Electronic Research Archive

Research article

Multiscale lung nodule segmentation based on 3D coordinate attention and edge enhancement

Jinjiang Liu[1,2], Yuqin Li[2,3], Wentao Li[1,2], Zhenshuang Li[1,2] and Yihua Lan[1,2,*]

[1] School of Artificial Intelligence and Software Engineering, Nanyang Normal University, Nanyang 473061, China
[2] Henan Engineering Research Center of Intelligent Processing for Big Data of Digital Image, Nanyang 473061, China
[3] School of Life Science and Agricultural Engineering, Nanyang Normal University, Nanyang 473061, China

* Correspondence: Email: yihualan@nynu.edu.cn; Tel: +8618537796726.

Abstract: An important prerequisite for improving the reliability of lung cancer surveillance and clinical interventions is accurate lung nodule segmentation. Although deep learning is effective at performing medical image segmentation, lung CT image heterogeneity, nodule size, shape, and location variations, convolutional localized feature extraction characteristics, the receptive field limitations of continuous downsampling, lesion edge information losses, fuzzy boundary segmentation challenges, and the low segmentation accuracy achieved when segmenting lung CT images using deep learning remain. An edge-enhanced multiscale Sobel coordinate attention-atrous spatial convolutional pooling pyramid V-Net (SCA-VNet) algorithm for lung nodule segmentation was proposed to solve these problems. First, a residual edge enhancement module was designed, which was used to enhance the edges of the original data. Using an edge detection operator in combination with a residual module, this module could reduce data redundancy and alleviate the gray level similarity between the foreground and background. Then, a 3D atrous spatial convolutional pooling pyramid module set different expansion rates, which could obtain feature maps under different receptive fields and capture the multiscale information of the segmentation target. Finally, a three-dimensional coordinate attention network (3D CA-Net) module was added to the encoding and decoding paths to extract channel weights from multiple dimensions. This step propagated the spatial information in the coding layer to the subsequent layers, and it could reduce the loss of information during the forward propagation process. The proposed method achieved a Dice coefficient of 87.50% on the lung image database

consortium and image database resource initiative (LIDC-IDRI). It significantly outperformed the existing lung nodule segmentation models (UGS-Net, REMU-Net, and multitask models) and compared favorably with the Med3D, CENet, and PCAM_Net segmentation models in terms of their Dice coefficients, which were 3.37%, 2.2%, and 1.43%, respectively. The experimental results showed that the proposed SCA-VNet model attained improved lung nodule segmentation accuracy and laid a good foundation for improving the early detection rate of lung cancer.

1. Introduction

Lung cancer has one of the highest mortality rates worldwide. Early symptoms are mostly small lung nodules with diameters that are usually less than 30 mm [1]. Computed tomography (CT) imaging is the technique of choice for screening lung nodules due to its fast imaging speed and high image resolution [2]. Although CT imaging has the advantages of being intuitive and effective, CT imaging technology generates hundreds of lung images in one second through a single scan, and radiologists must individually examine CT images containing thousands of slices for diagnosis purposes; this process is time-consuming, labor-intensive, and prone to misdiagnoses. Therefore, proposing an efficient and accurate lung nodule segmentation method is highly practical for improving the early detection rate of lung cancer. Clinical features such as the sizes and shapes of pulmonary nodules (including calcification, lung lobe, and ground-glass shadow features) vary greatly at different levels. The edge contour shapes of pulmonary nodules vary, and their contour lines are not sufficiently obvious. The morphological features and gray features of pulmonary nodules observed on CT images are similar to those of vascular tissues, which makes it difficult to accurately segment pulmonary nodules. As a result, the existing pulmonary nodule segmentation methods are not sufficiently accurate for pulmonary nodule segmentation applications. In this paper, we leverage the efficient image feature extraction capabilities of convolutional neural networks in deep learning models. We integrate the obtained features with edge detection methods derived from traditional image processing methods to segment lung nodules. Additionally, we incorporate a 3D CA-Net attention mechanism to enhance the focus of the network on the weight information across various spatial and channel dimensions, thus improving the feature representation capabilities of the neural network. Furthermore, we employ an enhanced null-space convolutional pooling pyramid module to construct convolution kernels with different receptive fields based on distinct null rates. This allows us to extract image features at multiple scales. Inaccurately segmented lung nodule edges are improved using the above method, thus providing enhanced lung nodule segmentation accuracy.

The current lung nodule image segmentation methods can be divided into two major categories: traditional manual feature extraction methods, such as threshold segmentation, region growth, edge detection, and morphological manipulation, and deep learning methods based on automatic feature extraction.

Because traditional methods are strongly affected by factors such as image background noise and textures, their segmentation effects are not ideal. Therefore, deep learning approaches based on

automatic feature extraction are widely used to conduct pulmonary nodule segmentation on medical images. This type of method trains a deep neural network model through a large quantity of image data to learn the shallow and deep features of pulmonary nodules to segment them. Ronneberger et al. [3] proposed the UNet segmentation model, which is composed of an encoding path that captures context and a decoding path. The model realizes precise positioning and greatly improves the accuracy of medical image segmentation. Fausto et al. [4] proposed the VNet model, which uses a residual module, an encoding path, and a decoding path to extract features from 3D medical images. They also proposed the Dice loss function. Li et al. [5] proposed the REMU-Net model. They introduced a spatial attention module and a feature enhancement module based on atrous spatial pyramid pooling (ASPP) so that the network could extract more diverse and efficient information and obtain richer contextual information. Finally, a multiscale hopping connection was used instead of the UNet hopping connection, overcoming the limitation that the decoder subnetwork could accept only feature information at the same scale. Their approach achieved a Dice coefficient of 0.8476 on the LIDC-IDRI dataset. Prasad et al. [6] proposed the SquExUNet model to combine a channel attention-based UNet model with a segmentation model for 2D lung nodule segmentation, and they achieved a segmentation Dice coefficient of 0.8. However, 2D segmentation requires the input data to be sliced while ignoring the depth information contained in the 3D dataset. Sundaresan et al. [7] proposed a multiscale fully convolutional three-dimensional UNet (MF-3D UNet) model and a maximum output aggregation method that fuses multiscale information in the encoding path to enhance the generalization ability of the model, enabling it to automatically segment lung nodules in CT images. The Dice coefficient achieved by this method on the LIDC-IDRI dataset was 0.83+ 0.05; however, this method cannot effectively segment nonsolid ground-glass lung nodules with fuzzy nodule boundaries, and the model is not sufficiently sensitive to nodule edge details, resulting in poor nodule segmentation effects. Wang et al. [8] proposed a lightweight segmentation network called SKV-Net. The overall network design adopts the original VNet structure, and a selective convolutional kernel with soft attention is incorporated into the selective kernel network to extract multiscale feature information. The method attained a Dice coefficient of 0.796 in a lung nodule segmentation task conducted based on the LIDC-IDRI dataset. This method is lightweight, but its segmentation effect needs to be improved. Zhou et al. [9] proposed a DSCMSF method for detecting and segmenting lung nodules. In the first three stages of the framework, the YOLOv5 model, a candidate nodule selection (CNS) algorithm, and a multisize 3D fusion model are used to locate nodules. In the final segmentation stage, a multiscale attention module is integrated into the 3D-based UNet autoencoder to finely segment the nodule regions. This technique realized a Dice similarity coefficient (DSC) of 0.8675 in a nodule segmentation task implemented on the LIDC-IDRI dataset. Zhou et al.[10] proposed an expansive convolutional neural network (MSDCNN) to better characterize the relationship between content and spatial information. Four parallel attentional expansion blocks were used to extract larger sensory field features while enhancing contextual information. Wang et al. [11] proposed the Attention-MVCNN model, which fuses channel attention, a residual structure and Mish activation to strengthen its 3D shape recognition performance. The excellent performance of this approach was experimentally proven on ModelNet40. Because the process of this network is divided into four stages, the number of required parameters is relatively large. The above methods use network structures that combine encoding and decoding paths, 2D or 3D network methods for segmenting lung nodules, residual modules, attention mechanisms, and multiscale information fusion to improve their lung nodule segmentation effects. These methods currently represent the mainstream direction of lung nodule segmentation research.

