Received August 17, 2020, accepted August 23, 2020, date of publication September 7, 2020, date of current version October 1, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3022249

Ultrasound Image Segmentation Method for Thyroid Nodules Using ASPP Fusion Features

YATING WU[1], **XUELIANG SHEN**[2], **FENG BU**[3], **AND JIN TIAN**[1]

[1]Department of Ultrasound, General Hospital of Ningxia Medical University, Yinchuan 750000, China
[2]Department of Otorhinolaryngology Head and Neck Surgery, General Hospital of Ningxia Medical University, Yinchuan 750000, China
[3]School of Electronic and Information Engineering, Soochow University, Suzhou 215006, China

Corresponding author: Jin Tian (erya916@126.com)

ABSTRACT Ultrasound imaging technology plays an important role to assist doctors in diagnosing thyroid nodules. The tissue structure around the thyroid is very complex, which makes it difficult to segment and extract the ultrasound image of thyroid nodules accurately. For address this problem, this paper proposes a model algorithm for thyroid nodule ultrasound image segmentation using ASPP fusion features. First, spatial pyramid pooling and depthwise separable convolution are combined in order to solve the problem that the size of the mapping feature will change in the process of better capturing the context information. Besides, Atrous Spatial Pyramid Pooling (ASPP) is proposed to achieve the purpose of processing input image channel and spatial information separately. In order to appropriately reduce the dimension and size of feature images, a 1 × 1 convolution operation is performed before each convolution calculation, and the model size is optimized. In the decoding stage, decoder module appropriately adjusts the feature map with a relatively low resolution previously from decoder module, and sets the output channel number of two convolutions to the same value. All features have the same dimension by adjustment, and features can be fused by element-wise summation. Finally, Dice Similarity Coefficient (DSC), Prevent Match (PM) and Correspondence Patio (CR) are used as evaluation criteria to compare with other model algorithms. The experimental results show that the proposed model can significantly improve the segmentation effect of ultrasound images for thyroid nodules compared with traditional models.

INDEX TERMS Thyroid nodule, medical image segmentation, atrous spatial pyramid pooling, dilated convolution, ultrasound image.

I. INTRODUCTION

In recent years, artificial intelligence technology and medical imaging have become more and more closely integrated [1]–[3]. The use of artificial intelligence to process medical images has increasingly become the main research focus. Ultrasound image segmentation is one of the research hotspots [4].

One of the most common diseases of endocrine system is thyroid nodules. Relevant studies have shown that 2% to 6% of adults in areas where iodine is not deficient have thyroid nodules [5], [6]. The incidence of ultrasound images is 19-35% [7], [8]. When segmenting ultrasound images of thyroid nodules, the methods generally used include the following: contour and shape-based segmentation methods, region-based segmentation methods, supervised and unsupervised segmentation methods, hybrid technology-based segmentation methods, threshold-based segmentation methods, segmentation methods based on Markov random field and segmentation methods based on deep learning [9].

Contour and shape-based segmentation methods can be divided into edge-based segmentation methods, probabilistic filtering-based segmentation methods and deformable model-based segmentation methods. In the process of image processing, various gradient filters are usually used to extract image edges. But the extraction process is often affected by noise, and the gradient filter often gets wrong edge results during the detection process. Therefore, it is particularly important to design a suitable algorithm and detect the edge by a large number of calculations [10]–[12]. In the traditional edge detection process, the contrast of ultrasound images is relatively low due to the presence of spots and noise, which causes the edge of shadow areas to be inaccurately obtained [13], [14]. In order to solve the above problems, Kwoh *et al.* used Fourier transform to Fourier decomposition

The associate editor coordinating the review of this manuscript and approving it for publication was Yizhang Jiang.

of images, and reduced false edges by obtained high-order harmonics [15]. In order to reduce the number of spots in original images, Aaraink *et al.* used local standard deviation as the basis to identify homogeneous and non-homogeneous regions in images under a multi-resolution framework. The result can provide a more reliable detection method for remote detection of thyroid images [16]. Yu *et al.* proposed a method to determine the initial contour of an image based on radial base embossing method, and based on this, proposed an algorithm that can remove false edges of images [17]. The algorithm is based on the deformation propagation of two-dimensional slices and can change the contour of each image slice. Gomez *et al.* achieved the enhancement of image contrast through a histogram equalization method with limited contrast. On this basis, the edge of the image is enhanced by an anisotropic diffusion filter. Finally, a watershed-based image segmentation method was proposed, and the boundary extraction of breast ultrasound images was realized [18]. Based on the U-net model, Pan Peike *et al.* realized the segmentation of MRI images of nasopharyngeal tumors. The principle was to obtain surrounding information by the contracted path, and on this basis, achieved precise positioning by expanding the path [19]. Most of the existing researches reduce the size of original images in the image segmentation process, and cannot obtain a full-resolution image. In addition, the shallower network during image segmentation will greatly reduce the accuracy of segmentation results. In order to solve the above problems, Chen *et al.* proposed an algorithm to increase the receptive field by expanding convolution. This algorithm inserted an appropriate number of zero into convolution kernel, which expands the convolution kernel. The expanded convolution kernel can obtain a larger receptive field and maintain the number of kernel parameters. And different expansion rates can extract the characteristics of different sizes of receptive fields. But when the expansion rate increases to a certain extent, this algorithm will fail [20]. Yang M *et al.* spliced the result obtained by expansion convolution algorithm in the previous layer with the result obtained by traditional convolution. Then the spliced image was transferred to the next layer of dilated convolutional layer, and a DenseASPP model was proposed. This model make up for the defect that ordinary dilated convolution will fail when the expansion rate increases to a certain extent, and can gradually increase the receptive field of each layer of dilated convolution. However, this model had certain defects in the application of ultrasound image segmentation, such as the unsmooth edges of segmentation [21]. Kumar *et al.* were based on Convolutional Neural Networks (CNN). In the process of nuclear segmentation in digital microstructure images, the segmentation result of entire images was obtained by predicting the category of each pixel in the form of a sliding window [22]. Lu Qiuju *et al.* proposed a global segmentation method for multi-threshold color image with adaptive step size for the segmentation of multi-threshold color image. This method improved the efficiency of segmentation by reducing the total number of image colors and does not reduce the quality of images. By listing the objective function and solving objective function based on swarm optimization algorithm, the optimal solution for color image threshold segmentation is obtained [23]. Anas *et al.* proposed a real-time prostate segmentation technique based on deep neural network during biopsy. It laid the foundation for the dynamic registration of mp-MRI and ultrasound data. In addition to extracting spatial features by convolutional networks, this technology also used recursive networks to collect and utilize time information between a series of ultrasound images. This system used residual convolution in the recurrent network to improve optimization, and finally proved the usability of fully convolutional neural network on ultrasound images [24].

