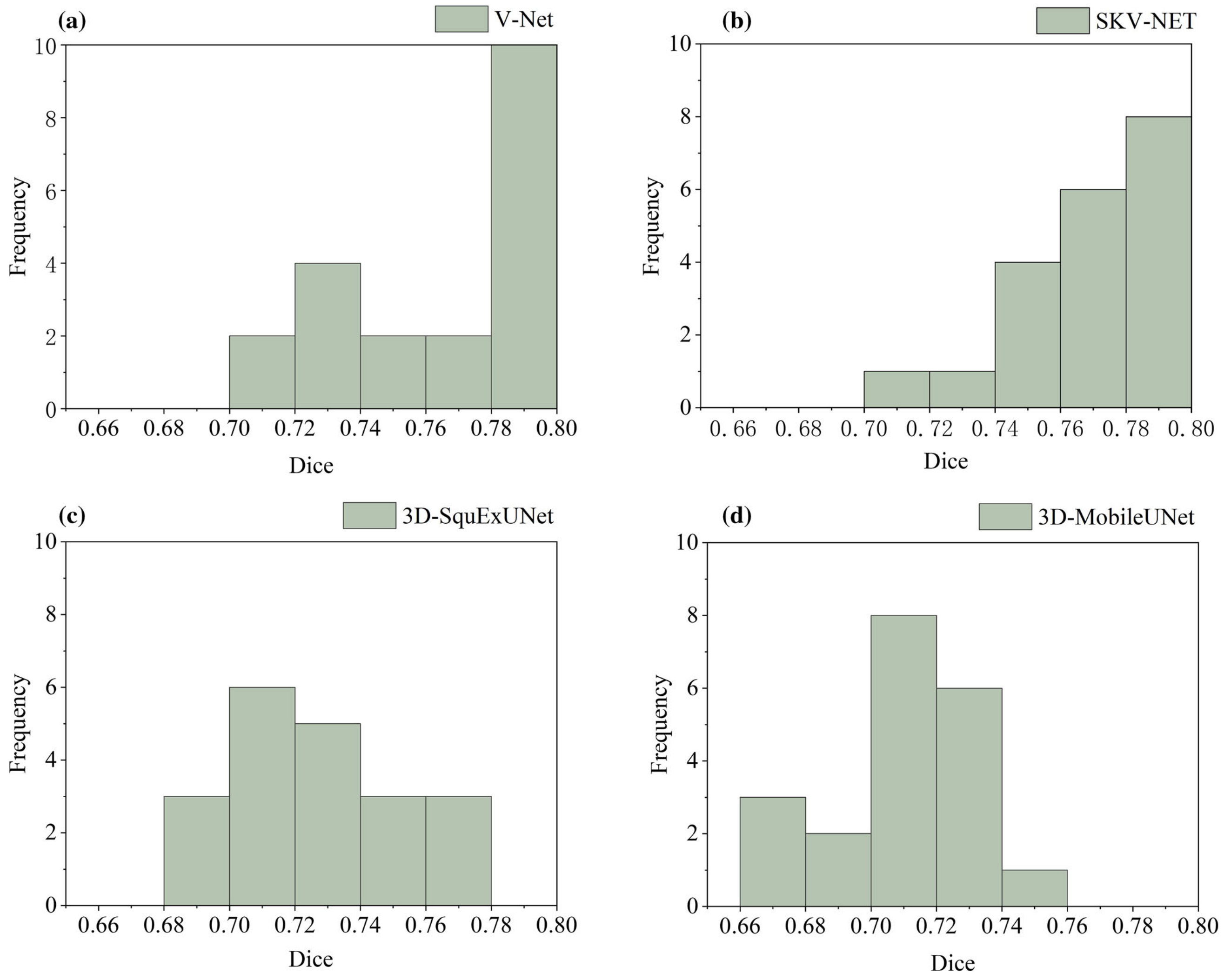

Fig. 6 Distribution statistics of dice similarity coefficients of different models, **a**, V-Net. **b**, SKV-Net. **c**, 3D-SquExUNet. **d**, 3D-SquExUNet

classified as large nodules. The segmentation effects of V-Net and SKV-Net on nodules with different sizes are shown in Table 4.

