Research Article

Research on Lung Tumor Cell Segmentation Method Based on Improved UNet Algorithm

Jun Sun [ID], Weimin Chen, Longzhen Zhang, and Xundong Yan

Yangzhou Hongquan Hospital, Jiangsu, Yangzhou 225000, China

Correspondence should be addressed to Jun Sun; jdsunjun@163.com

In order to improve the completeness of the computer-aided diagnosis system for the segmentation of large-sized lung tumors and the segmentation accuracy of small-sized lung tumors, a dual-attention 3D-UNet lung tumor segmentation network model was constructed. The upsampling operation in the traditional 3D-UNet network is replaced with the DUpsampling structure. By minimizing the loss between the pixels of the feature map and the compressed label image, a more expressive feature map is obtained, thereby improving the network convergence speed. On this basis, the spatial attention module and the channel attention module are integrated so that similar features in single channel and multichannel are related to each other, and the global correlation of feature maps is increased to improve the accuracy of segmentation results. The experimental results show that compared with methods such as 3D-UNet, the model effectively improves the accuracy of lung tumor cell segmentation, and the MIoU score on the public dataset LIDC-IDRI reaches 89.4%. The segmentation method will be closer to the facts and in line with the safety and health of human life.

1. Overview

As an important basis for the early diagnosis of lung cancer, the accurate segmentation of lung tumors is particularly important. With the exponential growth of computed tomography data, radiologists are faced with increasingly onerous tasks of reviewing CT images [1], even if physicians can quickly and accurately mark the location and boundaries of lung tumor cells, However, in the long-term high-intensity work process, misdiagnosis and missed diagnosis will inevitably occur. Therefore, there is an urgent need for an emerging technology to assist doctors in diagnosis, and the emergence of computer-aided diagnosis brings hope to imaging. The maturation of auxiliary diagnostic technology can not only reduce the workload of doctors but also improve the accuracy and efficiency of labeling lung tumor cells. However, due to the obvious differences in the size, shape, and other clinical characteristics of lung tumor cells [2] in lung CT images, some current segmentation [3] methods have low detection rates of lung tumor cells and are time-consuming. Therefore, it is difficult to construct an efficient lung tumor segmentation model.

In this paper, a deep neural network-based lung tumor segmentation model is constructed to improve the detection rate of lung tumors and reduce the detection time of lung tumors. At the same time, the dual-attention module is integrated into the deep neural network to optimize small-sized lung tumors, thereby improving multitype lung tumors as well as improving the segmentation accuracy of lung tumor cells.

2. Related Work

In recent years, the widespread use of deep learning [4] has led researchers to use neural networks to extract deep features of lung tumor cells for automatic diagnosis of lung tumor cells, instead of traditional lung tumor cell segmentation methods using hand-crafted features and descriptors. Reference [5] adjusted the contrast to enhance the part of lung tumor cells in CT images, then preprocessed the image by setting threshold and morphological parameters

based on experience, and finally used a simple region growing algorithm to segment lung tumor cells. Reference [6] first uses 2D deep neural network to roughly segment lung CT images and then uses a Markov model to optimize the rough segmentation results to obtain accurate segmentation results. Reference [7] proposed a multiview 2.5D convolutional neural network for the segmentation of lung tumor cells. The network consists of three CNN branches, respectively, from a set of axial, coronal, and sagittal views of lung tumor cells. Sensitivity features are captured in, and each branch consists of 7 stacked layers and takes multiscale lung tumor cell plaques as input. Three CNN branches are connected with a fully connected layer to predict whether the plaque center voxels belong to lung tumor cells. Reference [8] used the size of lung tumor cells as the main diagnostic criterion and used Mask R-CNN to segment lung tumor cells to obtain contour information. Reference [9] proposed to use the FCM algorithm as the basis, applied a wavelet transform to decompose the CT image, then used the decomposed low-frequency image pixels as the basic points of the FCM algorithm, and finally used the Mahalanobis distance to further correct the segmentation results. However, the above methods all have the following problems:

(1) Lung tumor cells have complex shapes and highly variable textures, and 2D low-level descriptors cannot capture discriminative features. Features extracted using only 2D convolutional neural networks cannot be mapped into high-quality segmentation feature maps, thus affecting the efficiency of network training. CT images are essentially three-dimensional data. Therefore, linking spatial context information plays an important role in lung tumor cell segmentation. References [6, 7] use 2D and 2.5D neural networks to segment lung tumor cells, respectively, but a single 2D lung CT image does not have the ability to effectively distinguish between small lung tumor cells and blood vessel profiles and neither makes full use of lung tumors. The spatial characteristics of cells lead to lower segmentation accuracy. Reference [8] used the size of lung tumor cells as the main feature to segment lung tumor cells, but ignored the texture features and shape features of lung tumor cell variability, and thus could not completely segment singular lung tumor cells.