Most of the methods mentioned in the above references have problems such as low contrast, blurred boundaries and speckle echo. Therefore, it is difficult to achieve ideal results when applied to ultrasound image segmentation of thyroid nodules. The main contribution of this paper is:

1) Based on the DenseNet-121 network structure model and combined with the Atrous Spatial Pyramid Pooling (ASPP), and proposes a new segmentation model for ultrasound images of thyroid nodules. Using spatial pyramid pooling for splicing to form the mapping feature solves the problem of changes in the size of mapping features.

2) In the encoding process, hierarchical feature fusion is proposed to generate semantic feature structure. The experimental results show that the segmentation model method proposed in this paper greatly improves the segmentation effect of ultrasound images of thyroid nodules. Besides, its performance is better than other comparison methods.

II. BASIC NETWORK STRUCTURE

The encoder-decoding structure is required to gradually reduce the spatial dimension of input data. Secondly, the structure can gradually restore the details of target and the spatial dimension of target based on the network layer such as deconvolution layer.

Atrous convolution can increase receptive field and maintain the number of kernel parameters at the same time, so as to achieve the purpose of effectively maintaining image resolution.

In the basic network structure of this paper, the encoder-decoding structure and atrous convolution are properly combined. In the encoding process, Fusion Atrous Spatial Pyramid Pooling (FASPP) is proposed, and Hierarchical Feature Fusion (HFF) is proposed in the decoding process to generate semantic feature structure (GSM). Based on the inherent characteristics of thyroid nodules ultrasound images, a targeted network structure Pronet is proposed. The final network structure is shown in Fig. 1.

III. ALGORITHM IMPLEMENTATION

A. ENCODER

1) DENSENET-121 NETWORK STRUCTURE MODEL

Huang Gao *et al.* proposed Dense Net for the first time in 2017 [25]. The structure principle diagram of Dense Block is shown in Fig. 2.

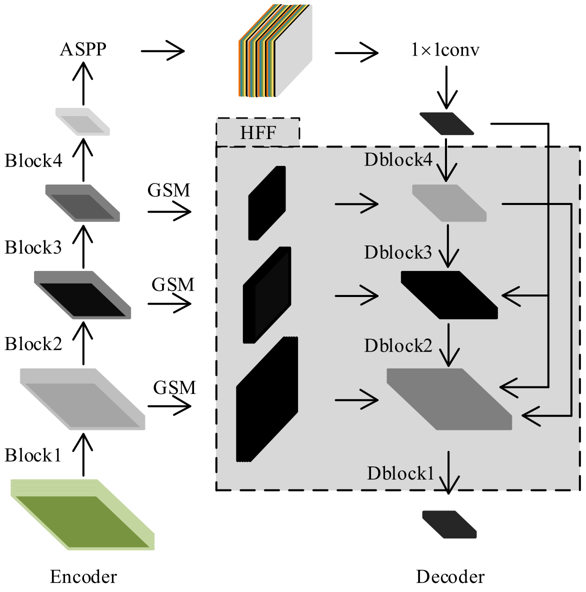


FIGURE 1. The basic structure of the proposed network.

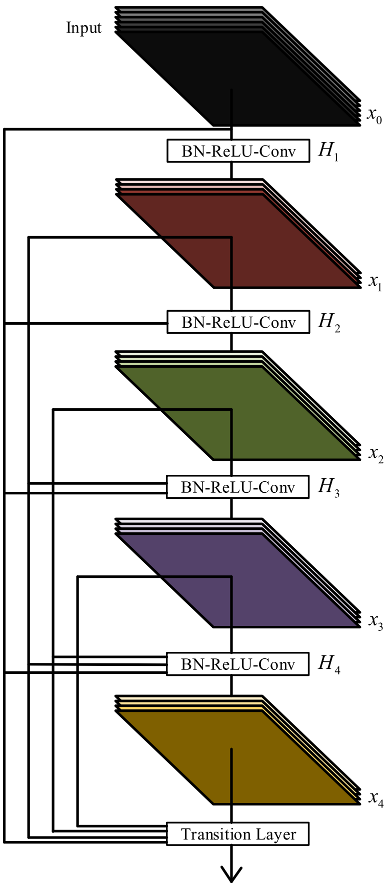


FIGURE 2. Schematic diagram of Dense Block structure.

Dense Net uses a similar idea to Res Net when dealing with problems such as network degradation and gradient disappearance, and uses short connections to deal with such problems. The difference between them when dealing with problems is that the core design of Dense Net uses Dense Block structure. The source of the name Dense Net is because this structure resembles a dense network. Its characteristic is that it can connect any two convolutional layers.

The Dense Block structure diagram in Fig. 2 contains 5 layers of structure, which form a densely connected block as a whole. It can be seen from Fig. 2 that any two convolutional layers are interconnected and connected. And the feature layer of the upper layer is the input value of lower layer. The structure diagram of densely connected blocks given in Fig. 2 can reduce the number of parameters of entire network to a certain extent. This makes the network narrower and can achieve the purpose of making full use of the characteristics of each layer. The connection between adjacent layers of Dense Block is to merge channels by Concatenation instead of simply adding them. This is quite different from Res Net network, which is also the essential difference between the two.

The basic settings in any Dense Block structure include growth rate parameters. The growth rate in a certain Dense Block structure represents the number of feature layers output by each layer in Dense Block. At the connection point between layers in Dense Block structure, a layer of bottleneck layer can be added to reduce the number of parameters in network and reduce its feature dimension. The new structure Dense Net formed after adding Bottleneck and Translation layer to Dense Block structure is named Dense Net-BC. The most common structure in Dense Net network is Dense Net-121 structure. The parameters and composition of the network structure are shown in Tab. 1.

