Technology in Cancer Research & Treatment
Volume 21: 1-12
© The Author(s) 2022
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/15330338221090847
journals.sagepub.com/home/tct
SAGE

Gross Tumor Volume Segmentation for Stage III NSCLC Radiotherapy Using 3D ResSE-Unet

Xinhao Yu, MS[1,2], Fu Jin, PD[2], HuanLi Luo, PhD[2], Qianqian Lei, MD[2], and Yongzhong Wu, MD[2]

Abstract

Introduction: Radiotherapy is one of the most effective ways to treat lung cancer. Accurately delineating the gross target volume is a key step in the radiotherapy process. In current clinical practice, the target area is still delineated manually by radiologists, which is time-consuming and laborious. However, these problems can be better solved by deep learning-assisted automatic segmentation methods.
Methods: In this paper, a 3D CNN model named 3D ResSE-Unet is proposed for gross tumor volume segmentation for stage III NSCLC radiotherapy. This model is based on 3D Unet and combines residual connection and channel attention mechanisms. Three-dimensional convolution operation and encoding-decoding structure are used to mine three-dimensional spatial information of tumors from computed tomography data. Inspired by ResNet and SE-Net, residual connection and channel attention mechanisms are used to improve segmentation performance. A total of 214 patients with stage III NSCLC were collected selectively and 148 cases were randomly selected as the training set, 30 cases as the validation set, and 36 cases as the testing set. The segmentation performance of models was evaluated by the testing set. In addition, the segmentation results of different depths of 3D Unet were analyzed. And the performance of 3D ResSE-Unet was compared with 3D Unet, 3D Res-Unet, and 3D SE-Unet.
Results: Compared with other depths, 3D Unet with four downsampling depths is more suitable for our work. Compared with 3D Unet, 3D Res-Unet, and 3D SE-Unet, 3D ResSE-Unet can obtain superior results. Its dice similarity coefficient, 95th-percentile of Hausdorff distance, and average surface distance can reach 0.7367, 21.39mm, 4.962mm, respectively. And the average time cost of 3D ResSE-Unet to segment a patient is only about 10s.
Conclusion: The method proposed in this study provides a new tool for GTV auto-segmentation and may be useful for lung cancer radiotherapy.

Keywords
lung cancer, CNNs, auto-segmentation, GTV, radiotherapy

Received: September 17, 2021; Revised: February 8, 2022; Accepted: March 3, 2022.

Introduction

Lung carcinoma(LC) is one of the most severe and widespread cancers in the world.[1] And statistics from the World Health Organization (WHO) in 2020 showed that there were 815,563 new cases of LC and 714,699 deaths in China. Currently, in addition to surgery and chemotherapy, radiotherapy(RT) is the most effective treatment for LC. And compared with other stages, patients with stage III non-small cell lung cancer are mainly treated by radiotherapy.[2]

In the radiotherapy workflow for patients of LC, precise delineation of gross tumor volume (GTV) in computed tomography(CT) images is the essential step. Other tumor target areas are based on GTV and consider the influence of potential invaded tissues, positioning errors, and other factors. Inaccurate delineation of GTV will result in unnecessary damage to normal tissues or undertreatment in the tumor target area. In clinical practice, GTV is usually manually delineated by radiologists. However, manual delineation is a time-consuming and laborious process, and the start of radiotherapy will be delayed as a result.[3] In addition, manual delineation is a

[1] College of Bioengineering, Chongqing University, Chongqing, China
[2] Department of radiation oncology, Chongqing University Cancer Hospital, Chongqing, China

Corresponding Author:
Yongzhong Wu, MD, Department of radiation oncology, Chongqing University Cancer Hospital, Chongqing, China.
Email: yongzhongw_cq@126.com

subjective process, and the radiologist's experience will have a great influence on the delineation results. Multiple studies have reported that this process has considerable inter-observer and intra-observer variability.[4–7] Thus, it is necessary to develop suitable automatic segmentation methods to relieve the workload of radiologists in the definition of the target volume and improve the consistency of the target area delineation.