(2) In the segmentation problem of relatively small objects, establishing the correlation between local features and global features helps to improve the feature representation, thereby improving the accuracy of segmentation. Although the probabilistic graphical model is used in the backend of the segmentation network to improve the segmentation accuracy, the literature [6] can only calculate the posterior probability more accurately to optimize the first-stage segmentation when the probabilistic graphical model obtains a better prior probability function. As a result, the method cannot adaptively segment the lung tumor cells according to the spatial features. References [5, 9] used the traditional region growing algorithm and the FCM algorithm as the main frame of the segmentation method, respectively, but they did not fully consider the correlation and dependence between the local and global features of lung tumor cells so that the irregular shape features of lung tumor cells cause undersegmentation.

In response to the above problems, this paper proposes a lung tumor cell segmentation method based on the 3D-UNet network [10] with a dual-attention mechanism [11]. The UNet network has excellent performance in the field of medical image segmentation. In order to adapt to the segmentation of lung tumor cells, this paper extends the original 2D-UNet network to a 3D network to capture the spatial information of lung tumor cells and introduces a dual-attention mechanism to make the network focus on key feature regions to improve segmentation accuracy for small-sized lung tumor cells.

3. 3D-UNet Method

The 3D-UNet network structure proposed in this paper is shown in Figure 1. First, in the main framework of the 3D-UNet network, this paper uses the newly proposed DUpsampling [12] structure to replace the traditional upsampling method in the decoding layer path to restore the features of lung tumor cells in the encoding path and improve the quality of lung tumor cell feature maps, to speed up network convergence. Second, a dual-attention module, namely the spatial attention module [13] and the channel attention module [14], is applied to the feature map of the penultimate layer of the 3D-UNet network to capture the correlation between local features and global features and dependency relationship, focus the network attention to the lesion area, and then improve the segmentation accuracy.

3.1. DUpsampling Structure. The DUpsampling structure is a new upsampling structure based on data correlation proposed in 2019. The upsampling structure is usually present in the decoding layer of the segmentation network, and its role is to restore the feature map to the size of the original image. Although the upsampling operation based on bilinear interpolation [15] and nearest neighbor interpolation [16] can capture and restore the features extracted by the convolutional layer to a certain extent, its process does not consider the difference between each predicted pixel. Correlation, such as weak data-dependent convolutional decoders [17], cannot produce relatively high-quality feature maps. In this paper, the DUpsampling structure based on data correlation is added to the features extracted by the 3D-UNet network reconstruction encoding path so that the obtained feature map has better expressive ability. In the process of upsampling, the most "correct" output is obtained by minimizing the loss between the pixels of the feature map and the compressed label image, which has a strong reconstruction ability. The structure of DUpsampling is shown in Figure 2.
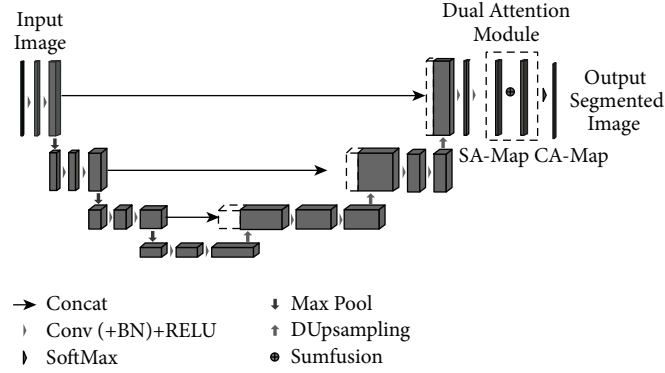
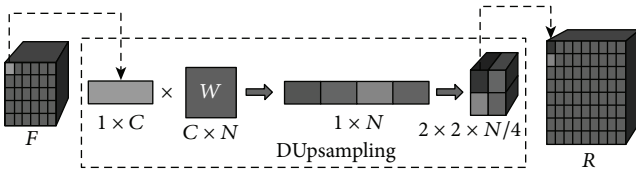FIGURE 1: The overall framework of the network.