2) ATROUS SPATIAL PYRAMID POOLING

In order to better obtain the contextual multi-scale information of input feature map, multiple convolutions with different expansion coefficients can be used in this process to achieve the purpose of obtaining multi-scale feature maps. But this will also bring some negative effects, such as changing the size of mapping features. In order to solve the above-mentioned problem that the size of mapping features changes, the mapping features can be formed by splicing by using Spatial Pyramid Pooling (SPP). In addition, depthwise separable convolution is generally used when processing the channels of input images. In summary, a new ASPP can be formed by combining spatial pyramid pooling and depthwise separable convolution to separate input images channel from the spatial information. In this paper, the operation used in the last layer of U-Net network coding in traditional computing is replaced with ASPP. The network structure of ASPP is shown as in Fig. 3.

In the calculation process, ASPP is used to perform convolution operation on the feature map of upper layer, which is mainly divided into the following five convolution processes:

(a) The first convolution uses 256 ordinary 1 × 1 convolution kernels to perform convolution calculation on the feature map, and add batch normalization layer operation after convolution.

TABLE 1. The parameters and composition of Dense Net-121 network structure.

Layers	Output Size	Dense Net-121(k=32)
Convolution	112×112	7×7conv, stride=2
Pooling	56×56	3×3max pool, stride=2
Dense Block (1)	56×56	$\begin{bmatrix}1\times1 & \text{conv}\\ 3\times3 & \text{conv}\end{bmatrix}\times 6$
Transition Layer (1)	56×56 28×28	1×1 conv 2×2 avg pool, stride=2
Dense Block (2)	28×28	$\begin{bmatrix}1\times1 & \text{conv}\\ 3\times3 & \text{conv}\end{bmatrix}\times 12$
Transition Layer (2)	28×28 14×14	1×1 conv 2x2 avg pool, stride=2
Dense Block (3)	14×14	$\begin{bmatrix}1\times1 & \text{conv}\\ 3\times3 & \text{conv}\end{bmatrix}\times 24$
Transition Layer (3)	14×14 7×7	1×1 conv 2x2 avg pool, stride=2
Dense Block (4)	7×7	$\begin{bmatrix}1\times1 & \text{conv}\\ 3\times3 & \text{conv}\end{bmatrix}\times 16$
Classification Layer	1×1	7x7 global avg layer
/	/	1000D fully-connected, soft max

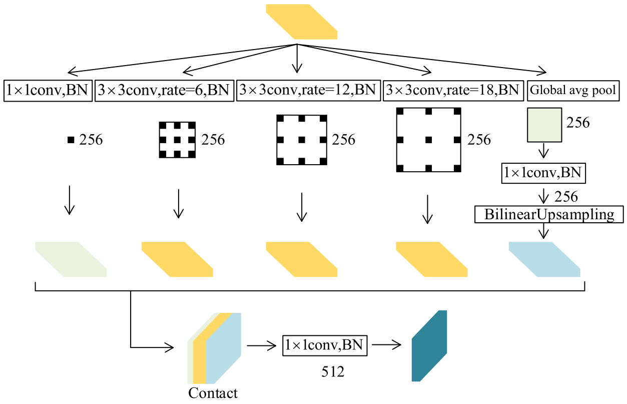


FIGURE 3. The basic network structure of ASPP.

(b) Use depthwise separable convolution calculations during the second to fourth convolution calculations. The depthwise separable convolution network structure in each convolution process can be expressed by the following formula:

$$\text{Depthconv}(3\times3)+\text{BN}+\text{Pointconv}(1\times1)+\text{BN}+\text{ReLU} \tag{1}$$

DepthConv —-3 × 3 dilated convolution with expansion coefficients of 6, 12, and 18.
PointConv —- Ordinary 1 × 1 convolution.

In the second to fourth convolution operations, using the network structure of depthwise separable convolution of equation (1) can greatly reduce the number of parameters in the model, thereby speeding up the convergence speed of model calculation.

(c) In the fifth convolution process, the size of original image needs to be reduced to 1/output step size of previous size (the output step size in this paper is 16). Then the feature map is sent to 1 × 1 convolution kernel with 256 output channels by performing the global mean pooling operation, then proceed to batch normalization layer operation. Finally, the bilinear interpolation method is used to restore the image size. Although ASPP with different sampling rates can capture multi-scale information well, but as the sampling rate gradually increases, the weight of filter will also decrease. When its weight is reduced to a certain extent, 3 × 3 convolution kernel can no longer fully capture the context information of images. The 3 × 3 convolution kernel will also degenerate into a simple 1 × 1 convolution kernel. According to the above-mentioned method, the fifth convolution can solve this problem to the greatest extent.

It should be noted that after five convolution operations are completed, the five multi-scale feature maps extracted need to be spliced. Its purpose is to be able to get the correlation between different feature maps. After the feature map is spliced, it is sent to 1 × 1 convolution kernel with 512 channels, then performs batch normalization layer operation to send the final feature map to decoding module for decoding.

3) MODEL SIZE OPTIMIZATION

The size of convolutional layer after fusion may be too wide. In order to prevent this, the size of model needs to be optimally controlled. In this paper, in order to appropriately reduce the dimension and size of feature images, a 1 × 1 convolution operation is performed before each convolution calculation. In this way, the dimensionality of feature maps can be reduced to half of original so as to achieve the purpose of reducing output size.

Assume that each convolutional layer initially has a0 input features, and each convolutional layer outputs m feature maps. Then the number of input feature maps ak of 1 × 1 convolution of k atrous convolution layer is:

$$a_k = a_0 + m \times (k-1) \tag{2}$$

Before each convolution calculation, a 1×1 convolution operation is performed, which reduces the number of channels to $a_0/2$. Set the output number of feature images of each convolutional layer to $m = a_0/8$. Assuming that there are Sum parameters in the network, then:

$$\begin{aligned} Sum &= \sum_{k}^{C}\left[m_k \times 1 \times 1 \times \frac{a_0}{2} \times m \times J^2\right] \\ &= \sum_{k}^{C}\left[\frac{a_0}{2}\cdot\left(a_0+(k-1)\times\frac{a_0}{8}\right)+\frac{a_0}{2}\times J^2\times\frac{a_0}{8}\right] \\ &= \frac{a_0 C}{32}\left(15+C+2J^2\right) \end{aligned} \tag{3}$$

C —-The number of layers of dilated convolution.
J —-The size of convolution kernel.

Each convolutional layer in ASPP outputs 64 feature maps (m = 64). If 1 × 1 convolution operation is not performed before the convolution calculation, the number of channels is 256. The calculation shows that the number of parameters at this time is about 107. Performing a 1 × 1 convolution operation before the convolution calculation can reduce the number of channels to 128. The number of final parameters calculated is about 1.556 × 106. Without loss of accuracy, the size of the model can be optimally controlled by reducing the amount of parameters, and the size of the model can be appropriately reduced while reducing the time required for network convergence.