Deep learning (DL) is a subfield of AI and machine learning, which has achieved tremendous success in recent years in various fields in science.[8–10] In medical image segmentation, DL-based auto-segmentation techniques have been shown to provide significant improvements over more traditional approaches.[11,12] Convolution neural networks (CNNs) are the most successful and popular DL architecture applied to image processing. A lot of researches have confirmed that CNNs are helpful for tumor target delineation for radiotherapy for head and neck cancer, breast cancer, and rectal cancer.[13–20] Some scholars have also conducted research on automatic segmentation of lung tumor target volume based on CNNs.[21–25] To explore the role of deep learning-assisted delineation, Bi N *et al* used a dilated residual network to delineate the CTV of NSCLC for postoperative radiation therapy. And compared with manual delineation, a CNN-assisted delineation can achieve better segmentation accuracy, segmentation consistency, and segmentation efficiency.[21] In order to facilitate the analysis of geometric tumor changes during radiotherapy, A CNN model named A-net was designed to delineate the GTV of LC with a DSC of 0.82.[22] Zhang F *et al* proposed an automatic segmentation method based on ResNet and analyze the role of the DL-assisted method for GTV segmentation of NSCLC.[23] To monitor tumor response to therapy, Jiang J *et al* extended the full resolution residual neural network and developed the multiple resolution residually connected network for the tumor segmentation of NSCLC.[24] To achieve the delineation of GTV for LC stereotactic body radiation therapy, Cui Y *et al* proposed CT-based dense V-networks with a DSC of 0.82.[25] Based on the above research, we reason that the automatic segmentation of GTV for LC radiotherapy can be achieved through CNNs. However, the above studies have three issues. First, most of the above studies use 2D CNNs and ignore the high-dimensional spatial features of tumors.[21–24] When delineating the GTV of LC, the radiologist needs to refer to adjacent CT slices to determine the trend of tumor growth. Therefore, it is worth designing a 3D CNN to mine three-dimensional spatial information from CT images to segment GTV. Second, with the increase of the network depth, CNNs are prone to the problem of vanishing gradients, and some studies did not consider this problem.[22] Third, the contribution of each channel feature in CNNs to the prediction result is different. The performance of the model can be effectively improved by using the appropriate attention mechanism. However, this point is ignored in the above research.[21–25]

In this work, we proposed a 3D CNN named 3D ResSE-Unet to achieve GTV segmentation of stage III NSCLC on computed tomography(CT) images. The main innovations of this article are as follows. First, 3D convolution operations were used to mine the three-dimensional spatial correlation of GTV. And the influence of the depth of the 3D Unet on the segmentation results was explored. Second, we introduced the residual connection mechanism and channel attention mechanism into the 3D Unet to improve the robustness of the model. The residual connection was adopted to address the optimization problem and vanishing gradients. The channel attention mechanism was used to strengthen the model's representational power by selectively emphasizing useful features and suppressing useless ones. The modified version of 3D Unet was proposed to segment GTV from CT images of 214 stage III NSCLC patients. And compared with 3D Unet, 3D Res-Unet, and 3D SE-Unet, 3D ResSE-Unet can obtain superior results. Third, to solve category imbalance, we designed a mixed loss function based on the Dice loss and the Focal loss for GTV segmentation. Fourth, batch normalization(BN) was adopted in the network training process. It could prevent overfitting and improve the accuracy of the target delineation. Finally, the Dice similarity coefficient(DSC), 95th-percentile of Hausdorff distance(HD_{95}), and mean surface distance(MSD) were used to evaluate the accuracy of the model's prediction. And the complexity and segmentation time of segmentation models were also compared and analyzed.

Methods

The experimental process of this article mainly includes three steps: data preprocessing, segmentation model training, and segmentation result evaluation. The flowchart of the method can be seen in Figure 1.