The original lung CT image is a 3D image. Due to resource limitations and other factors, many scholars process 3D CT images into 2D CT slice images for lung nodule segmentation and detection, which leads to considerable image information loss. To retain more spatial information on the nodules, 3D image data were used in this paper. To address problems such as loss of lesion edge information, fuzzy boundary segmentation, and low segmentation accuracy in deep learning-based lung nodule segmentation methods, this paper proposes an edge-enhanced multiscale lung nodule segmentation model called SCA-VNet, which can significantly improve the results of lung nodule segmentation by means of the edge-enhanced residuals module, the 3D CA-Net module, and the improved multiscale feature extraction null-space convolutional pooling pyramid. The major contributions of this study are as follows:

1) A residual edge enhancement module that combines the traditional edge detection method and a deep learning-based convolutional neural network is proposed. The module first enhances the input image data with edge data using an edge detection operator. Then, the edge-enhanced image is spliced with the original fusion result. Finally, it is fused with the feature image extracted by a convolutional neural network, which increases the sensitivity of the module to the edge contours of the lung nodule, providing more accurate edge information for model segmentation.

2) An optimized 3D coordinate attention (3D CA-Net) module is utilized to perceive nodule location information because the original 2D attention mechanism is not suitable for the 3D data used in this paper. The 3D CA-Net module obtained by improving the 2D attention mechanism module can realize multidimensional cross-channel interactions and extract multidimensional location information and channel weight information. Additionally, the adaptive pooling operation performed by this module is more flexible, and the input data are more robust so that the channel information that is useful for lung nodule segmentation can be considered from multiple dimensions. The generalization ability of the model is thus improved.

3) Atrous spatial pyramid pooling (ASPP) improves upon the 2D-suitable ASPP module to obtain the feature information of lung nodules at multiple scales, and different sensory fields are constructed by optimizing the sampling rate of ASPP according to the characteristics of lung nodule images. The convolution process captures the local and global features of the image, thus improving the performance and robustness of the model.

2. Related work

This section briefly describes the research related to optimizing edge detection and coordinate attention mechanisms.

2.1. Edge detection algorithms

Lung nodules are small objects on lung CT images that have diverse clinical features (e.g., calcification features, lobar features, and hairy glass shadow features); moreover, their edge contours are variable. Additionally, their contour lines are not sufficiently obvious, which leads to inaccurate nodule boundary segmentation results for the existing segmentation methods. Lisowska [12] proposed an efficient edge detection method for focused images that uses machine learning in combination with an edge detection algorithm to locally filter images in a region-by-region manner using a combination of the k-means algorithm and edge detection to solve the problem of inaccurate edge detection results for images with clear foregrounds and smooth backgrounds. Borba et al. [13] used the Gambini

algorithm to extract edge information features by fusing the edge feature information extracted through different channels that are more useful for this task, in turn enhancing the edge features of remote sensing data. Hait et al. [14] proposed an edge detection algorithm based on the Bezdek breakdown structure (BSS) by constructing a feature image and utilizing the intensity change direction and BM-type preaggregation operator. This method employing a combination of an edge detection operator and image intensity change information was validated by qualitative and quantitative evaluations to make it effective for image processing tasks. Wang et al. [15] proposed a boundary-based dual-path lung nodule segmentation model to solve this problem in 2D lung nodule segmentation tasks. This model uses edge detection operators in an edge detection branch and fuses the results with the outputs of the backbone lung nodule segmentation network, achieving good results but greatly increasing the complexity of the model. A lung nodule segmentation algorithm that fuses VNet and edge features was proposed by Jiang et al. [16]; their method uses an edge keypoint selection algorithm for edge feature extraction. However, this method suffers from incomplete edge connectivity and inaccurate edge localization during the detection process. The proposed residual edge enhancement module selects commonly used edge detection operators. It combines them with a residual module, enhancing the edge information of lung nodules without increasing the number of required model parameters.

The most commonly used and effective edge detection operators for image segmentation are the Sobel, Roberts, and Canny operators [17]. The best of these three edge detection operators is chosen through subsequent experiments as the edge detection operator in the proposed edge residual enhancement module for the lung nodule segmentation task.

2.2. Coordinate attention

In computer vision tasks, an attention mechanism is a technical means of enhancing the performance of convolutional neural networks, primarily by enabling models to focus on the relevant regions of the input task. These mechanisms selectively address the regions of interest rather than treating all locations as equally important. Squeeze-and-excitation (SE) attention calculates channel attention weights based on global information; however, this approach is less than ideal for some target tasks involving local feature information or small-scale information. Spatial attention focuses more on spatial information but ignores important channel and location information. The coordinate attention network (CA-Net) [18] mechanism dynamically weights the feature responses of different channels and spatial locations by learning the correlations between channels to enhance the attention given to important features. An et al. [19] proposed a novel deep framework that incorporates a new hybrid spatial channel attention module to facilitate cross-age face recognition tasks. Simultaneously executing average pooling and maximum pooling along the channel dimensions yields improved efficiency by optimizing the global maximum pooling operation in the channel attention mechanism to more efficiently capture the weights of each channel and compress the spatial dimensions of the feature map. Chen et al. [20] proposed an efficient two-gate attentional module to implement their dual-attention mechanism in the frequency domain. This approach yields improved deblurring performance and simultaneously reduces the computational cost of the model by avoiding redundant convolutions and feature channels. Li et al. [21] incorporated a global and local attention mechanism (GAL) into their model to provide an in-depth modeling approach for input images. This approach focuses on combining fine-grained local feature analysis with global contextual information processing, enabling the model to take both aspects into account to more accurately and comprehensively interpret images. Cao et al. [22] designed a ResSCBlock module based on the compression excitation

mechanism and a coordinate attention mechanism (ResNet). The fusion of the compression excitation module and the coordinate attention mechanism fully enables the model to make full use of the channel and location information of lung nodules to attain improved detection accuracy. In summary, attention mechanisms can improve the accuracy achieved in computer vision tasks. Considering the specifics of this study, we improve the coordinate attention mechanism. The original coordinate attention mechanism was proposed for 2D image data, and the dimensionality of CA-Net is improved in this paper so that it can be used in 3D image segmentation tasks. The core idea is to weigh the input feature map according to the observed spatial position and channel importance information. Then, the weighted feature map is input into the next layer of the network, and adaptive average pooling-based convolution is used on the three dimensions of the 3D data to more accurately locate the position and spatial information of the target nodule in the 3D image according to the channel information. This method enables the model to better focus on the lesion area.

3. The proposed method

3.1. Model structure design

The proposed SCA-VNet segmentation model is obtained by improving VNet, a three-dimensional fully convolutional neural network that adopts a U-shaped structure and is divided into an encoder, a decoder, and jump connections. The encoder and decoder correspond to image feature extraction and feature map resolution recovery processes, respectively. The most important feature of this network is that to achieve improved feature utilization, a residual module is used to realize jump connections for the output feature maps of different convolutional layers. The overall design of VNet is desirable for feature extraction; however, because only $5 \times 5 \times 5$ convolutional kernels are present in the VNet module, the sensing field is relatively singular, more detailed features are not considered, and splicing the same input data channels in the input module increases the number of channels, resulting in some information redundancy.

The improved SCA-VNet segmentation model uses VNet as its main model. A residual edge enhancement module is added at the beginning of the model to fully extract the edge contour information of the input image. The residual edge enhancement module also alleviates the information redundancy of the input module and better utilizes the input data. Adding the 3D CA-Net module to the encoding and decoding areas to model the relationships between channels allows the information interactions between channels to be captured more fully. This helps to mine the feature representations and better locate lesion information in the image.