B. DECODER

1) HIERARCHICAL FEATURE FUSION

The output step size of the output feature of thyroid nodule ultrasound image after passing by the decoder is 16. In the Deeplab v3 structure, bilinear interpolation is performed on the obtained feature images. The coefficient of bilinear interpolation is the same as the output step length of output features, which is 16. This structure is equivalent to a simple decoding module, and its feature is that the resolution of images can be restored to the same as original images. However, it has certain defects, such as losing part of the characteristic information, which will cause the ultrasound image of thyroid nodules to be unable to be completely segmented.

As can be seen from the basic structure of network in Fig. 1, the coding stage in the basic network structure of this paper is applied to ResNet-101 structure. It mainly includes four modules, block1, block2, block3 and block4. The encoder can continuously extract the characteristic information of each layer from block1 to block4. In the decoding stage of basic network structure, this paper proposes a Hierarchical Feature Fusion (HFF) structure. The decoder also includes four modules: dblock4, dblock3, dblock2 and dblock1. The decoder can provide feature images with multiple hierarchical levels from dblock1 to dblock4.

Each of the four modules in the decoder can be divided into the following two stages: encoder adaptation stage and image feature generation stage. Among the four modules, dblock1, dblock2 and dblock3 are integrated with each other, and merge the output of previous decoder with the output of encoder. The Dblock4 module is different from the previous three modules in that it only has inputs and does not perform fusion operations. In the process of fusion between dblock1, dblock2 and dblock3 modules, the output feature information of previous decoder needs to be fused with the matching feature obtained by the encoder. In the decoding stage, this paper uses feature maps obtained in the encoder stage as the basis in the decoder module to appropriately adjust feature maps with a relatively low resolution from the decoder module. The purpose of adjustment is to make the fusion features obtained by different convolutional layers have the same dimensions. That is, the spatial resolution and the number of channels of fused features obtained by different convolutional layers are the same.

The final operation of the encoder and previous decoder should be 3 × 3 convolution, which is done to ensure the same number of channels. Set the number of output channels that will be two convolutions. Their values are the same and they are both set to the minimum of the number of input channels of convolution. In addition, bilinear interpolation is used to sample low-resolution feature images and the maximum spatial resolution of features to be fused. After the adjustment, the dimensions of each feature information are the same, and the number of output channels of these two convolutions can be set to the same value, that is, the minimum value of the number of input channels of convolution, by fusing the features by element-wise summation. Then, in order to ensure that the features have the same spatial resolution, the bilinear interpolation method is used to up-sample low-resolution feature map and the maximum spatial resolution of features to be fused. Through adjustment, all features have the same dimension, and features can be fused by element-wise summation. The process is shown in Fig. 4.

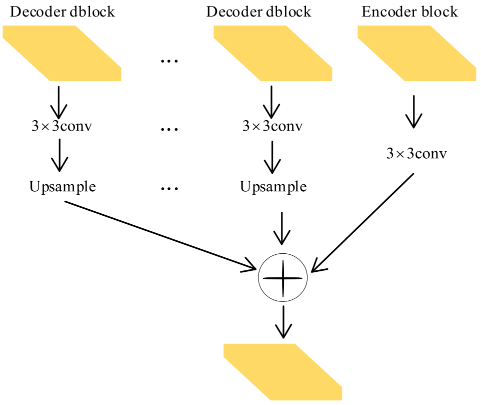


FIGURE 4. Hierarchical feature fusion.

2) GENERATE SEMANTIC FEATURES

Semantic features are generated based on contextual information, and the last part of each module in the decoder is responsible for capturing contextual information. In the process of capturing contextual information, four convolution pooling operations with 3 × 3 convolution are applied. Then there are the maximum pooling operation of 5×5 convolution and 3 × 3 convolution operation. Different fusion modules obtain the context information of image areas from different spatial positions of feature maps when capturing the context information, and merge the input at this stage with all the outputs of the set operation by connecting mapping features.

In order to appropriately reduce the dimension of feature maps from the fusion layer and the feature dimension of cascade structure, a 3 × 3 convolution operation is applied. The structure of semantic feature structure is shown in Fig. 5.

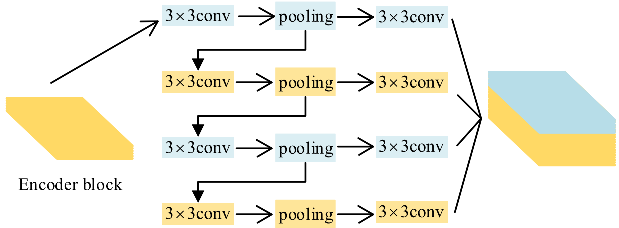


FIGURE 5. Generate semantic feature structure.

The four modules are stacked, and the final predicted segmentation result is implemented in dblock1 module. In the process of generating semantic features, firstly it reduces fitting by dropout operation. Then the number of output channels of feature maps is adjusted to be consistent with the number of output pixel classes through a 3 × 3 convolution operation. Then, based on softmax function, a semantic segmentation map of the thyroid is generated on all pixels. Finally, the low-resolution feature map is appropriately adjusted based on bilinear interpolation, and adjusted to the size of original images.

C. LOSS FUNCTION

The thyroid ultrasound image segmentation model proposed in this paper is based on the thyroid segmentation network of encoding and decoding. In the segmentation process, the model needs to be continuously trained to predict whether each pixel is a background. This problem is a pixel-level two classification problem.

The loss function is the cost function. It is usually used to measure the difference between the predicted results of model and true results. Its function is to judge the pros and cons of model. The smaller the value of loss function, the better the fitting ability of model, the richer the features learned by model, and the better the overall performance of model. The loss function usually involved in classification problems is Binary Cross Entropy (BCE) loss function, which can be expressed by the following formula:

$$BCE = -\sum_{j=1}^{N} a_{pre} \log a_{tru} + \left(1 - a_{pre}\right) \log\left(1 - a_{pre}\right) \tag{4}$$

a_{pre} —- The prediction value, its value is 0 or 1.
a_{tru} —- The true value, its value is 0 or 1.

When the predicted value is equal to the true value, loss value loss is 0. When the predicted value is not equal to the true value, loss is greater than 0. The more the probability is different, the greater the loss value.