SKV-Net is superior to V-Net in segmentation of the three types of nodules, especially in segmentation of small nodules and large nodules. The experimental results verify the effectiveness and importance of multi-scale information for pulmonary nodule segmentation.

This paper conducted also statistical analysis on the DSC values obtained from the test sets of four networks, as shown in Fig. 6. As it can be observed, the dice proposed in this paper was mainly concentrated between 0.7 and 0.8. Figure 6a shows the training results of the V-Net model. The effect of the model was relatively stable, with an average dice coefficient of 0.762 and a maximum dice coefficient of 0.791. Figure 6b exhibits the training results of the SKV-Net model. When the number of parameters was significantly reduced, the average dice coefficient was 0.772 and the maximum dice coefficient was 0.796. The SKV-Net model achieved the best dice metrics. Figure 6c displays the training results of the 3D-SquExUNet model, whose dice index was slightly lower than that of the V-Net model. This was mainly due to that the channel attention mechanism of SE [35] was equivalent to the feature loss induced by the inactivation of unimportant channels. Figure 6d presents the training results of the 3D-MobileUNet model. Since the structure of the model was simple, the feature extraction performance was insufficient, reaching a maximum dice coefficient of only 0.749.

Based on the above experimental results, it has been proved that SKV-Net has better segmentation performance and higher robustness than the other models, and can produce good segmentation results for different types of pulmonary nodules.

Figure 7 depicts the cross sections of the 3D segmentation images of 5 pulmonary nodules by 4 networks. The