FIGURE 2: DUpsampling structure.

In Figure 2, $F \in \mathbb{R}^{h \times w \times c}$ represents the feature map output by the CT image after encoding, h, w, and c represent the height, width, and number of channels of the feature map, respectively, and $\mathbf{R}$ represents the double upsampling of the DUpsampling structure. After the feature map obtained, $\mathbf{W}$ is a matrix that linearly compresses pixel vectors in the DUpsampling structure. Let each pixel of $x \in \mathbb{R}^{1 \times c}$ the feature map F be a vector, and then let the vector $\mathbf{x}$ and $W \in \mathbb{R}^{c \times n}$ multiply the matrix to get a vector $v \in \mathbb{R}^{1 \times n}$ and then multiply the vector $\mathbf{v}$ reorganized as $2 \times 2 \times N/4$; after rearranging, it is equivalent to 2 for each pixel of the original times the upsampling:

$$
\begin{aligned}
v &= x \times W, \\
\widetilde{x} &= v \times P.
\end{aligned}
\tag{1}
$$

Among them, the matrix $\mathbf{P}$ is the inverse transformation of the matrix $\mathbf{W}$, and $\widetilde{\mathbf{x}}$ is the vector obtained by the PCA method of dimension reduction of the manually labeled lung tumor cell segmentation area. The neural network uses the stochastic gradient descent method as the optimizer to minimize $\mathbf{x}$ on the training set. The reconstruction error between $\sim$ and $\mathbf{x}$ to find the optimal feature map reconstruction matrices $\mathbf{P}$ and $\mathbf{W}$ is shown as follows:

$$
P^+, W^+ = \underset{P,W}{\arg\min} \sum_v \|x - \widetilde{x}\|^\rho = \underset{P,W}{\arg\min} \sum_v \|x - xWP\|^2.
\tag{2}
$$

The traditional segmentation network only calculates the loss between the prediction result and the label image in the last Softmax layer [18] and then updates the weights through back-propagation to optimize the network. However, the DUpsampling structure calculates the loss between the feature map and the compressed label in advance in the upsampling part and then integrates the low-resolution feature map in the decoding layer into the high-level semantic features through the back-propagation of the network as a whole, thereby improving the features. The quality of the graph allows the dual-attention module to mine spatial information and channel information.

3.2. Dual-Attention Module. In the dual-attention module, this paper first uses the dilated convolution [19] operation with different expansion rates to capture the feature map information of different scales, fuses the resulting feature maps containing multiple scales, and uses the spatial attention module for the fusion results. The spatial attention module selectively aggregates the features of each location according to the weighted sum of all location features so that similar features are correlated with each other. Meanwhile, the channel attention module selectively emphasizes the interdependent channel feature maps by integrating the associated features among all channel maps. Finally, the outputs of the two attention modules are summed to further improve the feature representation, which in turn helps to improve the segmentation accuracy of small-sized lung tumor cells. The dual-attention module is shown in Figure 3.

3.2.1. Multiscale Feature Fusion. Extracting multiscale information of feature maps can improve the segmentation accuracy of small target objects. The usual method is to perform multiple maximum pooling [20] operations on the feature map to obtain output result maps of different resolutions and then extract features through the convolutional layer, but after multiple pooling operations, the detailed information or even all information of small target objects will be lost. Lung tumor cells account for a small proportion of lung CT images and belong to relatively small target-type segmentation. Therefore, this paper introduces atrous convolution with different expansion rates to extract feature maps. Atrous convolution can increase or decrease the receptive field by adjusting the expansion rate without shrinking the feature map and capture multiscale feature map information. Atrous convolution is defined as follows when given an input feature map $F \in \mathbb{R}^{h \times w \times c}$:
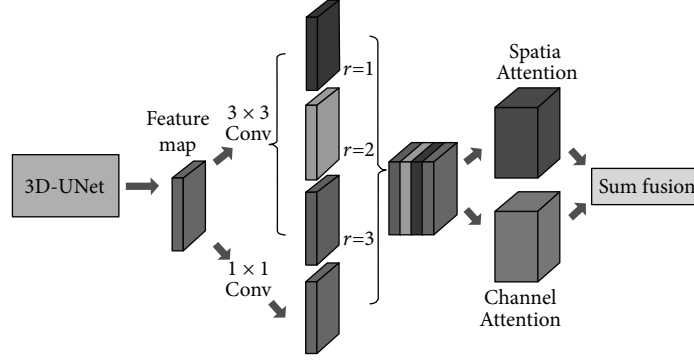
FIGURE 3: Dual-attention module.