In actual operation, the loss function is generally not used directly. In this paper, BCE loss function is replaced by Tversky Loss (TL) loss function [26] in the calculation. The TL loss function allows the flexibility to balance false negatives and false positives, and it can be expressed as follows:

TABLE 2. Platform parameters of the segmentation test.

Parameter name	Parameter value
CPU model	Intel(R)Xeon(R) CPU E5-2690 v4
CPU frequency	2.6 GHz
CPU core	28 cores
Memory size	188G
Video card model	NVIDIA Corporation GK210GL [Tesla K80]
Video memory	11G

$$TL = \frac{\sum_{j=1}^{N} a_{pre} \cdot a_{tru}}{\sum_{j-1}^{N} \left(a_{pre} \cdot a_{tru}\right) + \gamma_1 \sum_{j-1}^{N} \left(a'_{pre} \cdot a_{tru}\right) + \gamma_2 \sum_{j-1}^{N} \left(a_{pre} \cdot a'_{tru}\right)} \tag{5}$$

a_{pre} —- The predicted value of probability that pixel j belongs to the disease class.
a'_{pre} —-The predicted value of probability that pixel j belongs to the non-pathological class.
a_{tru} —-The true value of probability that pixel j belongs to the disease class.
a'_{tru} —-The true value of probability that pixel j belongs to the non-pathological class.
γ_1 —- Parameter, used to control the proportion of false negatives.
γ_2 —- Parameter, used to control the proportion of false positives.

Adjusting γ_1 and γ_2 can redistribute weights, which can improve sample imbalance and improve recall.

IV. SIMULATION EXPERIMENT

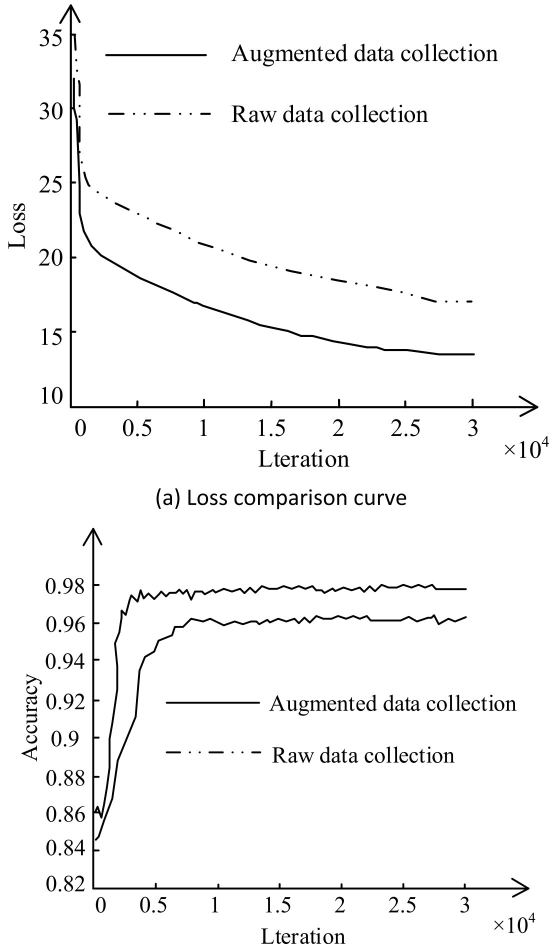
A. EXPERIMENTAL SCHEME SETTINGS

The parameters of experimental platform used in the simulation experiment are shown in Tab. 2.

Ultrasound images of thyroid nodules were taken from 30 patients in the hospital. Each patient can provide 20-40 usable ultrasound images as samples, and the final sample is 1,000. The size of ultrasound images for thyroid nodules is 548 × 456, and the pixel size is 0.35 mm. Due to the small number of effective samples in thyroid nodules dataset, the network training is over-fitting or the network generalization ability is poor. Therefore, it is necessary to optimize and select the original data first. Then we perform an augmentation operation on the filtered data to increase the number of samples to 4000, thereby improving training accuracy and overall network performance.

First, label the serial number of 1000 ultrasound image samples of thyroid nodules, then select 900 images for training, and the remaining 100 images for algorithm evaluation test. From 4000 ultrasound images of enhanced thyroid nodules, 3,600 images were taken for training, and the remaining 400 images were used for algorithm evaluation tests. Finally, taking the result of doctor's manual segmentation of images as the standard, the reliability of our algorithm is judged by comparing the image segmentation results of different algorithms.

The ultrasound images of thyroid nodules were trained on the basis of original dataset and enhanced dataset. The loss value and accuracy rate obtained by the network model are shown in Fig. 6.



(a) Loss comparison curve

(b) Accuracy comparison curve

FIGURE 6. Comparison of results between data augmentation training and original data.

It can be seen from Fig. 6(a) that the loss value comparison curves obtained on the basis of original dataset and enhanced dataset are very close in the initial training stage. However, as the number of iterations continues to increase, the loss value obtained based on enhanced dataset is much smaller than the loss value obtained based on original dataset.

It can be seen from Fig. 6(b) that as the number of iterations continues to increase, the accuracy obtained on the verification set based on enhanced dataset is much greater than the accuracy obtained on the verification set based on original dataset. Thus, the dataset after data enhancement can greatly improve the accuracy of the network model in training process and greatly reduce the loss value.

B. EVALUATION INDICATORS AND EXPERIMENTAL RESULTS

When testing the performance of the model, in order to quantitatively measure the performance of proposed model, this paper will use the following three standard indicators to evaluate the model performance: Dice Similarity Coefficient (DSC), Prevent Match (PM) and Correspondence Ratio (CR).

DSC is generally used to consider the similarity between labels and the predicted value, and its value range is (0, 1). The larger DSC value, the more similar labels and the predicted value. PM is generally used to consider the situation that ultrasound images are missed during segmentation, and its value range is (0, 1). The larger the value of PM, the less the ultrasound image is missed during segmentation. CR is generally used to consider the situation that ultrasound images are incorrectly segmented during segmentation, and its value range is (0, 1). The larger the value of CR is, the less the ultrasound image is mistakenly segmented during segmentation.