A 3D atrous spatial convolutional pooling pyramid module is used between the encoding and decoding areas. This module can extract lesion information at different scales and thus achieve multiscale information extraction. The area between the encoder and decoder is a key location in the model, where adding ASPP can integrate a larger range of contextual information into the feature extraction process, thus yielding improved segmentation performance. To better adapt to the feature characteristics of lung nodules, this study optimizes the parameters of the convolution kernel in the model to $3 \times 3 \times 3$ for feature extraction. The overall structural design of the model is shown in Figure 1.
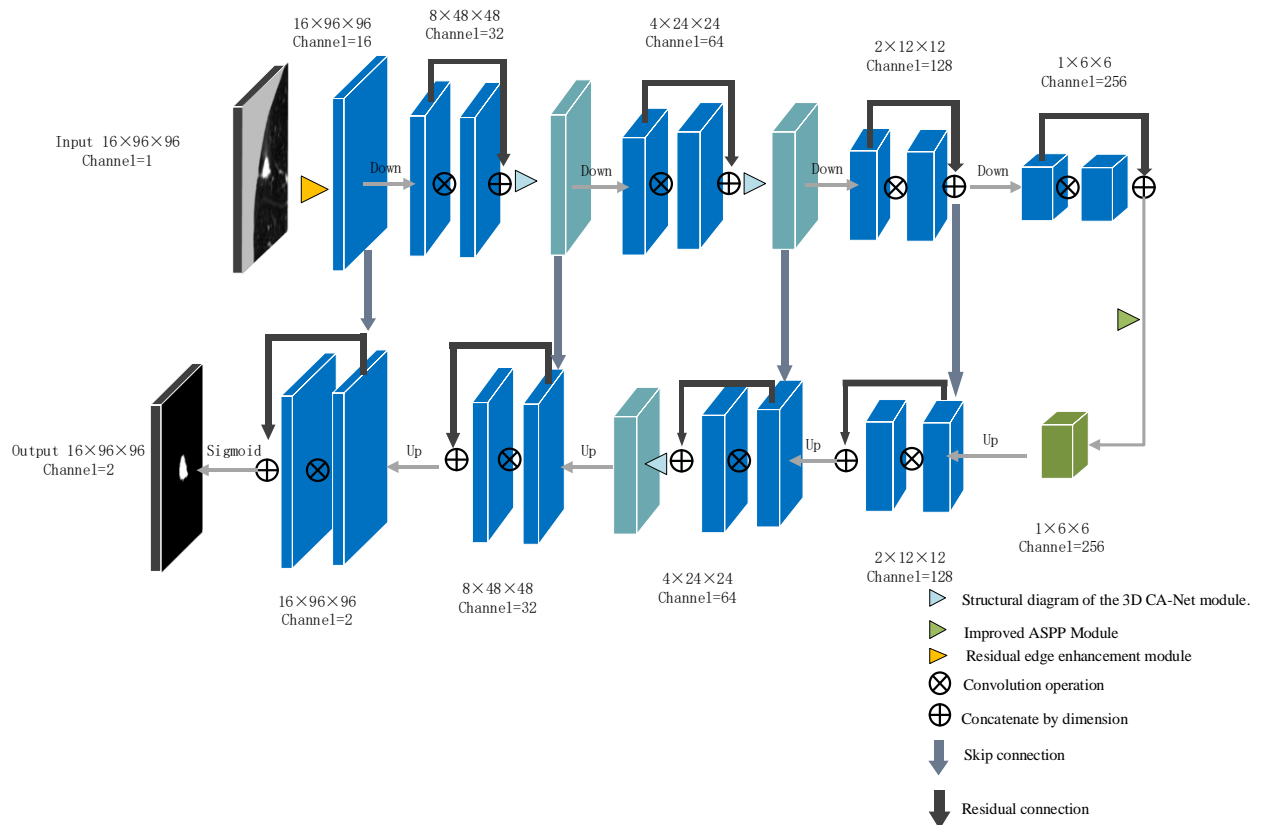
Figure 1. General structural design of the proposed model.

3.2. Residual edge enhancement module

Edge detection helps identify salient features and structures in an image, reduce complex textures and background noise to a certain extent, enhance the contours and features of the image, and separate the boundaries of different objects to locate the boundaries more accurately, all of which improve segmentation accuracy. The traditional segmentation algorithms are simple but less effective at segmenting nodules with complex backgrounds. Deep learning methods utilize convolutional neural networks to perform feature extraction; after continuously applying downsampling, the model loses boundary information, resulting in low model attention. The SCA-VNet segmentation model combines the traditional edge segmentation algorithm with the deep learning approach. First, an edge detection operator is used to sharpen the edges of the original image, which reduces the similarity between the foreground and background grayscales by strengthening the edge features. Then, the edge-enhanced image and the original image are spliced together, which ensures that the original image features are not lost. Finally, the residual module is used to fuse the spliced image with the convolved feature extraction map. This approach can strengthen the edge detail information while retaining the global feature information of the original image. Moreover, it can alleviate the redundancy of the information in the input module of the baseline VNet model used in this paper and make full use of the input information. Figure 2 shows the structure of the residual edge enhancement module.
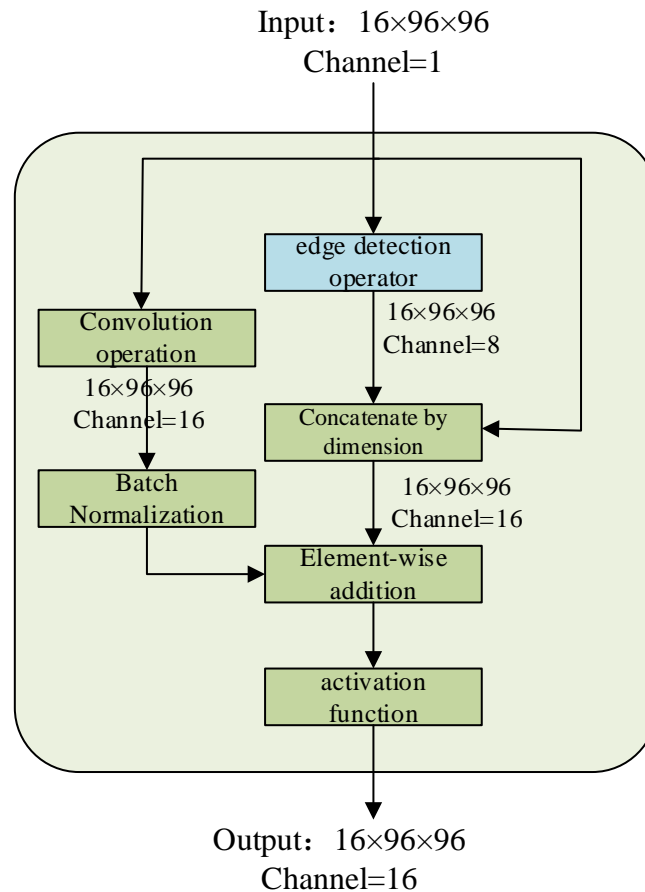
Input：16×96×96
Channel=1

edge detection operator

16×96×96
Channel=8

Convolution operation

16×96×96
Channel=16

Concatenate by dimension

16×96×96
Channel=16

Batch Normalization

Element-wise addition

activation function

Output：16×96×96
Channel=16

Figure 2. Residual edge enhancement module.

3.3. 3D coordinate attention module

Coordinate attention embeds position information into channel attention, models channel correlations and remote dependencies, complements channel- and direction-related position information, and enhances the ability of the model to express input features. The 2D coordinate attention mechanism decomposes the observed channel information into two 1D feature encoding processes that aggregate features in different directions. This approach has the advantage of capturing long-range dependencies along one spatial direction and retaining precise position information along the other direction. The resulting feature maps are subsequently encoded separately to form a pair of direction-aware and location-sensitive feature maps, which can be complementarily applied to the input feature map to enhance the representation of the target of interest.

To adapt this type of mechanism to the 3D lung nodule segmentation task, we use a three-dimensional channel attention module (CA-Net). This module performs an average pooling operation on each dimension of the feature map based on adaptive 3D pooling convolution to efficiently extract the weight information of each channel dimension. The weights are used between the channels to measure the importance of the mapped convolutional features so that the important and useful features are reinforced during the feature extraction process to improve the accuracy of the model.

The CA attention mechanism first inputs channel information. It aggregates the input features in each three-dimensional direction into three separate direction-aware feature maps using a 3D adaptive

average pooling operation. For an input X, three perceptual feature maps are first generated by encoding each channel in the horizontal and vertical depth directions using pooling kernels with dimensions of (d,1,1), (1,h,1), and (1,1,w).