$$y(x) = \sum_d (x + r\ d) \cdot W(x),\qquad(3)$$

where x is the position of the current pixel, W is the convolution kernel weight, r is the dilation rate, and d is the pixel value in the current convolution process. The standard form of atrous convolution is defined as Dconvr (F), where Dconvr represents the atrous convolution operation on the feature map F when the dilation rate is r. As shown in Figure 3, the feature map of the penultimate layer of the 3D-UNet network is used as an input in the dual-attention module, and then a cascade hole convolution operation is performed on the feature map, which is defined as follows:

$$O\left(C^l\right) = \left\{ Dconv_{r_1}(F); Dconv_{r_2}(F)Dconv_{r_3}(F); M(F)\right\},\quad(4)$$

where **M** represents the output feature map obtained by 1×1 convolution of the input image, where the 1×1 convolution operation is to ensure that the channels between the result maps after different dilated convolutions remain consistent to fuse different scales characteristics of lung tumor cells. After the cascade hole convolution operation, a feature map fused with multiple scale features is finally obtained, which will be used as the input of the dual-attention module.

3.2.2. Spatial Attention Module.
Location features play an important role in segmentation tasks, which are obtained by capturing contextual information between pixels. Local features generated by traditional feature extraction networks that do not consider the influence of neighboring pixels may lead to erroneous segmentations. Therefore, to build rich interpixel positional relationships on local features, a spatial attention module is introduced in this paper, as shown in Figure 4. This module enhances feature map representation by encoding a wider range of contextual information into local features and highlighting the locations of key features.

As shown in Figure 4, the input feature map **A** is the lung tumor cell feature map fused with the dilated convolution results of different expansion rates, which is first copied into 5 new feature maps. Now the feature map number of pixels are $\mathbf{A}^1$, $\mathbf{A}^2$, $\mathbf{A}^3$, and $\{A^1, A^2, A^3\} \in \mathbb{R}^{c \times h \times w}$ and reshape its dimensions to $\mathbb{R}^{c \times n}$, $n = h \times w$ be the number of pixels. Then, perform matrix multiplication with the transposed
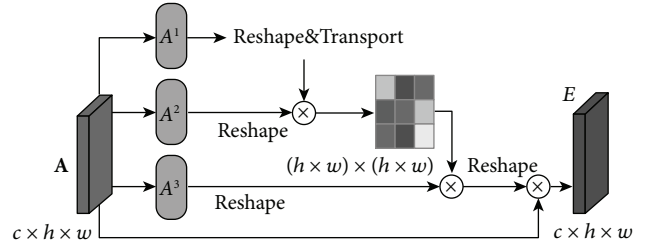

FIGURE 4: Spatial attention module.

matrix of matrix $\mathbf{A}^1$ and matrix $\mathbf{A}^2$ and then apply the Softmax layer to calculate the spatial attention map $S \in \mathbb{R}^{n \times n}$:

$$S_{ji} = \frac{\exp\left(A_i^1 \cdot A_j^2\right)}{\sum_{i=1}^{N} \exp\left(A_i^1 \cdot A_j^2\right)},\qquad(5)$$

where S_{ji} represents the influence of i pixel position to j pixel location features. The more similar the feature representations of two locations, the greater the correlation between them and vice versa. Then, reshape the reshaped matrix A3 and matrix S. The transpose of does matrix multiplication and reshapes the result is $\mathbb{R}^{c \times h \times w}$. Finally, multiply the result of the matrix operation by a scaling parameter α and with feature map **A**. Perform an element-wise sum operation to obtain the final output E as follows:

$$E_j = \alpha \sum_{i=1}^{C} \left(s_{ji} A_i^3\right) + A_j,\qquad(6)$$

where α is initialized to 0 and gradually assigns more weights during training. From the above formula, the resulting feature E_j at each position in the spatial attention map is the weighted sum of the features at all positions and the original features. Therefore, it has contextual information and selectively aggregates contexts according to spatial attention maps, highlighting key feature regions and improving segmentation accuracy.

3.2.3. Channel Attention Module.
Each feature map channel of high-level features can be viewed as the response of a specific segmentation result, and different semantic responses are correlated with each other. By mining the interdependencies between the channel graphs, the

dependencies of the feature graphs can be expressed, and the feature representation of specific semantics can be improved. Therefore, this paper constructs a channel attention module to explicitly establish the dependencies between channels, as shown in Figure 5.