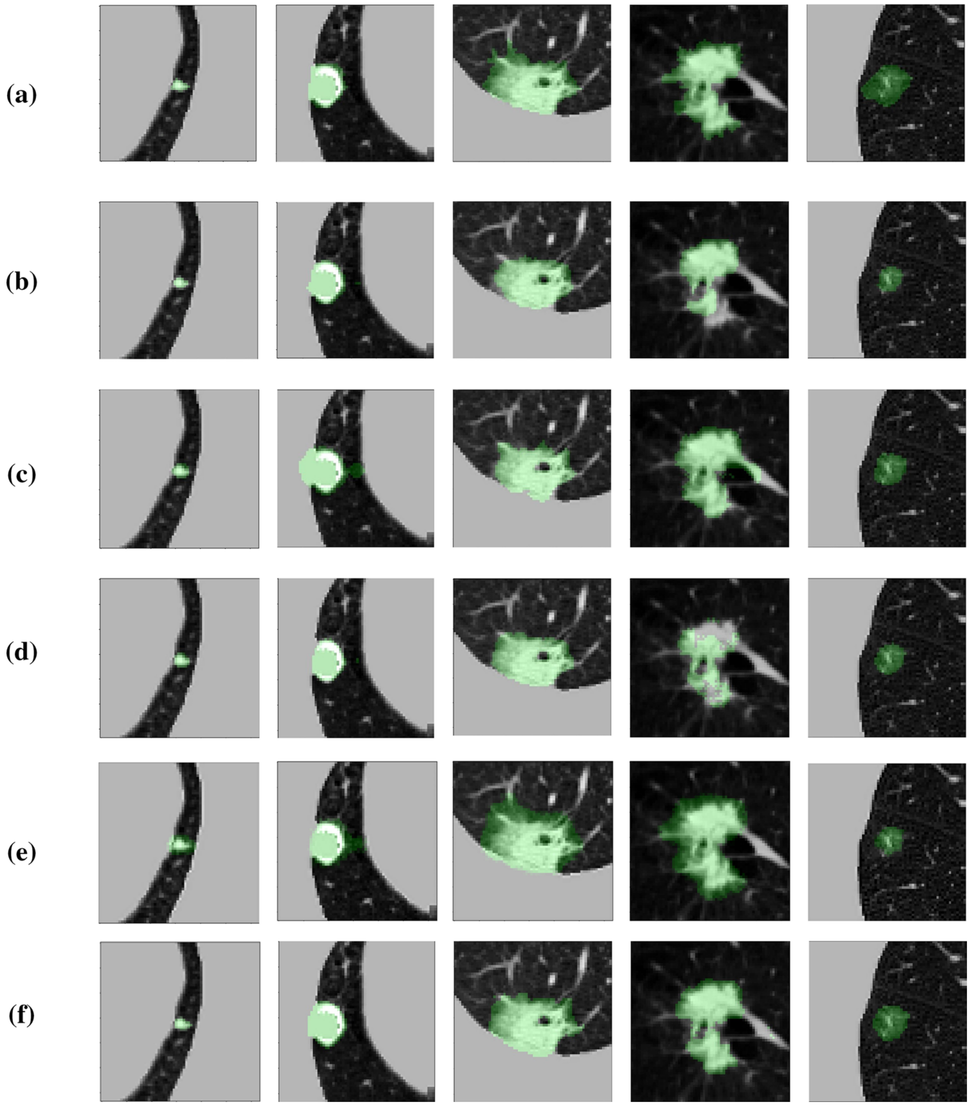

Fig. 7 Cross section of three-dimensional segmentation results of different networks, **a**, Label. **b**, V-Net. **c**, 3D-MobileUNet. **d**, 3D-SquExUNet. **e**, SK-Net. **f**, SKV-Net

non-nodular diameters from left to right correspond to small, medium, large, adhesion, and ground glass shadow nodules.

The five models of small nodules performed well, indicating that all had good sensitivity and effectiveness for the lesion region. The 3D-SquExUNet model for medium and large nodules began to exhibit over-segmentation and several false positive points. This can be mainly attributed to that the channel attention overexpressed some features, leading to the inability of the feature extractor to effectively converge within a limited amount of time. The V-Net and 3D-SquExUNet models are used to under-segment vascular adhesion nodules, the research results show unclear morphological contour segmentation and unsegmented or incomplete vascular adhesion contour segmentation, the 3D-MobileUNet model presented an over-segmentation problem in vascular adhesions. The above results are due to the lack of multi-scale information correction and the model pays too much attention to local information in point classification and judgment. But SK-Net method has the problem of over-segmentation in adhesion nodules and ground glass shadow nodules, which is caused by the lack of detailed information. Since the boundaries of opaque nodules of frosted glass are very fuzzy, the model can extract pathological features, but the contour information is not clear. Thus, the five models cannot achieve a sufficient segmentation effect. The SKV-Net can adequately solve this problem. Through the information fusion of different sensory domains, the under- and over-segmentation problem is well balanced. At the same time, the segmentation effect of ground glass nodules of the SKV-Net is superior to the overall over-segmentation of 3D-MobileUNet and other models, which indicate the superiority of SKV-Net regarding multi-scale feature extraction. Based on the above experimental results, it has been proved that the method proposed in this paper has better segmentation performance and robustness than other models, and can provide a better segmentation effect for different types of pulmonary nodules.

4 Discussion and conclusion

In this paper, a high-precision and lightweight pulmonary nodule segmentation model called SKV-Net has been proposed. The SKV-Net improves the maximum dice of the segmentation results under experimental conditions to 0.796 and the average dice to 0.772, which is 1.3% higher than the baseline. The segmentation results filled in the missing morphological information well, and the number of parameters was decreased to 42% of the original. Its structure inherits the overall design of V-Net, and the 3D convolution of V-Net is replaced by SK-Block. The latter has a multi-scale selective core and uses soft attention to fuse multi-scale information to improve the model. Based on the traditional selective kernel, this paper innovatively incorporates an unprocessed route to better extract detailed features. Compared with V-Net, the SKV-Net improves the segmentation effect, and reduces the number of parameters by using block convolution. Compared to other models, SKV-Net is more robust and can achieve a significant optimization effect for a variety of nodules, such as vascular adhesion and ground glass shadow nodules. Its small number of parameters enables its potential deployment on mobile devices. In future research, alternative data blocks will be selected from the entire lung CT images in combination with pulmonary nodular detection to achieve end-to-end pulmonary CT nodular segmentation.