The output depth expression for the cth channel with a depth of d is shown in Eq (1) below:

$$Z_c^d(d) = \frac{1}{w \times h} \sum_{0 \leq i \leq (w \times h)} X_c(d, i) \tag{1}$$

The output height expression for the cth channel with a height of h is shown in Eq (2) below:

$$Z_c^h(h) = \frac{1}{w \times d} \sum_{0 \leq j \leq (w \times d)} X_c(h, j) \tag{2}$$

The output width expression of the cth channel with a width of w is shown in Eq (3) below:

$$Z_c^w(w) = \frac{1}{h \times d} \sum_{0 \leq k \leq (h \times d)} X_c(w, k) \tag{3}$$

The three feature maps generated above are cascaded, and then a shared $1 \times 1 \times 1$ convolution is used to perform a transformation F_1 to generate intermediate feature maps for spatial information acquisition in the depth, height, and width directions f. This transformation is expressed as shown in Eq (4) below:

$$f = \delta\big(F_1([Z^d, Z^h, Z^w])\big) \tag{4}$$

Then, f is sliced into three separate tensors f^d, f^h and f^w along the spatial dimension, and the sizes of f^d, f^h and f^w are increased to the number of input data channels using the three $1 \times 1 \times 1$ convolutions f^d, f^h and f^w, respectively, as shown in Eqs (5)−(7):

$$g^d = \sigma\left(F_d(f^d)\right) \tag{5}$$

$$g^h = \sigma\left(F_h(f^h)\right) \tag{6}$$

$$g^w = \sigma\left(F_w(f^w)\right) \tag{7}$$

Finally, after executing the activation function with g^d, g^h and g^w as the attention weights, the final output of the 3D channel attention module can be expressed as shown in Eq (8):

$$y_c(i, j, k) = x_c(i, j, k) \times g_c^d(i) \times g_c^h(j) \times g_c^w(k) \tag{8}$$

c represents the channel, x_c represents the feature graph matrix for channel c, and this formula represents the output of 3D CA-Net for channel c.

Adaptive average pooling (AAP) does not require a preset pooling kernel size or additional parameters to adapt to inputs with different sizes. It reduces the number of model parameters needed to retain more global and local spatial information, helps to reduce the risk of model overfitting, and improves the generalizability of the model. Figure 3 shows the structural diagram of the 3D CA-Net module.
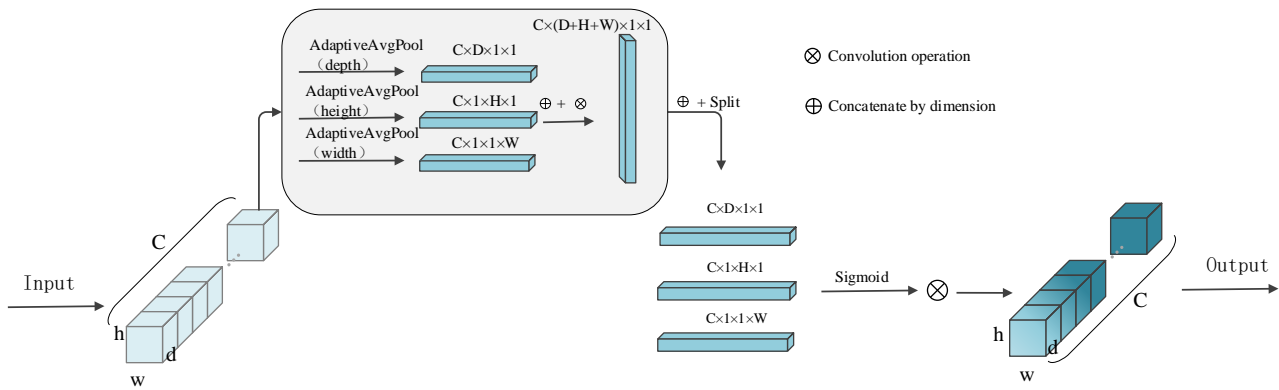
Figure 3. Structural diagram of the 3D CA-Net module.

3.4. Atrous spatial convolutional pooling pyramid

The ASPP module is a neural network module for image semantic segmentation tasks that was originally proposed for use with DeepLabV2 [23]. It aims to capture semantic information at different sensory field scales. The main idea of the ASPP module is to perform feature extraction with different null convolution rates to capture contextual information at different scales and then fuse these features. This helps the model better understand the objects and structures of an image at different scales.
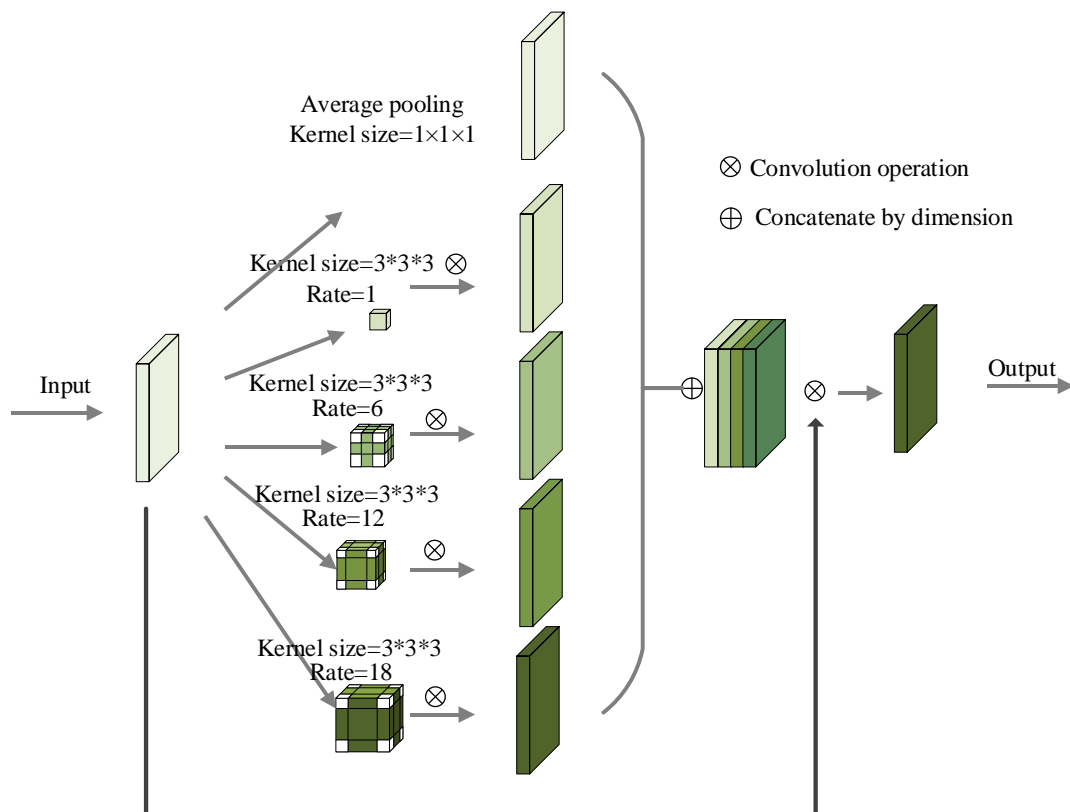

Figure 4. Improved ASPP module.

ASPP was proposed to address the problem that traditional convolutional networks only focus on a single class of lesion feature information in segmentation tasks while ignoring the contextual information of different lesion classes. ASPP enhances the sensory field of the model and thus improves the output segmentation results by performing convolutional operations in parallel at different nulling rates (known as expansion rates), thus obtaining contextual information at multiple scales. Smaller expansions are used to capture local details, medium expansions are used to capture local structures and targets, and larger expansions are used to capture global contextual information. Moreover, combined with the small sizes of the input image blocks in lung nodule segmentation tasks, a $3 \times 3 \times 3$ convolution kernel is used, and the expansion rates are set to 1, 6, 12, and 18. Then, the branches with the largest expansion rate are removed to adapt to and extract the image features related to the lung nodule; the improved ASPP module is shown in Figure 4.

4. Experimental results and analysis

This section describes the employed dataset, data processing method, evaluation metrics, and experimental setup before presenting the experimental results.