Different from the spatial attention module, the channel attention module first reshapes the feature map **A** as $\mathbb{R}^{c \times n}$, then performs matrix multiplication with the transposed matrix of **A** and **A**, and finally still uses a Softmax layer to obtain the channel attention map $x \in \mathbb{R}^{c \times c}$:

$$x_{ji} = \frac{\exp\left(A_i \cdot A_j\right)}{\sum_{i=1}^{C} \exp\left(A_i \cdot A_j\right)}, \tag{7}$$

where x_{ji} measure the first i channel pair j influence of a channel. Also, take the matrix multiplication of **x** and the transpose of matrix **A** and reshape the result to $\mathbb{R}^{c \times h \times w}$, multiply the result of the matrix operation by a scale parameter β, and perform an element-wise sum operation with the feature map A, to obtain the final output $E \in \mathbb{R}^{c \times h \times w}$:

$$E_j = \beta \sum_{i=1}^{c} \left(x_{ji} A_i\right) + A_j, \tag{8}$$

where β is initialized to 0 and gradually assigns more weights during training. The final feature of each channel is the weighted sum of all channel features and the original features, thus establishing a long-term semantic dependency model between feature maps, which helps to improve the distinguishability of features and thus the completeness of segmentation results.

4. Experimental Results and Analysis

4.1. Lab Environment.
The experimental data involved in this paper come from LIDC (Lung Imaging Database Consortium), excluding slice thicknesses greater than 2.5 CT of mm. Scan the images and use the remaining 888 cases of lung images as a dataset. These 888 cases of CT images contain a total of 1 186 lung tumor cells with a diameter range of 3.170–27.442 mm. CT image acquisition parameters are 15 0 mA, 140 kV, average layer thickness is 1.3 mm, and the image resolution is 512 pixel × 512 pixel. The training data and test data are 800 and 88 cases, respectively.

During training, in DA 3D-UNet taking the preprocessed 10 consecutive CT images as a set of input data, the weights are randomly initialized using the MSRA method. In the standard back-propagation update, the learning rate is initialized as 0.1, every 1 Epoch completed Decay 5%, set batch size to 64 and momentum to 0.9. A 10-fold cross-validation strategy is used to evaluate the performance of the method, maintaining a similar data distribution in the training and testing datasets to avoid over- and undersegmentation due to data imbalance.

The environment built by the DA 3D- UNet network is Python3.7, TensorFlow Frame, CentOS 7.4, NVIDIA GeForce1080TiGPU, Processor Intel®Xeon™CPU E5-2630 v4 @ 2.20 GHz.
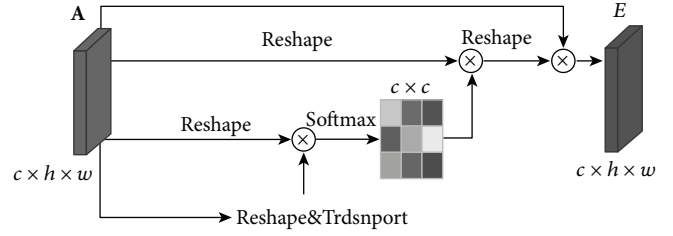

FIGURE 5: Channel attention module.

4.2. Data Preparation and Evaluation Criteria

4.2.1. Data Preprocessing.
In this paper, the mask map of the left and right lung lobes is extracted as the model input, ignoring the thoracic cavity and other noise parts. The extraction process is shown in Figure 6.

Extraction process of lung parenchyma is as follows: (1) binarize CT image and find the threshold that can distinguish lung area and nonlung area by the clustering method; (2) Kmeans clustering, distinguish lung area as one type, and nonlung area as another type; (3) corrosion operation is performed on the highlighted part of the image to remove tiny granular noise; (4) dilation operation is performed to engulf blood vessels into lung tissue and remove black noise, especially black lungs caused by opaque rays; and (5) perform the numerical model and operation of process (4) with the original image and crop it to the same size to obtain the lung parenchyma area.