Data sets. Data of patients with the stage of III NSCLCs from January 2017 to October 2020 in the department of radiation oncology, Chongqing University Cancer Hospital, were collected selectively. The clinical staging of tumors was based on the eighth edition of the International Association for Lung Cancer (IASLC). This work was approved by the ethics committee of Chongqing University Cancer Hospital(No. CZLS2021231-A, Date:13-Sep-2021). And written consents were provided by all patients to store their medical information in the hospital database. In addition, all patient details have been de-identified. A total of 214 patient data were collected selectively. 148 patient cases were randomly selected as the training set, 30 cases were used as the validation set, and 36 cases were used as the test set. The training set was used to train the segmentation model and learn the feature distribution of GTV. The validation set was used to filter the best segmentation model. And the segmentation performance of models on new data was tested by the testing set. The general characteristic of the training, the validation, and the testing sets are shown in Table 1.

The patients' data were acquired on Philips BigBore CT simulator(Philips Medical Systems, Madison, WI) set on helical scan mode(120kV,30mA), and slice thicknesses of 5mm or 3mm. Iodine contrast agents were used for all patients. And CT images were obtained with free breathing. Planning CT images and radiotherapy structure of each patient were all exported and they were all Digital Imaging and

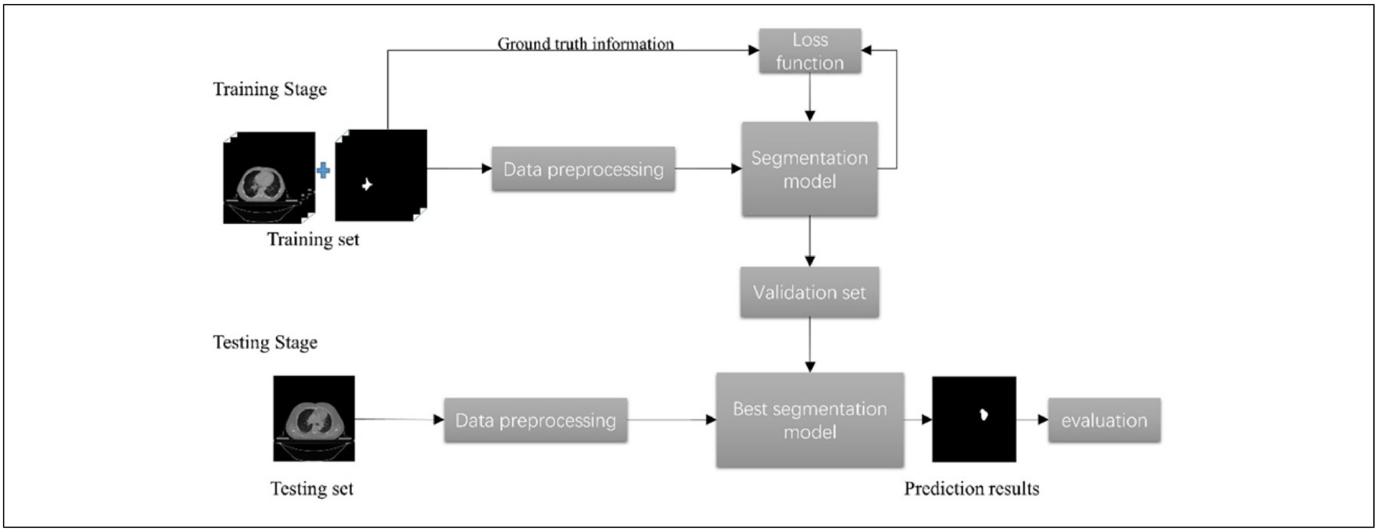

Figure 1. Flowchart of the 3D CNN-based segmentation method

Communications in Medicine(DICOM) files. Delineation of the GTV was carried out by a senior lung cancer radiologist who has more than 10 years of work experience and then peer-reviewed by two other experts. In this study, these GTV contours delineated by radiologists were referred to as the ground truth. The criteria for radiologists to delineate GTV of stage III NSCLC was based on NCCN Clinical Practice Guidelines in Oncology – Non-Small Cell Lung Cancer. And the primary gross tumor volume and the lymph node gross tumor volume were all included.