Acknowledgements Not applicable.

Author contribution Z-RW performed conceptualization. Z-RW and J-RM were involved in software. Z-RW and F-CZ were involved in data curation, methodology, writing—original draft preparation and funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding The current study was supported by the Scientific and Technology Innovation Team of Yan'an University (2017CXTD-01), the Graduate Teaching Reform Project of Yan'n University (YDYJG2019016) and the Special Research Project of Yan'an University (YCX2022078).

Data availability The code described in this study is available from the first author upon reasonable request. Data supporting the results of this study will be provided by the first author upon reasonable request upon approval by the review Committee of the School of Physics and Electronic Information, Yan'an University.

Declarations

Conflict of interest The authors declare that they have no competing interests.

Consent for publication Not applicable.

Ethical approval Not applicable.

References

1. Sherafatian, M., Arjmand, F.: Decision tree-based classifiers for lung cancer diagnosis and subtyping using TCGA miRNA expression data. Oncol. Lett. **18**, 2125–2131 (2019)
2. Qinghua, Z., Yaguang, F., Ying, W., Youlin, Q., Guiqi, W., Yunchao, H., Xinyun, W., Ning, W., Guozhen, Z.: Chinese guidelines for classification, diagnosis and treatment of pulmonary nodules. Chin. J. Lung **19**, 793–798 (2016)
3. Xiaoju, Z., Li, B., Faguang, J., Qun-ying, H., Jie, H., Chun-xue, B., Liang-an, C., Weimin, L.: Lung nodule diagnosis and treatment in China. Chin. J. Tuberc. Respir. **41**, 763–771 (2018)
4. Armato, S.G., Giger, M.L., Moran, C.J., Blackburn, J.T., Doi, K., MacMahon, H.: Computerized detection of pulmonary nodules in CT scans. Radiology **209**, 161–161 (1998)
5. Ying, W., Xinhe, X., Tong, J., Dazhe, Z.: Extraction of suspected pulmonary nodules from CT images based on multi-scale morphological filtering. J. Northeastern Univ. (Nat. Sci.) **3**, 312–315 (2008)
6. Hui, L., Caiming, Z., Kai, D., Zhi-yuan, S.: J. Comput. Aided Des. Comput. Graph. **26**, 1727–1736 (2014)
7. Yehang, C., Zhi, L., Bao, F., Xiangmeng, C., Wansheng, L.: Segmentation of pleural contact pulmonary nodules based on improved active contour model. Chin. J. Sci. Instrum. **40**, 107–116 (2019)
8. Kostis, W.J., Reeves, A.P., Yankelevitz, D.F.: Threedimensional segmentation and growth-rate estimation of small pulmonary nodules in helical CT images. IEEE Trans. Med. Imaging **22**, 1259–1274 (2003)
9. Kommrusch, S., Pouchet, L.N.: Synthetic lung nodule 3D image generation using autoencoders. Int. Joint Conf. Neural Netw. **10**, 1–9 (2019)
10. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **39**, 640–651 (2015)
11. Korez, R., Likar, B., Pernus, F., Vrtovec, T.: Model-based segmentation of vertebral bodies from MR images with 3D CNNs. MICCAI 2016. In: Lecture Notes in Computer Science. vol. 9901 (2016).
12. Nguyen, R.D., Smyth, M.D., Zhu, L., Pao, L.P., Swisher, S.K., Kennady, E.H., Mitra, A., Patel, R.P., Lankford, J.E., Von Allmen, G., et al.: A comparison of machine learning classifiers for pediatric epilepsy using resting-state functional MRI latency data. Biomed. Rep. **15**, 77 (2021)
13. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. Med. Image Comput. Comput. Assist. Interv. (MICCAI) **9351**, 234–241 (2015)
14. Feng, X., Bin, Z., Jinxiang, G., Libo, L.: Segmentation of nodules based on u-net. Softw. Guide **17**, 161–164 (2018)
15. Wenjun, Z., Ming, C., Ying, Z., Mengyi, Z., Lihang, F.: A new method for segmentation of lung nodules based on Mobile-Unet network. J. Nanjing Univ. Technol. (Nat. Sci.) **09**, 1–8 (2021)
16. Sihua, Z., Xingming, G., Yineng, Z.: Improved segmentation method of pulmonary nodules using u-net network. Comput. Eng. Appl. **56**, 203–209 (2020)
17. Ozgun, C., Abdulkadir, A., Lienkamp, S.S.: 3D Net: learning dense volumetric segmentation from sparse annotation. Med. Image Comput. Comput. Assist. Interv. (MICCAI): 424–432, 2016.
18. Hui, Z., Pinle, Q.: U-net algorithm for pulmonary nodules detection based on multi-scale characteristic structure. Comput. Eng. Program **45**, 254–261 (2019)