DCS can be expressed by the following formula (6):

$$\mathrm{DCS} = 2 \times \frac{\left|Z \cap \hat{Z}\right|}{|Z| + \left|\hat{Z}\right|} \tag{6}$$

PM can be expressed by the following formula (7):

$$\mathrm{PM} = \frac{Z_{tru}}{Z} \times 100\% \tag{7}$$

CR can be expressed by the following formula (8):

$$\mathrm{CR} = \frac{Z_{tru} - \frac{1}{2} Z_{fau}}{Z} \times 100\% \tag{8}$$

Z —- The measured area manually divided by doctors.
$\hat{Z}$ —- The measured area segmented by the model proposed in this paper.
Z_{tru} —- The area that is segmented correctly.
Z_{fau} —- The area that is segmented wrongly.

In this paper, original dataset and enhanced dataset are used as the basis for training to obtain the corresponding model, and then the corresponding test set is processed with the obtained model. The segmentation results obtained on the test set were evaluated with DSC, PM and CR as the standards. The evaluation results based on the contour distance and evaluation results based on the contour area are shown in Tab. 3 and Tab. 4 respectively.

It can be seen from Tab. 3 and Tab. 4 that DSC based on original dataset in the evaluation results based on contour distance is 0.9551, PM is 0.9261, and CR is 0.9271. Based on the enhanced dataset, DSC obtained is 0.9724, PM is 0.9787, and CR is 0.9621. In the evaluation result based on contour area, DSC based on original dataset is 0.9481, PM

TABLE 3. Evaluation results based on contour distance.

Method	DSC	PM	CR
Original dataset	0.9551	0.9261	0.9271
Enhanced dataset	0.9724	0.9787	0.9621

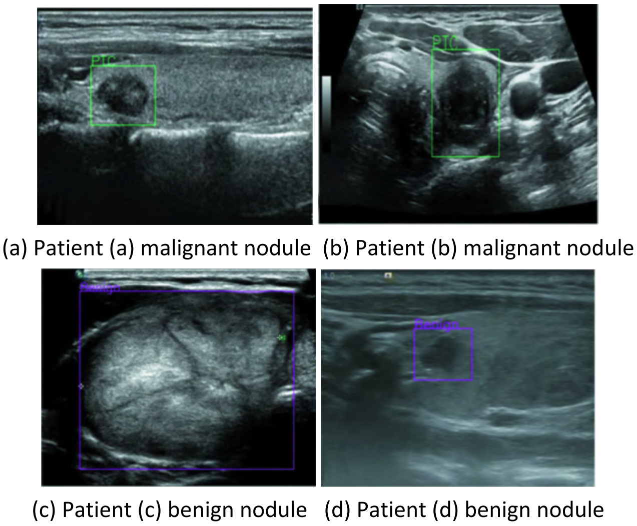
TABLE 4. Evaluation results based on contour area.

Method	DSC	PM	SN
Original dataset	0.9481	0.9621	0.9281
Enhanced dataset	0.9612	0.9879	0.9558

is 0.9621, and SN is 0.9281. Based on the enhanced dataset, DSC obtained is 0.9612, PM is 0.9879, and SN is 0.9558.

It can be seen from the data in the table that whether it is in evaluation results based on contour distance or the evaluation results based on contour area, the model trained on the basis of enhanced dataset is better than the training based on original dataset on the whole resulting model. Therefore, it can be verified that data growth can effectively improve the generalization ability of the network model and the accuracy of tests.

The figure shows the ultrasound images of thyroid nodules of different nature (malignant or benign) in 4 different patients in samples.



(a) Patient (a) malignant nodule (b) Patient (b) malignant nodule

(c) Patient (c) benign nodule (d) Patient (d) benign nodule

FIGURE 7. Ultrasound images of thyroid nodules in different patients.

According to the selected samples, the ultrasonic image segmentation experiment of thyroid nodules is carried out. Fig. 8 shows the segmentation results of different thyroid ultrasound image samples.

The first row of images in Fig. 8 are different thyroid ultrasound image samples, and the second row corresponds to the image drawn manually by doctors. The third row is the segmentation result of the network model proposed in this paper, and the fourth row is the difference graph of division probability. It can be seen from Fig. 7 that the nodule segmentation model of thyroid ultrasound images proposed in this paper can segment images relatively accurately.

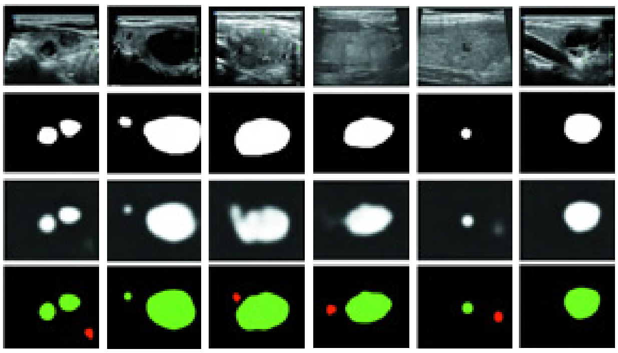


FIGURE 8. Schematic diagram of the segmentation effect of thyroid nodules.

C. OPTIMIZER SELECTION EXPERIMENT

In the process of optimization selection, SGD optimizer, RMSprop optimizer and Adam optimizer are used to test the segmentation data of the ultrasound image of thyroid nodules after coarse positioning. We select the appropriate optimizer by comparing and analyzing different test results.

The variation of the intersection ratio of test sets under different optimizers with the number of iterations is shown in Fig. 9.

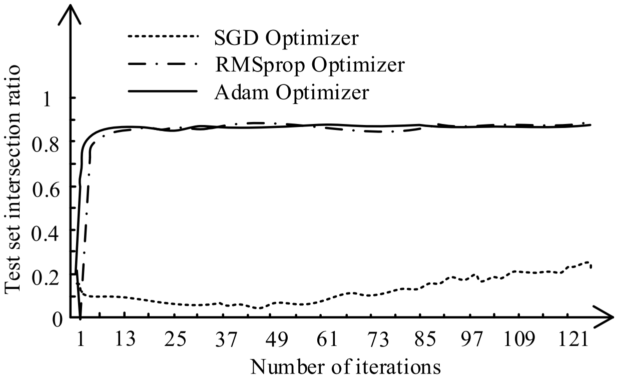


FIGURE 9. The curve of the intersection ratio of test sets under different optimizers with the number of iterations.