4.2.2. Data Augmentation.
Each CT normalized scan was set to have a mean of −600 and a standard deviation of 300 before data augmentation. Data augmentation strategies are as follows. (1) Cropping: for each 512-pixel × 512-pixel CT image, we crop every 2 pixels into smaller slices of 500×500, so the amount of data per candidate region increases by 36 times. (2) Flip: for each CT image, flipping was performed from 3 orthogonal dimensions (coronal, sagittal, and axial positions), thus ultimately increasing the amount of data by a factor of $8 \times 36 = 288$ per CT image. (3) Repeat: to balance the number of positive and negative sample slices in the training set, the positive sample slices are replicated 8 times.

4.2.3. Evaluation Standard.
This article uses pixel accuracy (PA), mean pixel accuracy (MPA), and mean intersection over union (MIoU). Three international semantic segmentation metrics are used to evaluate the segmentation results. The calculation formulas are shown in formulas (9) to (11), respectively.

Pixel accuracy:

$$PA = \frac{\sum_{i=0}^{k} p_{ii}}{\sum_{0}^{k} \sum_{j=0}^{k} p_{ij}}. \tag{9}$$

Average pixel accuracy:

$$MPA = \frac{1}{k+1} \left(\frac{p_{ii}}{\sum_{j=0}^{k} p_{ij}} \right). \tag{10}$$
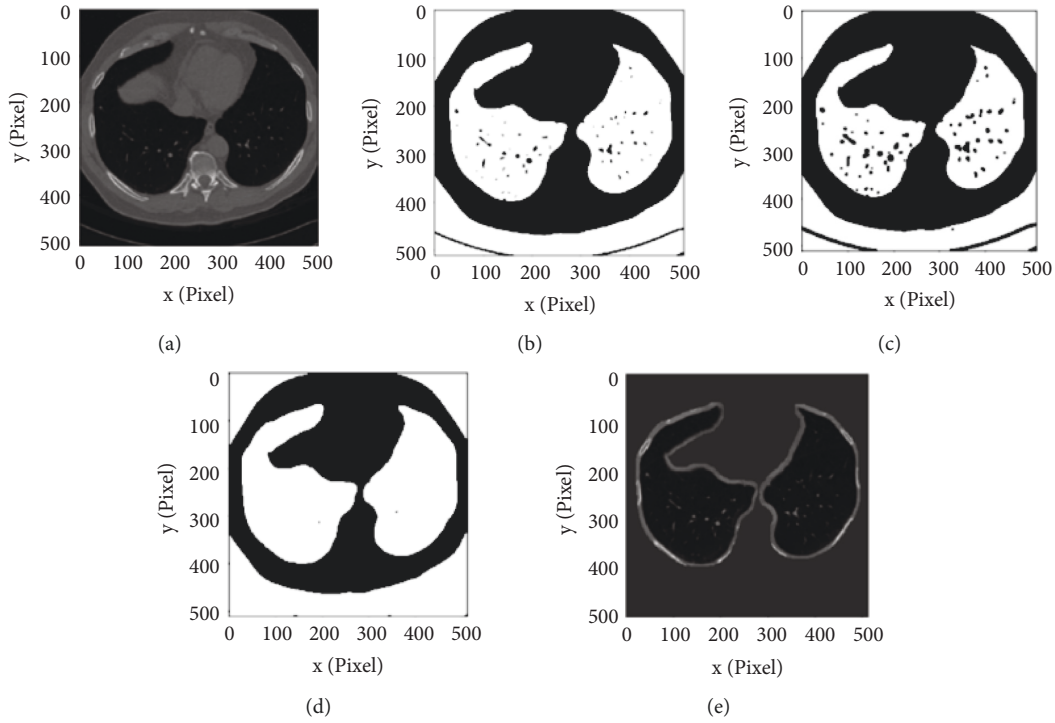
(a)

(b)

(c)

(d)

(e)

FIGURE 6: Lung parenchyma extraction process.

Average intersection ratio:

$$MIoU = \frac{1}{k+1} \sum_{i=1}^{k} \left(\frac{p_{ii}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} (p_{ji} - p_{ii})} \right). \quad (11)$$

The segmentation of lung tumor cells only needs to obtain one class of semantic segmentation results, including lung tumor cells and background, so $k = 1$ here. p_{ij} means that this belongs to i class and is predicted to be j number of pixels for the class. Similarly, p_{ii} and p_{ji}, respectively, represent that this belongs to class i and is predicted to be i number of pixels in the class and the number of pixels that belong to class j but is predicted to be i, the number of pixels for the class.