Preprocessing. To make full use of the three-dimensional spatial information of CT images, the images need to be processed according to the following steps. As shown in Figure 2, GTV contours are extracted from the radiotherapy structure of each patient by using python. And the CT images and GTV contours of each patient are converted into 3D matrices using the SimpleITK module. Then to maintain consistency across different patients, resampling operations were applied to the image and contour matrices so that each has a slice thickness of 5.0 mm and a pixel pitch of 1.0 mm. In order to reduce the computational burden and memory consumption, input images are randomly cropped into 3D volume with $160 \times 160 \times 32$ pixels. And to make full use of the spatial information of CT data, the input data is prepared as overlapping batches. The overlapping technique ensures that the segmentation model can utilize as much information over the third axis as possible. In addition, the overlap stride is set to 8 images for training data, but this method is not used in the validation data and the testing data. An example can be seen in Figure 3. In the end, 1159 blocks of 3D data are obtained in the training set, and 90 blocks of 3D data are obtained in the validation set.

In addition, considering the difference in CT value distribution between subjects, the pixel intensity of CT images is normalized to 0-1 by using Hounsfield(HU) window [-180,220]. Hounsfield(HU) window [-180,220] is the mediastinal CT window, and the radiation oncologist observes this window when delineating the GTV of LC. Finally, since the limited data resources, data augmentation is an unavoidable choice to get better performance on unseen data. Therefore, random zoom and random rotation are adopted to augment the training data. And this process is achieved by using the multi-dimensional image processing package(.ndimage) in the Scipy.

Architecture. In the field of medical image segmentation, U-net[26] has become one of the most well-known structures. 3D U-net[27] is an improved version of the basic U-net model and enables 3D volumetric segmentation using very few annotated examples. More importantly, the information on adjacent slices of an image can be transmitted through the network to provide more consistent predictions. The delineation of GTV is mainly dependent on the patient's anatomical structure and tumor presentation on the CT images. Thus, we propose to apply the 3D U-net model as the base model for GTV segmentation, and the influence of depth of 3D Unet on segmentation performance is analyzed. In order to further strengthen the ability to extract features and aspired by the ResNet[28] and SE-Net,[29] the residual connection and the channel attention mechanism are introduced into 3D Unet. The effects of these improvement methods are also compared.

In this paper, a model called 3D ResSE-Unet is proposed for target segmentation, which is an improved version of 3D Unet. The network diagram is shown in Figure4. It is composed of a contracting path to capture context and a symmetric expanding path that enables precise localization. Four max pool operations are stacked in the contracting path to reduce image resolution, expand the receptive field, and explore more detailed features. And in the expanding path, the image resolution is recovered by upsampling operations. To localize precisely, high-resolution features from the contracting path are combined with the upsampled output. Our network architecture contains 7 ResSE blocks, four max pool operations, and four upsampling operations. The last layer is a $1 \times 1 \times 1$

convolution to produce the predicted map. The network parameters are summarized in table 2.

The design of the ResSE block is presented in Figure4. The following expression can denote the details of residual connection:

$$y_l = x_l + F(x_l) \quad (1)$$

$$x_{l+1} = f(y_l) \quad (2)$$

x_l and x_{l+1} correspond to the input of the l_{th} layer and $(l+1)_{th}$ layer respectively. $F(\cdot)$ is denoted as the residual function, which is composed of several operations, including convolution, batch normalization (BN), rectified linear unit(ReLU), and SE block. $f(\cdot)$ is denoted as the activation function, and ReLU was used in this work. The residual block integrates x_l with the $F(x_l)$ to improve the information flow. This behavior allows the network to preserve feature maps in deeper neural networks, addressing vanishing gradients and making networks easier to optimize.