4.1. Introduction to the dataset

We used the lung image database consortium and image database resource initiative (LIDC-IDRI) [24], the largest publicly available dataset of lung CT images in existence. This dataset consists of 1018 study cases derived from seven academic centers. Up to four physicians labeled the lesion information in the study cases. The screening process applied to the data samples in this paper followed the sample screening strategy of the Luna16 dataset, excluding 9 patients with inconsistent slice spacing or missing slices. To facilitate data analysis to effectively detect lung nodules with diameters exceeding 3 mm, the experiments in this paper excluded 121 patients with CT layer thickness values that were greater than 2.5 mm and screened 888 available patients [25] for a total of 1186 training samples of lung nodules [26].
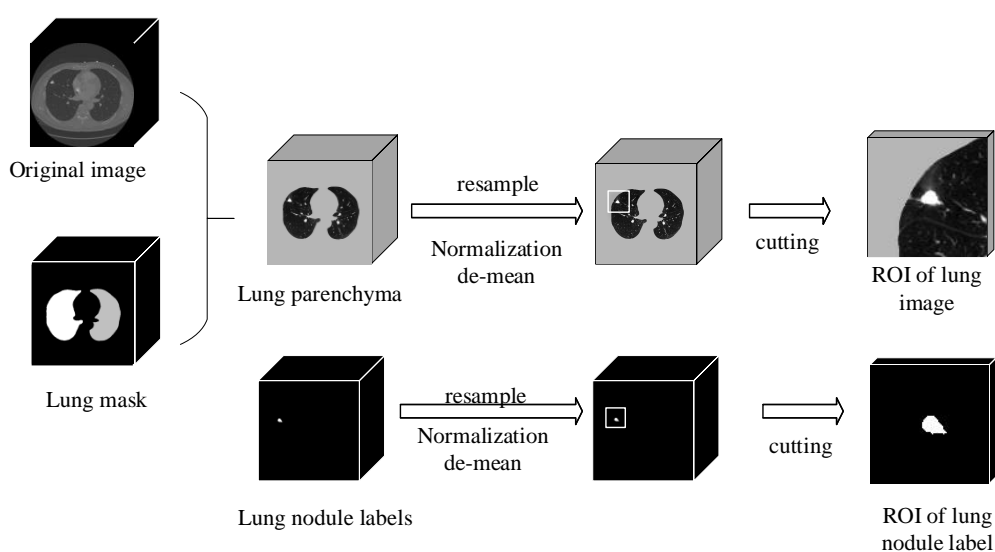

Figure 5. Image preprocessing flowchart.

Before conducting the experiments, data preprocessing was performed on the LIDC-IDRI dataset. The preprocessing steps were as follows. (1) The lung parenchyma was extracted, and lung nodule segmentation was performed to obtain labeled images. (2) The data were resampled, and the pixel intervals were standardized. Due to the different collection sources of the dataset, the pixel intervals of all the data were standardized to prevent experimental errors. (3) The pixel values of the data were normalized and de-meaned. To facilitate the model calculation process, the value range of the adjusted lung window for CT was set to [−1000,400], and the image pixel values were linearly transformed to be within the range of [0,1] using deviation normalization. Finally, a de-meaning operation was performed on the normalized data. Due to the computational resource limitations, $16 \times 96 \times 96$ data blocks were used as model inputs [8]. Figure 5 shows the image preprocessing flowchart.

4.2. Evaluation methodology

The accuracy (Acc), positive predictive value (PPV), Dice similarity coefficient (Dice), intersection over union (IoU), and sensitivity (SEN) evaluation indices [27] were used, among which the Dice coefficient was the most important. The formula for each index is as follows:

$$Acc = \frac{TP+TN}{TP+FP+TN+FN} \tag{9}$$

$$PPV = \frac{TP}{TP+FP} \tag{10}$$

$$Dice = \frac{2TP}{2TP+FP+FN} \tag{11}$$

$$IoU = \frac{TP}{TP+FP+FN} \tag{12}$$

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

TP is the number of correctly predicted positive samples, that is, true positives; TN is the number of negative samples predicted as negative, that is, true negatives; FN is the number of positive samples predicted as negative, that is, false negatives; and FP is the number of negative samples predicted as positive, that is, false positives [27].

4.3. Experimental settings

In this paper, the above network model was used to segment the lung nodules observed in lung CT images. The experiment used the LIDC-IDRI dataset for model training and increased the number of training samples by translation, rotation, and other methods to improve the robustness of the network. The PyTorch1.12.1 framework was used to construct the network model. During the experiment, the initial learning rate was set to 0.001, 10% of the dataset was used for testing, 90% of the dataset was used for training the network model, and the cross-entropy loss function and Adam optimizer were employed [28].

4.4. Experimental results and analysis

4.4.1. Verification of the edge detection operators

The edge enhancement operation was performed on the original LIDC-IDRI dataset using different edge detection operators on to select the most appropriate edge detection operator. Figure 6 visualizes the results obtained for the unprocessed LIDC-IDRI image and the edge-enhanced images generated by different edge detection operators.
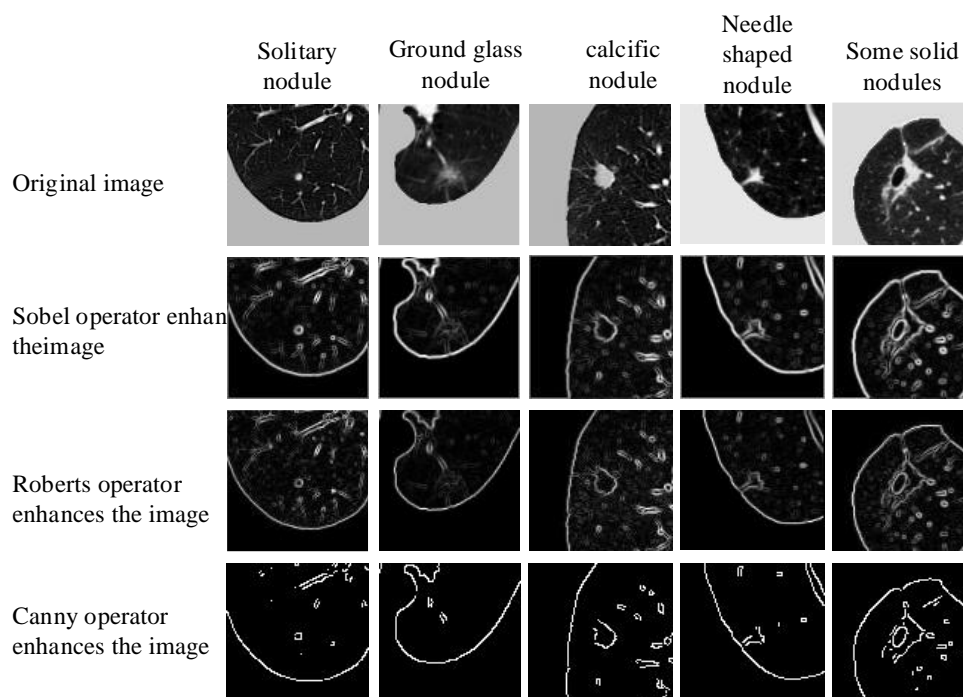

Figure 6. Visualization of the original image and the edge-enhanced images generated by edge detection operators.

As Figure 6 shows, both the Sobel and Roberts operators have strong abilities to sharpen the edges of different types of lung nodules. When the highest gray level threshold of the Canny operator was set to 650 and the lowest gray level threshold was set to 500, the edges of the nodules were clearly identified, and the background was effectively screened. Additionally, the edge boundaries of solitary, calcified, needle and some solid nodules were more clearly detected. However, the edge detection effect attained for nonsolid ground-glass nodule structures was poor, and their boundaries could not be identified.

To select the most suitable edge detection operator for lung nodule segmentation, an ablation experiment was designed on the LIDC-IDRI dataset. In the experiment, the original VNet was selected as the basic model. The Sobel, Roberts and Canny operators were used in the residual edge enhancement module, and the corresponding models are represented as VNet+Sobel, VNet+Roberts, and VNet+Canny, respectively.

Table 1. Comparison among the segmentation results produced by different edge enhancement operators on the LIDC-IDRI dataset.

Model	Acc (%)	Dice (%)	IoU (%)	Sen (%)	PPV (%)
VNet	99.77	85.98	75.41	**88.43**	83.71
VNet+Sobel	**99.80**	**86.76**	**76.62**	85.15	**88.47**
VNet+Roberts	99.78	86.47	76.17	86.19	86.84
VNet+Canny	99.78	85.91	75.30	85.59	86.28

The segmentation results are shown in Table 1. The segmentation results show that the residual edge enhancement module impacted the segmentation results obtained for the lung nodules. The best edge enhancement operator was the Sobel operator, and its segmentation Dice coefficient and IoU were greater than those of the basic model. This means that the residual edge enhancement module based on the Sobel operator could compensate for the loss of edge information during the segmentation process of the original model to a certain extent. The integrity of the lung nodule segmentation procedure was ensured. The segmentation results of the VNet+Canny model were worse than those of the VNet+Sobel and VNet+Roberts models because some types of pulmonary nodule boundaries, such as nonsolid ground-glass pulmonary nodules, have fuzzy boundaries. The Canny operator could not detect them, and the model could not obtain the edge information of some pulmonary nodules with fuzzy edges during training. Compared with that of the basic model in the input module, the fuzzy edge feature information of the input data enhanced by the Canny operator was reduced, so the resulting segmentation effect was slightly worse than that of the basic model. The experimental results showed that the Sobel operator was the most suitable edge detection operator for lung nodule segmentation, and the edge detection operators used in the residual edge enhancement module in subsequent experiments were all Sobel operators.