It can be seen from Fig. 9 that in the curve of test set intersection ratio with the number of iterations corresponding to SGD optimizer, as the number of iterations continues to increase, the test set intersection ratio first decreases and then rises. But it did not converge after 125 iterations. In the curve of test set intersection ratio with the number of iterations corresponding to RMSprop optimizer, as the number of iterations increases, the test set intersection ratio rises rapidly and tends to converge stably when the number of iterations is about 20. Its value is about 86.6%. In the curve of the intersection ratio of test set corresponding to Adam optimizer with the number of iterations, the intersection ratio of test set rises rapidly and tends to converge stably when the number of iterations is about 10. Its value is about 87.1%.

The cross entropy loss function curves under different optimizers are shown in Fig. 9.

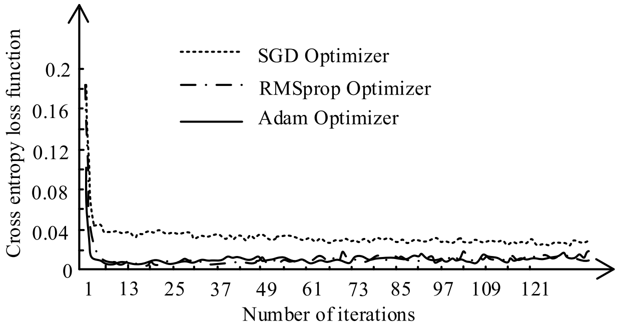


FIGURE 10. Cross entropy loss function curve under different optimizers.

It can be seen from Fig. 10 that in the cross entropy loss function curve corresponding to SGD optimizer, the loss function drops rapidly in the first iteration. As the number of iterations increases, the loss function shows a steady and slow downward trend. In the cross entropy loss function curve corresponding to RMSprop optimizer, the loss function also drops rapidly in the first iteration. But as the number of iterations increases, the loss function tends to converge smoothly at the seventh iteration. The cross entropy loss function curve corresponding to Adam optimizer is similar to RMSprop optimizer. However, as the number of iterations increases, the loss function has begun to converge smoothly by the second iteration.

In summary, the overall performance of Adam optimizer is the best. Therefore, Adam optimizer should be selected for optimization in the ultrasound image segmentation model of thyroid nodules.

D. INFLUENCE OF DIFFERENT LOSS ON EXPERIMENTAL RESULTS

Due to the large differences between individuals in the ultrasound image samples of thyroid nodules selected during experiment, there are more slices containing small targets when making slices. Eventually, it may destroy the data balance of positive and negative samples, making it difficult to continue training. This paper proves that using Tversky loss as a loss function can greatly improve the performance of segmentation by comparing the segmentation results of Tversky loss and BCE loss function in the network structure. The calculation results of different loss functions are shown in Tab. 5.

TABLE 5. Comparison of segmentation results of different loss functions.

Loss	DSC	PM	CR
BCE loss function	0.9954	0.9894	0.9872
Tversky loss	0.9961	0.9931	0.9874

It can be seen from Tab. 2 that the evaluation index of segmentation results using Tversky loss as loss function is higher than the segmentation results of BCE loss function. Thus, using Tversky loss as a loss function can effectively improve the performance of segmentation.

E. COMPARISON OF SEGMENTATION RESULTS OF DIFFERENT MODELS AND ALGORITHMS

Reference [27] proposed an active contour model, which can efficiently segment images. In addition, reference [28], reference [29], reference [30] and reference [31] also proposed image segmentation models with different performance. In this paper, evaluation calculations are made for the above-mentioned different segmentation models and algorithms, the calculation results are shown in Tab. 6.

TABLE 6. Platform parameters of the segmentation test.

Model	DSC	PM	CR
Snake [27]	0.9713	0.9665	0.9701
VGG16 [28]	0.9858	0.9912	0.972
Deeplabv3+[29]	0.9722	0.9803	0.945
U_Net [30]	0.984	0.9753	0.9711
CE-Net[31]	0.99	0.982	0.979
Model in this paper	0.9961	0.9931	0.9874

It can be seen from the above results that DSC, PM and CR calculated by the model algorithm proposed in this paper are 0.9961, 0.9931 and 0.9874 respectively. And the three standards are better than the results of other algorithms. Therefore, it can be known that the segmentation model algorithm proposed in this paper has better segmentation performance and strong generalization ability, and has a certain improvement in the segmentation effect of ultrasound images for thyroid nodules.

V. CONCLUSION

The tissue structure around the thyroid is complex, the resolution of thyroid ultrasound images is low, and image segmentation is difficult due to external interference. Thus, it is difficult to segment and extract the ultrasound images of thyroid nodules accurately. Aiming at these problems, this paper proposes an ultrasound image segmentation model algorithm for thyroid nodules based on ASPP fusion features. Fusion atrous convolution pyramid structure is proposed in the encoding process by properly combining the encoder-decoding structure and atrous convolution. Furthermore, the possibility of fused convolutional layer size being too wide is eliminated by optimizing the control of model size. In the decoding process, hierarchical feature fusion is proposed and semantic features are generated. The feature images with low resolution and the maximum spatial resolution of features to be fused are sampled by bilinear interpolation, and the fused features are calculated by element-wise summation. According to the basic structure of proposed network, DSC, PM and CR are used as evaluation criteria to compare and analyze with other methods.

The experimental results show that the ultrasound image segmentation effect of thyroid nodules is greatly improved compared with traditional segmentation method, and the effectiveness of the model algorithm proposed in this paper is verified. We will further improve the model based on this work in the future. The ultrasound image segmentation effect is further improved, and the detection of corresponding thyroid nodules on this basis is carry out considering increasing the depth of convolutional layer.

REFERENCES

[1] D. Wu *et al.*, "Medical image segmentation based on visual saliency of feature fusion," *Chin. J. Med. Phys.*, vol. 35, no. 6, pp. 670–675, 2018.

[2] W. Haiou *et al.*, "Design of superpiexl U-Net network for medical image segmentation," *J. Comput.-Aided Des. Comput. Graph.*, vol. 31, no. 6, pp. 1007–1017, 2019.

[3] C. Ma, Y. S. Liu, G. N. Luo, and K. Q. Wang, "Combining concatenated random forests and active contour for the 3D MR images segmentation," *Acta Automatica Sinica*, vol. 45, no. 5, pp. 1004–1014, 2019.

[4] L. Miao *et al.*, "Automatic segmentation of liver tumor in CT volume using, nonlinear enhancement and graph cuts," *J. Comput.-Aided Des. Comput. Graph.*, vol. 31, no. 6, pp. 1030–1038, 2019.