The SE module can selectively strengthen useful features and suppress useless features by learning to use global information, thereby achieving feature recalibration. As shown in Figure 4, C,H,W represent the channel number of the feature, the height, and the width of the feature image, respectively. And r represents the reduction ratio, and the value in this

Table 1. Characteristics of 214 patients with stage III NSCLC

Characteristics	Training Set	Validation Set	Testing Set
Patients number	148	30	36
Age			
Median (range)	61(39-78)	56(43-82)	61.5(45-76)
Gender			
Male/female	134/14	24/6	35/1
Tumor site			
Left/right	69/79	13/17	13/23
Tumor volume			
Median(range)	60.06cm^3(3.884cm^3-839.5 cm^3)	76.15cm^3(9.820cm^3-414.7 cm^3)	106.6cm^3(6.303 cm^3-527.3 cm^3)
Treatment			
IMRT/TOMO	143/5	26/4	35/1
Subtypes		17	
Squamous cell carcinoma	98	13	31
Adenocarcinoma	45		5
Large cell carcinoma	1		
Sarcomatoid carcinoma	2		
Adenosquamous carcinoma	1		
Unknown	1		
T stage			
T1	14	1	8
T2	34	8	5
T3	23	5	22
T4	73	16	1
Tx	4		
N stage			
N0	11	2	3
N1	6	2	18
N2	51	7	15
N3	80	19	

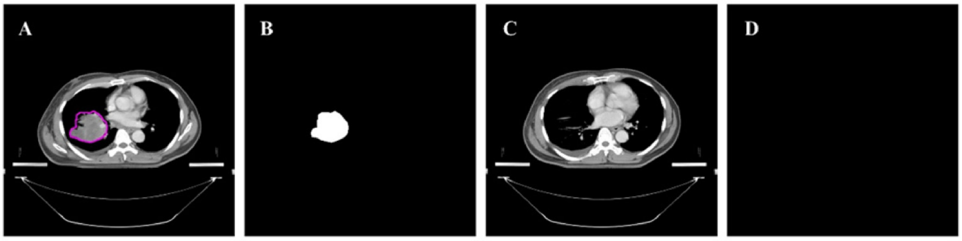

Figure 2. CT images and corresponding labels. A. CT images with GTV (red contour is the manually delineated GTV). B. Label images for the image presented in *A*. C. CT images without GTV. D. Label images for the image presented in *C*.

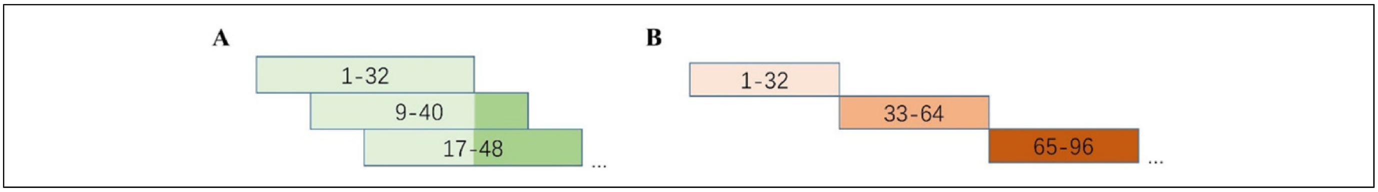

Figure 3. Examples of data cropping of training, validation, and testing set (Number show which CT image slices are included in 3D data). A. Example of data cropping of the training set. B. Example of data cropping of validation and testing set.

work is 2. This method implements attention weighting on channels in three steps. Firstly, global spatial information is squeezed using global average pooling, and the $1 \times 1 \times C$ channel feature map is generated in the end. The second is the excitation operation, in which a bottleneck with two fully connected (FC) layers around the ReLU unit is formed. In this process, first, compress the channel feature number to C/r, then go through a ReLU function to increase non-linearity, then restore the channel feature number to C, and finally go through a sigmoid function to obtain the weight of each channel feature. During this process, important channel features get larger weights, and unimportant channel features get smaller weights. Finally, each channel feature and the corresponding weight are multiplied as the output of the SE block.