4.4.2. Validation of the proposed optimization module

To verify that the proposed module contributes to improving the segmentation accuracy achieved for lung nodules, ablation experiments involving the proposed SCA-VNet model were performed on the LIDC-IDRI dataset. The original VNet was chosen as the baseline network, and ablation experiments were performed on each module to prove its effectiveness. In this group of experiments, Experiment 1 considered the base model. Experiment 2 added the residual edge enhancement module to the base model. Experiment 3 added the residual edge enhancement module and the 3D CA-Net module. Experiment 4 added all the proposed modules, i.e., the proposed SCA-VNet model. The specific segmentation results are shown in Table 2.

Table 2. Results of ablation experiments conducted for each module on the LIDC-IDRI dataset.

Model				Acc	Dice	IoU	Sen	PPV
Baseline	Edge-enhanced	3D CA-Net	Improved ASPP	(%)	(%)	(%)	(%)	(%)
√				99.77	85.98	75.41	**88.43**	83.71
√	√			99.80	86.76	76.62	85.15	88.47
√	√	√		99.80	87.02	77.02	85.04	**89.13**
√	√	√	√	**99.81**	**87.50**	**77.80**	86.80	88.32

Table 2 shows that the proposed module facilitated lung nodule segmentation. Additionally, the proposed module increased the lung nodule segmentation accuracy of the model. The experimental results show that the Dice index, IoU and Acc improved with the addition of each module proposed in this paper. The IoU index of the SCA-VNet model was 2.39% greater than that of the baseline model. The Sen index of the baseline model was higher than that of the model proposed in this paper, while its PPV index was lower than that of the model proposed in this paper. This indicates that the baseline model is more sensitive to positive sample data, but its segmentation accuracy is not as good as that of the SCA-VNet model. The above experiments prove that the model proposed in this paper effectively segments lung nodules.

4.4.3. Comparison with the existing methods

Table 3 compares the results of the existing lung nodule segmentation methods in the literature with those of the method proposed in this paper. It also summarizes the models, datasets, input sizes, numbers of samples, segmentation metrics, and lung nodule sample selection strategies used by the different methods [27]. Among the existing lung nodule segmentation methods, the UGS-Net, REMU-Net, and multitask methods achieved better segmentation results in the lung nodule segmentation task. The method proposed in this paper selected the LIDC-IDRI dataset for training to validate its segmentation effect. It achieved 87.50% accuracy, and its segmentation results were better than those of the existing lung nodule segmentation methods in the literature.

4.4.4. Comparison with the existing models

To further verify the effectiveness of the proposed method, this paper compared it with the open-source VNet, PCAM_Net [36], Med3D [37], and CENet [38] models that were developed in recent years for 3D medical image segmentation tasks. The segmentation results of each model are shown in Table 4. Among the open-source segmentation models, the PCAM_Net model yielded the best segmentation effect, with its Dice and IoU indices reaching 86.07% and 75.55%, respectively. The Med3D model had poorer segmentation effects than those of the other models, with Dice and IoU metrics reaching 84.13% and 72.61%, respectively. The Dice and IoU indices of the proposed method were 1.49% and 2.25% greater than those of the PCAM_Net model, respectively. The Dice and IoU metrics achieved in this paper were 3.37% and 5.19% greater than those of the Med3D model, respectively.

Table 3. Comparison among different lung nodule segmentation methods. IS indicates the size of the input data, NS denotes the number of data samples for model training, and NSS indicates the lung nodule selection strategy applied on the dataset.

Method	Dataset	IS	NS	Dice (%)	NSS
UGS-Net [29]	LIDC-IDRI	64×64	1859	86.12	Number of doctors > 1
REMU-Net [5]	LIDC-IDRI	64×64	1487	84.76	Number of doctors > 2
SAtUNet [30]	LIDC-IDRI	64×64	3132	81.10	All data in LIDC-IDRI
ResAANet [31]	Private1	256×256	565(train)	—	CT layer thickness = 0.625-3.0 mm, ground-glass structure
	LIDC-IDRI		145(test)	83.36	CT layer thickness = 0.45−5.0 mm, ground-glass structure
	Private2		84(test)	83.46	CT layer thickness 0.625−1.0 mm, ground-glass structure
LeisionNet [32]	LIDC-IDRI	$128 \times 128 \times 128$	1131	80.89	Nodule diameter ≥ 3 mm, marked by at least three doctors
CRU^{2}Net [27]	LIDC-IDRI	$64 \times 64 \times 64$	1186	83.83	Nodule diameter ≥ 3 mm, CT layer thickness < 2.5 mm
MS-UNet [33]	LIDC-IDRI	$64 \times 96 \times 96$	1625	77.40	Number of doctors > 3
	LNDb		1968	70.62	All data in LNDb
	Private		6864	79.62	2 mm $<$ Nodule diameter < 64 mm
H-DL [26]	LIDC-IDRI	$64 \times 64 \times 64$	2885	75±13.5	7 mm $<$ Nodule diameter < 45 mm, Number of doctors ≥ 2
CSE-GAN [34]	Luna16	$64 \times 64 \times 64$	888	80.74	CT layer thickness < 2.5 mm, CT with missing slices removed
	Private		113	76.36	3 mm $<$ Nodule diameter < 30 mm, CT with missing slices removed
Multitask [35]	LIDC-IDRI	$3 \times 96 \times 96$	2616	86.43	—
Method of this paper(SCA-VNet)	LIDC-IDRI	$16 \times 96 \times 96$	1186	87.50	Nodule diameter ≥ 3 mm, CT layer thickness < 2.5 mm

Table 4. Comparison between the segmentation results of the proposed method with those of open-source 3D medical image segmentation models.

Model	Acc (%)	Dice (%)	IoU (%)	Sen (%)	PPV (%)
VNet	99.77	85.98	75.41	**88.43**	83.71
PCAM_Net	99.79	86.07	75.55	86.23	86.02
Med3D	99.76	84.13	72.61	82.81	85.51
CENet	99.78	85.30	74.40	81.10	**90.07**
SCA-Vnet	**99.81**	**87.50**	**77.80**	86.80	88.32

Figures 7 and 8 show box plots of the obtained Dice and IoU indicators, respectively, for evaluating the segmentation results of the proposed method and different open-source 3D segmentation models. From these plots, we can see that among the models tested in this paper, the PCAM_Net and Med3D models were more stable, and the Dice and IoU values of the model proposed in this paper were greater than those of the other models. The proposed model not only was more stable during training but also outperformed the other segmentation models.
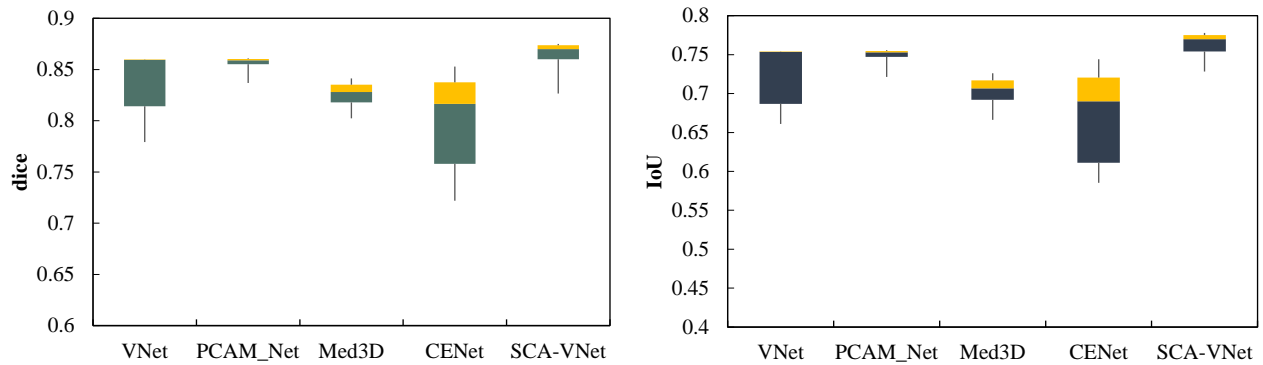

Figure 7. Box plots of the Dice coefficients and IoU values produced by the method proposed in this paper and the open-source 3D segmentation models.