4.3. Experimental Results. Table 1 is the experimental comparison results of various experimental methods on 88 test data. Table 2 shows the experimental comparison results of 35 cases of small-sized lung tumor cells with diameters ranging from 3.170 mm to 7.5 mm extracted from 88 test datasets by various experimental methods. Table 3 shows the number of iterations and loss of the neural network, from Table 3. It can be seen that the loss value of the method in this paper has reached a relatively low level when the Bestepoch is 124 and keeps a small fluctuation. The loss values of the other methods are higher than those of the method in this paper.

Figure 7 shows the segmentation results of various types of lung tumor cells, in which the first and second columns are relatively common solitary lung tumor cells, the third and fourth columns are vascular adhesion lung tumor cells, the fifth and sixth columns are the pleural traction type lung

TABLE 1: Comparison of experimental results of different methods.

Method	PA	MPA	MIoU
Literature [1] method	0.801	0.742	0.711
Literature [2] method	0.871	0.855	0.845
Literature [3] method	0.893	0.862	0.853
Literature [4] method	0.912	0.890	0.876
Literature [5] method	0.884	0.864	0.849
3D-UNet method	0.903	0.882	0.870
The method of this paper	**0.921**	**0.907**	**0.894**

TABLE 2: Comparison of segmentation results of small-sized lung tumor cells.

Method	PA	MPA	MIoU
Literature [1] method	0.594	0.559	0.527
Literature [2] method	0.786	0.761	0.747
Literature [3] method	0.819	0.804	0.792
Literature [4] method	0.840	0.838	0.814
Literature [5] method	0.800	0.791	0.782
3D-UNet method	0.823	0.812	0.810
The method of this paper	**0.859**	**0.840**	**0.822**

TABLE 3: Optimal number of iterations and loss.

Method	Bestepoch	Loss
Literature [2] method	208	0.2504
Literature [3] method	221	0.3494
Literature [4] method	157	0.3125
3D-UNet method	186	0.2811
The method of this paper	**124**	**0.1025**

FIGURE 7: Segmentation results of various types of lung and lung tumor cells.

tumor cells, the seventh column is the rare ground glass type lung tumor cells, and the second, third, and sixth columns are all small-sized lung tumor cells with a diameter of less than 7.5 mm. The method proposed in this paper enables a complete segmentation of large-sized lung tumor cells, including columns 1, 5, and 7, and small-sized lung tumor cells, including columns 2, 3, and 6. For more accurate segmentation, the remaining comparison methods will be more or less over- and undersegmented. The experimental results show that the segmentation network proposed in this paper is superior, and the MIoU value of lung tumor cell segmentation reaches 89.4% under the LIDC standard [21] lung tumor cell dataset. In Figure 7, the 1st to 9th rows are CT image, physician annotated image, literature [5] method, literature [6] method, literature [7] method, literature [8] method, literature [9] method, 3D-UNet method, and our method.

5. Concluding Remarks

Aiming at the problems of low segmentation accuracy and long time-consuming in the current segmentation network, this paper constructs an attention mechanism 3D-UNet network structure. The DUpsampling structure is integrated into the 3D-UNet network to improve the quality of the feature map generated by the upsampling operation during the network training process so that the feature map after each upsampling is closer to the label data, and at the same time, the convergence speed of the network is accelerated. On this basis, a spatial attention module and a channel attention module are proposed to capture the global dependencies in the spatial and channel dimensions,

respectively. The experimental results show that the network structure can effectively integrate long-range context information, and improve the segmentation integrity of large-sized lung tumor cells and the segmentation accuracy of small-sized lung tumor cells. The next step will be to analyze the characteristics of various types of lung tumor cells to achieve accurate localization and tracking of all types of lung tumor cells. The future segmentation of lung cancer will move toward real-time segmentation methods.

Data Availability

The datasets used and analyzed during the current study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

[1] C. Long, H. Xu, Q. Shen, and X. Zhang, "Diagnosis of the Coronavirus disease (COVID-19): rRT-PCR or CT?" *European Journal of Radiology*, vol. 126, Article ID 108961, 2020.

[2] J. Jiang, Y.-C. Hu, C.-J. Liu, and D. Halpenny, "Multiple resolution residually connected feature streams for automatic lung tumor segmentation from CT images," *IEEE Transactions on Medical Imaging*, vol. 38, no. 1, pp. 134–144, 2019.

[3] X. Dong, Y. Lei, T. Wang et al., "Automatic multiorgan-segmentation in thoraxCTimages using U-net-GAN," *Medical Physics*, vol. 46, no. 5, pp. 2157–2168, 2019.