Loss function. When training a CNN model, choosing an appropriate loss function can improve network performance. Considering that there is a problem of foreground-background class imbalance in this task. Thus, we designed a mixed loss function, as defined in Eq. 1:

$$L_{mix} = L_{dice} + L_{focal} \tag{1}$$

Where L_{dice} and L_{focal} represent the Dice loss and the Focal loss, respectively. And they are explained as follows:

$$L_{dice}(X, Y) = 1 - \frac{2|X \cap Y|}{|X| + |Y|} \tag{2}$$

Where X and Y represent the ground truth and the prediction result, the dice loss is suitable for severe class imbalance tasks. However, in the routine task, dice loss will influence the backpropagation and lead to training difficulty.

$$L_{focal}(p_t) = -\alpha_t(1 - p_t)^r \log(p_t) \tag{3}$$

$$p_t = \begin{cases} p & if \;\; y = 1 \\ 1 - p & otherwise \end{cases} \tag{4}$$

Where α_t is the variant to balance the importance of positive/negative examples, $(1 - p_t)^r$ is a modulating factor. The focal loss can be seen as the variation of Binary Cross-Entropy. Due to down-weight the contribution of easy examples and focusing more on learning hard examples, it works well for highly imbalanced class scenarios.

Evaluation. The testing set was used to evaluate the predictive performance of the 3D ResSE-Unet. The ground-truth volumes were contoured manually by experienced senior radiation oncologists. And the difference between auto-delineated GTV and the ground-truth was calculated by dice similarity coefficient(DSC), the 95th-percentile Hausdorff distance(HD$_{95}$), and mean surface distance(MSD).

Model training. The proposed models were implemented by Pytorch framework on Linux operating system using Python Application Programmable Interface and then accelerated by the NVIDIA graphics card. To prevent overfitting, a batch normalization operation was performed after each convolution operation. And the Kaiming function was used to configure network initialization parameters. In the training stage, the learning rate was set to 0.00015 in the Adam optimizer, the batch size was set to 2, and the mixed loss function was the training loss function. The max number of epochs was 90, and the loss value decreased with the epoch number. After each training epoch, validation was performed on the validation set, and only the best parameters would be saved. All experiments in this article were performed on Intel Xeon E5-2650 V4 (2.2GHz) processor and NVIDIA tesla T4 graphics card.

Result

After training the models, CT images of the testing set were imported into the best-performing model to perform GTV delineation and delineation results were evaluated qualitatively and quantitatively.

Comparison of different depths of 3D unet. To find a suitable depth of 3D Unet for GTV segmentation, different depths of 3D Unet were trained respectively. Different depths of 3D Unet include 3D Unet_3B, 3D Unet_4B, 3D Unet_5B, which respectively include three downsamplings, four downsamplings, and five downsamplings. The number of convolution channels in each layer of 3D Unet_3B from shallow to deep is 16, 32, 64, 128. Similarly, the number of convolutional channels in each layer of 3D Unet_4B is 16, 32, 64, 128, 256. And the number of convolution channels in each layer of 3D Unet_5B is 16, 32, 64, 128, 256, 512.

Quantitative evaluation results of three different depths of 3D Unet are summarized in Table 3. As shown, the 3D Unet_4B has realized better segmentation results. Its average values of DSC, Hausdorff distance, and average surface distance can reach 0.7090, 33.89mm, and 7.030mm, respectively.

And the three quantitative evaluation results of 3D Unet_3B and 3D Unet_5B are not as good as 3D Unet_4B.

The partial segmentation results of the testing set are shown in Figure 5. Intuitively, 3D Unet segmentation results of different depths all have the problem of false positives. However, compared with 3D Unet_3B and 3D Unet_5B, there are fewer false positives and false negatives in 3D Unet_4B.