Figure 8 shows the loss plots induced by the training results of the proposed method and the open-source 3D segmentation methods. This figure shows that the Vnet model had the slowest convergence speed, the loss indices of the CENet and Med3D models were relatively turbulent during training, and the PCAM_Net and SCA-Vnet models exhibited fast convergence speeds and stable loss indices during training.
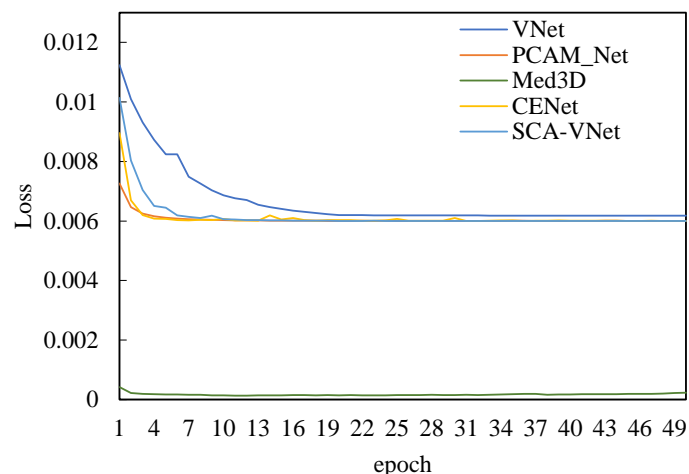

Figure 8. Training loss curves of the proposed method and the open-source 3D segmentation models.
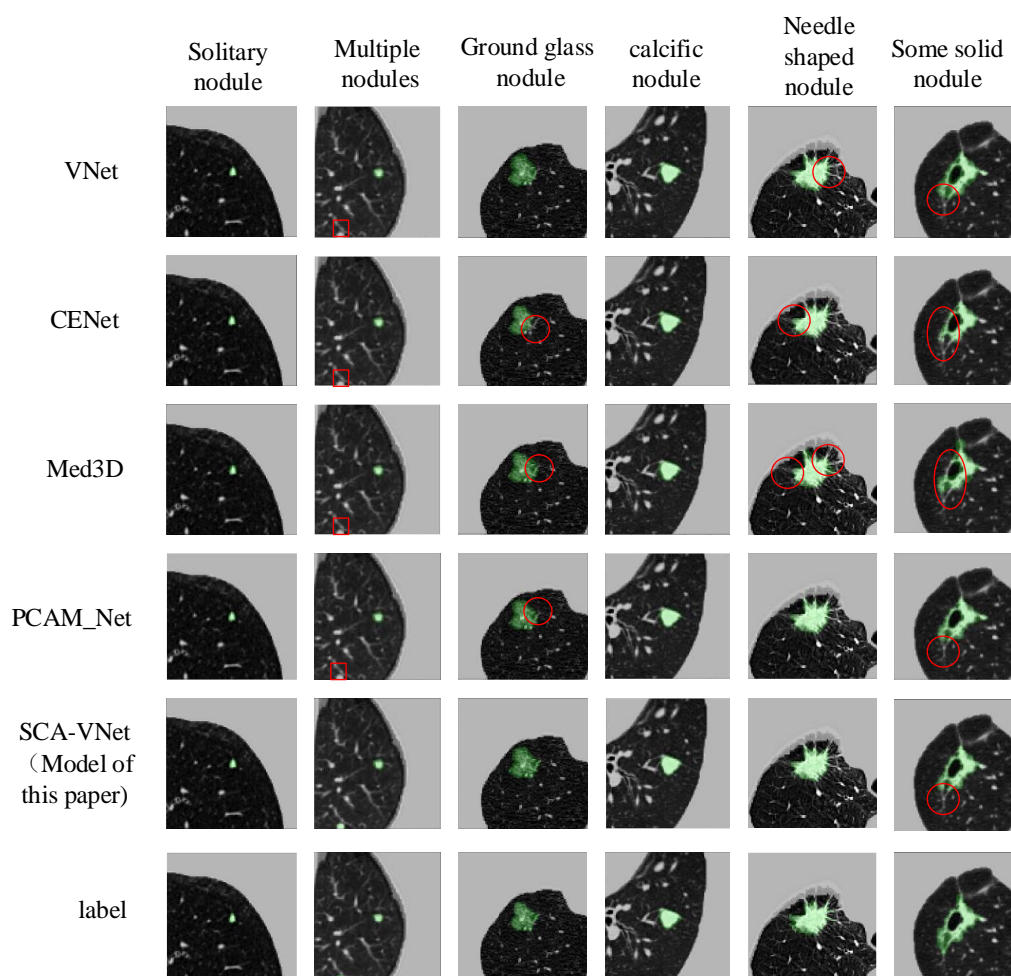
4.4.5. Segmentation results

Figure 9. Comparison among the segmentation results of different segmentation models.

Figure 9 shows the obtained segmentation results, from which it can be seen that each model had a strong effect on segmenting isolated and calcified nodules. The backgrounds of the isolated nodules were relatively simple, and compared with other types of features, they were more obvious and easier to segment. Additionally, the edges of the calcified nodules had more obvious contours and were easier to segment. The segmentation result graph shows that only the method proposed in this paper segmented all the nodes in the multimode graph, and all the other models yielded missed detections. The segmentation results of the Vnet model and the proposed model were better than those of the CENet, Med3D, and PCAM_Net models. PCAM_Net and the proposed model outperformed the other models in terms of segmenting specular nodules. When segmenting partial solid nodules, the detection effects of PCAM_Net and the proposed method were better than those of the other three models. In general, the segmentation effect of the segmentation method proposed in this paper was better than those of the other segmentation models, and the edge contour segmentation results of the proposed model were more accurate.

5. Conclusions and ideas for future work

The SCA-VNet segmentation model proposed in this paper combines the Sobel edge detection operator with a residual edge enhancement module. This method can not only fully use the edge information contained in lung CT images but also replace part of the information-redundant data with edge-enhanced data during the input phase, which can mitigate the information redundancy problem encountered when splicing the input data of the model from a channel perspective. Ablation experiments showed that the proposed method and module achieved improved lung nodule segmentation accuracy. A 3D CA-Net module is used to enhance the performance of the convolutional neural network so that the model can focus on pulmonary nodule-related channels and spatial locations in the input image. The improved ASPP module has receptive fields with different scales and effectively extracts nodule information at these different scales. Experiments showed that the segmentation results of the model were significantly improved after adding the 3D CA-Net module and the improved ASPP module. A comparison with a variety of lung nodule segmentation methods in the literature and several open-source 3D segmentation models revealed that the proposed method is superior to the existing lung nodule segmentation methods. This study improves the automation level of the existing computer-aided diagnosis systems and more effectively assists radiologists in clinical practice. The algorithm in this paper has room for improvement and was only analyzed based on the training data. In the experiment, LIDC-IDRI was selected as the training dataset for the model. Although the scale of the training data was large, only a single input image (a CT image) was used. Segmentation methods based on transformer technology have received extensive attention in the medical segmentation domain. Subsequent research can use a combination of transformer technology and convolutional neural networks.

AI tool use declaration

The authors declare that they have not used Artificial Intelligence (AI) tools for the creation of this article.

Acknowledgments

Conflicts of interest

The authors declare that there are no conflicts of interest.