[4] R. Mu and X. Zeng, "A review of deep learning research," *KSII Transactions on Internet and Information Systems (TIIS)*, vol. 13, no. 4, pp. 1738–1764, 2019.

[5] S. A. El-Regaily, M. A. M. Salem, and M. H. A. Aziz, "Lung nodule segmentation and detection in computed t o m og r a ph y," in *Proceedings of the 7th International Conference on Intelligent Computing and Information Systems*, pp. 257–269, IEEE Press, Washington DC, USA, 2017.

[6] S. Mukherjee, X. Huang, and R. R. Bhagalia, "Lung nodule segmentation using deep learned prior based graphcut," in *Proceedings of the 201 IEEE International Symposium on Biomedical Imaging*, pp. 178–199, IEEE Press, Washington D. c USA, 2017.

[7] S. Wang, M. Zhou, and O. Gevaert, "A multi view deep convolutional neural networks for lung nodule segmentation," in *Proceedings of the 39th Annu al International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 2232–2245, IEEE Press, Washington D. C. , US A, 2017.

[8] M. H. Hesamian, W. Jia, X. He, and P. Kennedy, "Deep learning techniques for medical image segmentation: achievements and challenges," *Journal of Digital Imaging*, vol. 32, no. 4, pp. 582–596, 2019.

[9] L. I. Yue, "Application of wavelet transform and improved FC Malgori thminmedical CT image se gmentation," *Electronic Test*, vol. 23, no. 22, pp. 42-43, 2017.

[10] W. Wu, L. Gao, H. Duan, and G. X. S. Huang, "Segmentation of pulmonary nodules in CT images based on 3D-UNET combined with three-dimensional conditional random field optimization," *Medical Physics*, vol. 47, no. 9, pp. 4054–4063, 2020.

[11] Z. Huang, Z. Chen, J. Chen, and P Lu, "DaNet: dose-aware network embedded with dose-level estimation for low-dose CT imaging," *Physics in Medicine and Biology*, vol. 66, no. 1, Article ID 015005, 2021.

[12] S. Minaee, Y. Y. Boykov, F. Porikli, A. J. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image Segmentation Using Deep Learning: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 7, pp. 3523–3542, 2021.

[13] X. Zhu, D. Cheng, and Z. Zhang, "An empirical study of spatial attention mechanisms in deep networks," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 6688–6697, Souel Korea, 2019.

[14] F. Wang, M. Jiang, C. Qian, and S. Yang, "Residual attention network for image classification," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6450–6458, Soeul Korea, 2017.

[15] H. R. Yun, M. J. Lee, and H. Hong, "Super-resolution image generation for improvement of orbital thin bone segmentation," in *Proceedings of the International Workshop on Advanced Imaging Technology (IWAIT) 2020*, International Society for Optics and Photonics, Yogyakarta, Indonesia, 2020.

[16] P. Yang, L. Xu, Y. Wan et al., "Deep neural network-based approach to improving radiomics analysis reproducibility in liver cancer: effect on image resampling," *Physics in Medicine and Biology*, vol. 66, no. 16, Article ID 165009, 2021.

[17] Q. Tang, F. Liu, T. Zhang, and J. Y. Jiang, "Attention-guided chained context aggregation for semantic segmentation," *Image and Vision Computing*, vol. 115, Article ID 104309, 2021.

[18] S. Kanai, Y. Fujiwara, Y. Yamanaka, and S. Adachi, "Sigsoftmax: reanalysis of the softmax bottleneck," *Advances in Neural Information Processing Systems*, vol. 31, 2018.

[19] T. W. Chen, D. Wang, and W. Tao, "CASSOD-Net: cascaded and separable structures of dilated convolution for embedded vision systems and applications," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3182–3190, Soeul Korea, 2021.

[20] M. Oya, S. Sugimoto, K. Sasai, and K. Yokoyama, "Investigation of clinical target volume segmentation for whole breast irradiation using three-dimensional convolutional neural networks with gradient-weighted class activation mapping," *Radiological Physics and Technology*, vol. 14, no. 3, pp. 238–247, 2021.

[21] L. Pehrson, M. Nielsen, and C. Ammitzbøl Lauridsen, "Automatic pulmonary nodule detection applying deep learning or machine learning algorithms to the LIDC-IDRI database: a systematic review," *Diagnostics*, vol. 9, no. 1, p. 29, 2019.