Comparison of 3D ResSE-Unet, 3D Res-Unet, 3D SE-Unet and 3D Unet. To investigate the effectiveness of the proposed

Table 2. Network parameter

Layer	Operation	Kernel size	Stride	Num. of filters	Input size ($C_{in} \times D \times H \times W$)	Output size ($C_{out} \times D \times H \times W$)
Double conv1	(Conv3D + BN + Relu) × 2	3 × 3 × 3	(1,1,1)	16	1 × 32 × 160 × 160	16 × 32 × 160 × 160
Max pool 1	MaxPool3D	2 × 2 × 2	(2,2,2)		16 × 32 × 160 × 160	16 × 16 × 80 × 80
ResSEblock1	Conv3D + BN + Relu + SE block	3 × 3 × 3	(1,1,1)	32	16 × 16 × 80 × 80	32 × 16 × 80 × 80
Max pool 2	MaxPool3D	2 × 2 × 2	(2,2,2)		32 × 16 × 80 × 80	32 × 8 × 40 × 40
ResSEblock2	Conv3D + BN + Relu + SE block	3 × 3 × 3	(1,1,1)	64	32 × 8 × 40 × 40	64 × 8 × 40 × 40
Max pool 3	MaxPool3D	2 × 2 × 2	(2,2,2)		64 × 8 × 40 × 40	64 × 4 × 20 × 20
ResSEblock3	Conv3D + BN + Relu + SE block	3 × 3 × 3	(1,1,1)	128	64 × 4 × 20 × 20	128 × 4 × 20 × 20
Max pool 4	MaxPool3D	2 × 2 × 2	(2,2,2)		128 × 4 × 20 × 20	128 × 2 × 10 × 10
ResSEblock4	Conv3D + BN + Relu + SE block	3 × 3 × 3	(1,1,1)	256	128 × 2 × 10 × 10	256 × 2 × 10 × 10
Trans Conv1	ConvTranspose3D	2 × 2 × 2	(2,2,2)		256 × 2 × 10 × 10	128 × 4 × 20 × 20
ResSEblock5	Conv3D + BN + Relu + SE block	3 × 3 × 3	(1,1,1)	128	256 × 4 × 20 × 20	128 × 4 × 20 × 20
Trans Conv2	ConvTranspose3D	2 × 2 × 2	(2,2,2)		128 × 4 × 20 × 20	64 × 8 × 40 × 40
ResSEblock6	Conv3D + BN + Relu + SE block	3 × 3 × 3	(1,1,1)	64	128 × 8 × 40 × 40	64 × 8 × 40 × 40
Trans Conv3	ConvTranspose3D	2 × 2 × 2	(2,2,2)		64 × 8 × 40 × 40	32 × 16 × 80 × 80
ResSEblock7	Conv3D + BN + Relu + SE block	3 × 3 × 3	(1,1,1)	32	64 × 16 × 80 × 80	32 × 16 × 80 × 80
Trans Conv4	ConvTranspose3D	2 × 2 × 2	(2,2,2)		32 × 16 × 80 × 80	16 × 32 × 160 × 160
Double conv2	(Conv3D + BN + Relu) × 2	3 × 3 × 3	(1,1,1)	16	32 × 32 × 160 × 160	16 × 32 × 160 × 160
Last conv	Conv3D	3 × 3 × 3	(1,1,1)	2	16 × 32 × 160 × 160	2 × 32 × 160 × 160