19. Minjie, H., Gang, W., Xingchen, L., Juncheng, J., Xinhao, Y.: Detection of pulmonary nodules based on attention mechanism. Comput. Eng. Des. **42**, 83–88 (2021)
20. Gao, X., Xiaojun, B.: Detection of pulmonary nodules in low-dose CT images based on improved faster R-CNN. Appl. Res. Comput. **37**, 404–406 (2020)
21. Bo, W., Xupeng, F., Lijun, L., Qingsong, H.: Detection of pulmonary nodules in CT images based on improved YOLO algorithm. Beijing Biomed. Eng. **39**, 615–621 (2020)
22. Singadkar, G., Mahajan, A., Thakur, M., Talbar, S.: Deep deconvolutional residual network based automatic lung nodule segmentation. J. Digit. Imaging **33**, 4 (2020)
23. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: fully convolutional neural networks for volumetric medical image segmentation. In: Fourth International Conference on 3D Vision (3DV), IEEE, pp. 565–571 (2016).
24. Jintao, H., Dongwei, H., Boming, X., Pan, S., Miao, T.: MRI visualization system for human brain and hippocampus segmentation based on VNet neural network. Mechatronics **27**, 24–31 (2021)
25. Sihua, Z., Meng-lu, W., Xingming, G., Yao, Z., Yineng, Z.: Research on segmentation of pulmonary nodules based on improved VNet. Chin. J. Sci. Instrum. **41**, 206–215 (2020)
26. Yixin, C., Lei, C., Shu-guang, C., Jing-yang, G.J.: Beijing Univ. of Chem. Technol. (Nat. Sci.) 48: 86–95 (2017)
27. Li, X., Wang, W., Hu, X.: Selective Kernel networks. In: Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, pp. 510–519 (2020).
28. He, K., Zhang, X., Ren, S.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778 (2016).
29. Armato, S.G., Roberts, R.Y., Mcnitt-Gray, M.F.: The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans. Acad. Radiol. **14**, 1455–1463 (2007)
30. Howard, A., Sandler, M., Chen, B.: Searching for MobileNetV3. In: International Conference on Computer Vision (ICCV), pp. 1314–1324 (2019).
31. Dutande, P., Baid, U., Talbar, S.: LNCDS: A 2D–3D cascaded CNN approach for lung nodule classification, detection and segmentation. Biomed. Signal Process Control **67**, 6 (2021)
32. Gpa, B., Vrr, A., Prb, C.: CoLe-CNN:Context-learning convolutional neural network with adaptive loss function for lung nodule segmentation. Comput. Methods Programs Biomed. **1**, 198 (2021)
33. Szegedy, C., Ioffe, S., Vanhoucke, V.: Inception-v4, inception-ResNet and the impact of residual connections on learning. AAAI **31**, 4278–4284 (2017)
34. Jinzhu, Y., Bo, W., Lanting, L., Peng, C., Osmar, Z.: MSDS-UNet: a multi-scale deeply supervised 3D U-Net for automatic segmentation of lung tumor in CT. Comput. Med. Image Grap. **10**, 1–16 (2021)
35. Jie, H., Li, S., Gang, S., et al.: Squeeze-and-excitation networks. IEEE Trans. Pattern Anal. Mach. Intell. **42**, 2011–2023 (2020)