References

1. Q. Zhou, Y. Fan, Y. Wang, Y. Qiao, G. Wang, Y. Huang, et al., Chinese national guidelines for classification, diagnosis and treatment of pulmonary nodules (2016 version), *Chin. J. Lung Cancer*, **19** (2016), 793−798. https://doi.org/10.3779/j.issn.1009-3419.2016.12.12

2. T. Dong, L. Wei, S. Nie, Research progress of pulmonary nodule segmentation in CT images, *J. Image Graphic*, **26** (2021), 751−765. https://doi.org/10.11834/jig.200201

3. O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, (2015), 234−241. https://doi.org/10.1007/978-3-319-24574-4_28

4. F. Milletari, N. Navab, S. Ahmadi, V-net: Fully convolutional neural networks for volumetric medical image segmentation, in *2016 Fourth International Conference on 3D Vision (3DV)*, (2016), 565−571. https://doi.org/10.1109/3DV.2016.79

5. D. Li, S. Yuan, G. Yao, Pulmonary nodule segmentation based on REMU-Net, *Phys. Eng. Sci. Med.*, **45** (2022), 995−1004. https://doi.org/10.1007/s13246-022-01157-9

6. P. Dutande, U. Baid, S. Talbar, LNCDS: A 2D-3D cascaded CNN approach for lung nodule classification, detection and segmentation, *Biomed. Signal Process. Control*, **67** (2021), 102527. https://doi.org/10.1016/j.bspc.2021.102527

7. A. A. Sundaresan, A. Jeevanayagam, Efficient multiscale fully convolutional UNet model for segmentation of 3D lung nodule from CT image, *J. Med. Imag.*, **9** (2022), 052402. https://doi.org/10.1117/1.JMI.9.5.052402

8. Z. Wang, J. Men, F. Zhang, Improved V-Net lung nodule segmentation method based on selective kernel, *Signal Image Video Process.*, **17** (2023), 1763−1774. https://doi.org/10.1007/s11760-022-02387-w

9. Z. Zhou, F. Gou, Y. Tan, J. Wu, A cascaded multi-stage framework for automatic detection and segmentation of pulmonary nodules in developing countries, *IEEE J. Biomed. Health Inf.*, **26** (2022), 5619−5630. https://doi.org/10.1109/JBHI.2022.3198509

10. W. Zhou, F. Zheng, Y. Zhao, Y. Pang, J. Yi, MSDCNN: A multiscale dilated convolution neural network for fine-grained 3D shape classification, *Neural Networks*, **172** (2024), 106141. https://doi.org/10.1016/j.neunet.2024.106141

11. Y. Wang, W. Zhong, H. Su, F. Zheng, Y. Pang, H. Wen, et al., An improved mvcnn for 3D shape recognition, in *2021 IEEE International Conference on Emergency Science and Information Technology (ICESIT)*, (2021), 469−472. https://doi.org/10.1109/ICESIT53460.2021.9696941

12. A. Lisowska, Efficient edge detection method for focused images, *Appl. Sci.*, **12** (2022), 11668. https://doi.org/10.3390/app122211668

13. A. A. De Borba, A. Muhuri, M. Marengoni, A. C. Frery, Feature selection for edge detection in PolSAR images, *Remote Sens.*, **15** (2023), 2479. https://doi.org/10.3390/rs15092479

14. S. R. Hait, R. Mesiar, P. Gupta, D. Guha, D. Chakraborty, The Bonferroni mean-type pre-aggregation operators construction and generalization: Application to edge detection, *Inf. Fusion*, **80** (2022), 226−240. https://doi.org/10.1016/j.inffus.2021.11.002

15. S. Wang, A. Jiang, X. Li, Y. Qiu, M. Li, F. Li, DPBET: A dual-path lung nodules segmentation model based on boundary enhancement and hybrid transformer, *Comput. Biol. Med.*, **151**(2022), 106330. https://doi.org/10.1016/j.compbiomed.2022.106330

16. Y. Jiang, Pulmonary nodule segmentation algorithm based on Vnet and edge features, *Chin. J. Med. Phys.*, **39** (2022), 705−712. https://doi.org/10.3969/j.issn.1005-202X.2022.06.009

17. Y. Zhang, Y. Li, Y. Zhang, W. Hu, H. Liu, X. Gu, Research progress of laser spot edge detection operator, *J. Quantum Opt.*, (2019), 109−116. https://doi.org/10.3788/JQO20192501.0901

18. Q. Hou, D. Zhou, J. Feng, Coordinate attention for efficient mobile network design, in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2021), 13713−13722. https://doi.org/10.1109/CVPR46437.2021.01350

19. W. An, G. Wu, Hybrid spatial-channel attention mechanism for cross-age face recognition, *Electronics*, **13** (2024), 1257. https://doi.org/10.3390/electronics13071257

20. J. Chen, S. Ye, Z. Jiang, Z. Fang, Image deblurring using feedback mechanism and dual gated attention network, *Neural Process. Lett.*, **56** (2024), 88. https://doi.org/10.1007/s11063-024-11462-x

21. Y. Li, Z. Zhou, G. Qi, G. Hu, Z. Zhu, X. Huang, Remote sensing micro-object detection under global and local attention mechanism, *Remote Sens.*, **16** (2024), 644. https://doi.org/10.3390/rs16040644

22. Z. Cao, R. Li, X. Yang, L. Fang, Z. Li, J. Li, Multi-scale detection of pulmonary nodules by integrating attention mechanism, *Sci. Rep.*, **13** (2023), 5517. https://doi.org/10.1038/s41598-023-32312-1

23. L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A. L. Yuille, Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs, *IEEE Trans. Pattern Anal. Mach. Intell.*, **40** (2017), 834−848. https://doi.org/10.1109/TPAMI.2017.2699184

24. A. Iii, S. G. Mclennan, G. Bidaut, L. Mcnitt-Gray, M. F. Meyer, C. R. Reeves, 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** (2011), 915−931. https://doi.org/10.1118/1.3528204

25. A. A. A. Setio, A. Traverso, T. De Bel, M. S. Berens, C. Van Den Bogaard, P. Cerello, et al., Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge, *Med. Image Anal.*, **42** (2017), 1−13. https://doi.org/10.1016/j.media.2017.06.015

26. Y. Wang, C. Zhou, H. P. Chan, L. M. Hadjiiski, A. Chughtai, E. A. Kazerooni, Hybrid U-Net-based deep learning model for volume segmentation of lung nodules in CT images, *Med. Phys.*, **49** (2022), 7287−7302. https://doi.org/10.1002/mp.15810

27. W. Jiang, L. Zhi, S. Zhang, T. Zhou, Segmentation of pulmonary nodules in CT images based on channel residual nested U structure, *J. Graphics*, **44** (2023), 879−889. http://www.txxb.com.cn/CN/10.11996/JG.j.2095-302X.2023050879

28. D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, preprint, arXiv: 1412.6980.

29. H. Yang, L. Shen, M. Zhang, Q. Wang, Uncertainty-guided lung nodule segmentation with feature-aware attention, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, (2022), 44−54. https://doi.org/10.1007/978-3-031-16443-9_5

30. S. Selvadass, P. M. Bruntha, K. M. Sagayam, H. Günerhan, SAtUNet: Series atrous convolution enhanced U-Net for lung nodule segmentation, *Int. J. Imaging Syst. Technol.*, (2023). https://doi.org/10.1002/ima.22964

31. D. Ting, W. Long, Y. Xiaodan, C. Yang, H. Xuewen, N. Shengdong, A full convolution residual network ground-glass pulmonary nodule segmentation method based on empty space convolution pooling pyramid structure and attention mechanism, *J. Biomed. Eng.*, **39** (2022), 11. https://doi.org/10.7507/1001-5515.202010051

32. X. Yi, X. Jun, X. Gang, X. Xinying, Multi-task pulmonary nodule detection and segmentation with attention feature fusion, *Comput. Eng. Design*, **43** (2022), 8. https://doi.org/10.16208/j.issn1000-7024.2022.09.017

33. Z. Li, J. Yang, Y. Xu, L. Zhang, W. Dong, B. Du, Scale-aware test-time click adaptation for pulmonary nodule and mass segmentation, in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2023*, (2023), 681−691. https://doi.org/10.1007/978-3-031-43898-1_65

34. S. Tyagi, S. N. Talbar, Medicine, CSE-GAN: A 3D conditional generative adversarial network with concurrent squeeze-and-excitation blocks for lung nodule segmentation, *Comput. Biol. Med.*, **147** (2022), 105781. https://doi.org/10.1016/j.compbiomed.2022.105781

35. W. Liu, X. Liu, H. Li, M. Li, X. Zhao, Z. Zhu, Integrating lung parenchyma segmentation and nodule detection with deep multi-task learning, *IEEE J. Biomed. Health Inf.*, **25** (2021), 3073−3081. https://doi.org/10.1109/JBHI.2021.3053023

36. D. Liang, J. Liu, K. Wang, G. Luo, W. Wang, S. Li, Position-prior clustering-based self-attention module for knee cartilage segmentation, in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2022*, (2022), 193−202. https://doi.org/10.1007/978-3-031-16443-9_19

37. S. Chen, K. Ma, Y. Zheng, Med3D: Transfer learning for 3D medical image analysis, preprint, arXiv:1904.00625.

38. Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, et al., Ce-net: Context encoder network for 2d medical image segmentation, *IEEE Trans. Med. Imaging*, **38** (2019), 2281−2292. https://doi.org/10.1109/TMI.2019.2903562