[5] M. Shao, J. Yan, X. Cui, and Z. Yu, "Ultrasound image segmentation of thyroid nodule based on CV-RSF algorithm," *J. Biomed. Eng. Res.*, vol. 38, no. 3, pp. 336–340, 2019.

[6] D. Ran, J. Yan, Y. Cui, and Z. Yu, "Ultrasound image segmentation of thyroid gland based on bilateral filters-distance regularized level set evolution algorithm," *J. Biomed. Eng. Res.*, vol. 38, no. 2, pp. 170–175, 2019.

[7] S. A. Paschou, A. Vryonidou, and D. G. Goulis, "Thyroid nodules: A guide to assessment, treatment and follow-up," *Maturitas*, vol. 96, pp. 1–9, Feb. 2017.

[8] D. S. Dean and H. Gharib, "Epidemiology of thyroid nodules," *Best Pract. Res. Clin. Endocrinol. Metabolism*, vol. 22, no. 6, pp. 901–911, 2008.

[9] S. Ghose, A. Oliver, R. Martí, X. Lladó, J. C. Vilanova, J. Freixenet, J. Mitra, D. Sidibé, and F. Meriaudeau, "A survey of prostate segmentation methodologies in ultrasound, magnetic resonance and computed tomography images," *Comput. Methods Programs Biomed.*, vol. 108, no. 1, pp. 262–287, Oct. 2012.

[10] R. Dong-Mei, Y. Jia-Yong, C. Xiao-Yao, and Y. Zhen-Kun, "Automatic segmentation of thyroid 3D ultrasound images based on an improved DRLSE model," *Comput. Eng. Softw.*, vol. 40, no. 4, pp. 61–66, 2019.

[11] Z. Chunyu and C. Xianyi, "Automatic segmentation of thyroid ultrasound images based on adaptive filtering combined with T-snake model," *Appl. Res. Comput.*, vol. 37, no. 3, pp. 944–946, 2020.

[12] B. Zheng, F. Xu, J. Guo, and L. Liu, "Thyroid nodule ultrasound image segmentation method based on Chan-Vese model," *Comput. Telecommun.*, vol. 6, pp. 4–7, 2018.

[13] W. Xin and L. Liang, "Ultrasound image segmentation algorithm for thyroid nodules," *Appl. Electron. Techn.*, vol. 43, no. 3, pp. 192–196 and 200, 2017.

[14] L. Yong-Luo, W. Wen-Qiang, and M. Li-Wu, "Segmentation of ultrasonic phased array NDT image based on hybrid active contour model," *Softw. Guide*, vol. 18, no. 1, pp. 741–748, 2019.

[15] C. K. Kwoh, M. Y. Teo, W. S. Ng, S. N. Tan, and L. M. Jones, "Outlining the prostate boundary using the harmonics method," *Med. Biol. Eng. Comput.*, vol. 36, no. 6, pp. 768–771, Nov. 1998.

[16] R. Aarnink, S. D. Pathak, J. J. M. C. H. de la Rosette, F. M. J. Debruyne, Y. Kim, and H. Wijkstra, "Edge detection in prostatic ultrasound images using integrated edge maps," *Ultrasonics*, vol. 36, nos. 1–5, pp. 635–642, Feb. 1998.

[17] Y. Yu, Y. Chen, and B. Chiu, "Ultrasound images based propagation, computers fully automatic prostate segmentation on radial bas-relief initialization from transrectal and slice-based in biology and medicine," vol. 74, pp. 74–79, 2016.

[18] W. Gómez, L. Leija, A. V. Alvarenga, A. F. C. Infantosi, and W. C. A. Pereira, "Computerized lesion segmentation of breast ultrasound based on marker-controlled watershed transformation," *Med. Phys.*, vol. 37, no. 1, pp. 82–95, Dec. 2009.

[19] P. Pan, Y. Wang, Y. Luo, and J. Zhou, "Automatic segmentation of nasopharyngeal neoplasm in MR image based on U-Net model," *J. Comput. Appl.*, vol. 39, no. 4, pp. 1183–1188, 2019.

[20] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[21] M. Yang, K. Yu, C. Zhang, Z. Li, and K. Yang, "DenseASPP for semantic segmentation in street scenes," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 3684–3692.

[22] N. Kumar, R. Verma, S. Sharma, S. Bhargava, A. Vahadane, and A. Sethi, "A dataset and a technique for generalized nuclear segmentation for computational pathology," *IEEE Trans. Med. Imag.*, vol. 36, no. 7, pp. 1550–1560, Jul. 2017.

[23] Q. Lu and S. Tuo, "Global segmentation method for multi-threshold color images under adaptive step size," *J. Jilin Univ.(Sci. Ed.)*, vol. 57, pp. 82–88, 2019.

[24] E. M. A. Anas, P. Mousavi, and P. Abolmaesumi, "A deep learning approach for real time prostate segmentation in freehand ultrasound guided biopsy," *Med. Image Anal.*, vol. 48, pp. 107–116, Aug. 2018.

[25] Z. Liangjun *et al.*, "Research oncolor fundus image blood vessel segmentation based on Dense U-net," *J. Biomed. Eng. Res.*, vol. 38, no. 4, pp. 397–402, 2019.

[26] N. Abraham and N. M. Khan, "A novel focal Tversky loss function with improved attention U-Net for lesion segmentation," in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2019, pp. 683–687.

[27] E. E. Kumar *et al.*, "Automatic lung segmentation with Juxta-pleural nodule identification using active contour model and Bayesian approach," *Alexandria Eng. J.*, vol. 55, no. 3, pp. 2583–2588, 2016.

[28] L. Geng, S. Zhang, J. Tong, and Z. Xiao, "Lung segmentation method with dilated convolution based on VGG-16 network," *Comput. Assist. Surg.*, vol. 24, no. 2, pp. 27–33, Oct. 2019.

[29] L. Wenya *et al.*, "Urban green space extraction from GF-2 remote sensing image based on DeepLabv3+ semantic segmentation model," *Remote Sens. Land Resour.*, vol. 32, no. 2, pp. 120–129, 2020.

[30] R. Bensch *et al.*, "U-Net: Deep learning for cell counting, detection, and morphometry," *Nature Methods*, vol. 16, no. 1, pp. 67–72, 2019.

[31] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu, "CE-net: Context encoder network for 2D medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 38, no. 10, pp. 2281–2292, Oct. 2019.