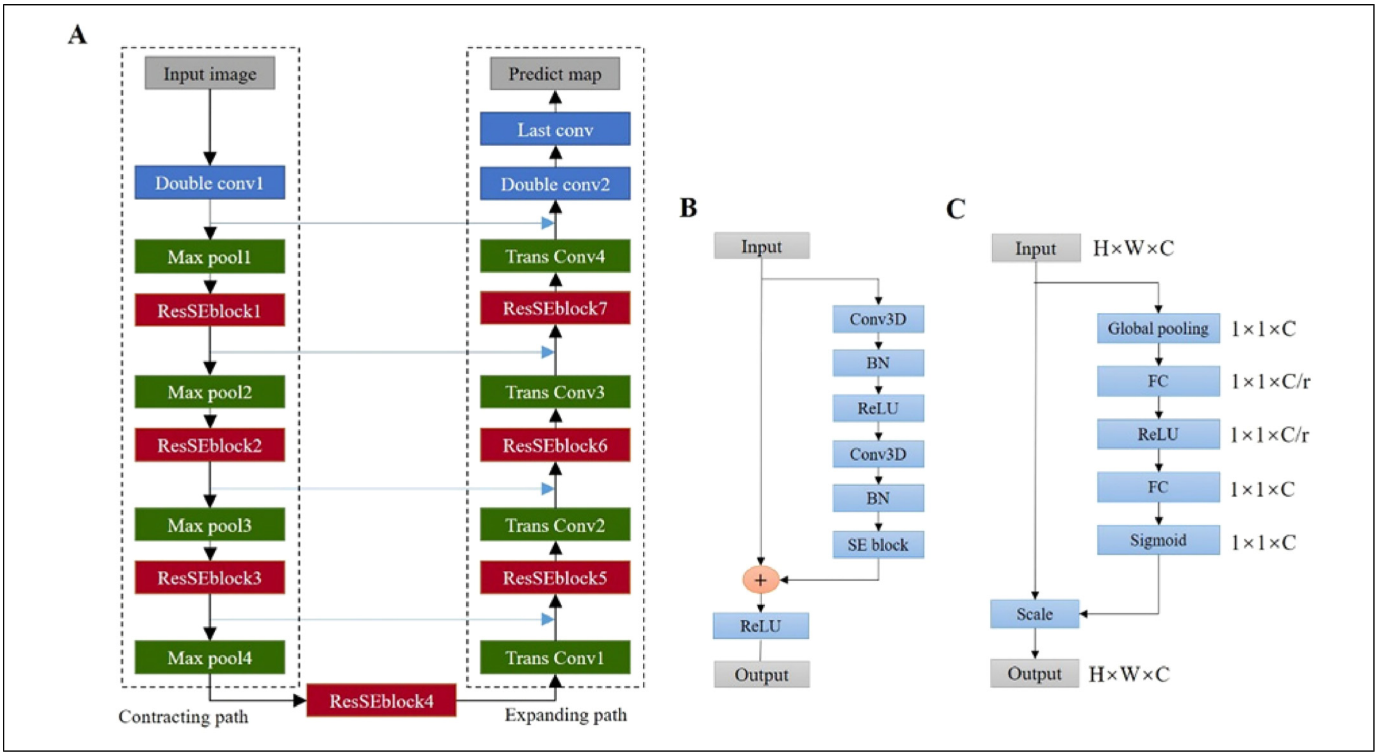

Figure 4. The diagram of the 3D ResSE-Unet structure. A. The architecture of 3D ResSE-Unet B. The design of ResSE block. C. The structure of SE block(H represents the height of the input feature, W represents the width of the input feature, C represents the channel number of the input feature, r represents reduction ratio, and the value in this work is 2).

segmentation model, 3D Unet, 3D Res-Unet, 3D SE-Unet, and 3D ResSE-Unet were trained respectively. Compared with 3D Unet, 3D Res-Unet introduced residual connection, 3D SE-Unet introduced channel attention mechanism, 3D ResSE-Unet introduced residual connection and channel attention mechanism at the same time. For useful comparison, when training 3D SE-Unet, 3D Res-Unet, and 3D ResSE-Unet, the hyperparameters were the same as those used in 3D Unet.

As shown in Table 4, the quantitative evaluation results of four networks on the testing set are summarized. It can be found that compared with the 3D Unet, the introduction of

Table 3. Comparison of quantitative evaluation metrics of 3D Unet with different depths ($\bar{x}$).

Method	DSC	HD_{95}(mm)	MSD(mm)
3D Unet_3B	0.6979	40.72	9.392
3D Unet_4B	0.7090	33.89	7.030
3D Unet_5B	0.6936	48.94	9.427

Table 4. Quantitative evaluation metrics comparison of different model($\bar{x}$).

Method	DSC	HD_{95}(mm)	MSD(mm)
3D Unet	0.7090	33.89	7.030
3D SE-Unet	0.7222	23.46	5.487
3D Res-Unet	0.7247	21.64	5.121
3D ResSE-Unet	0.7367	21.39	4.962

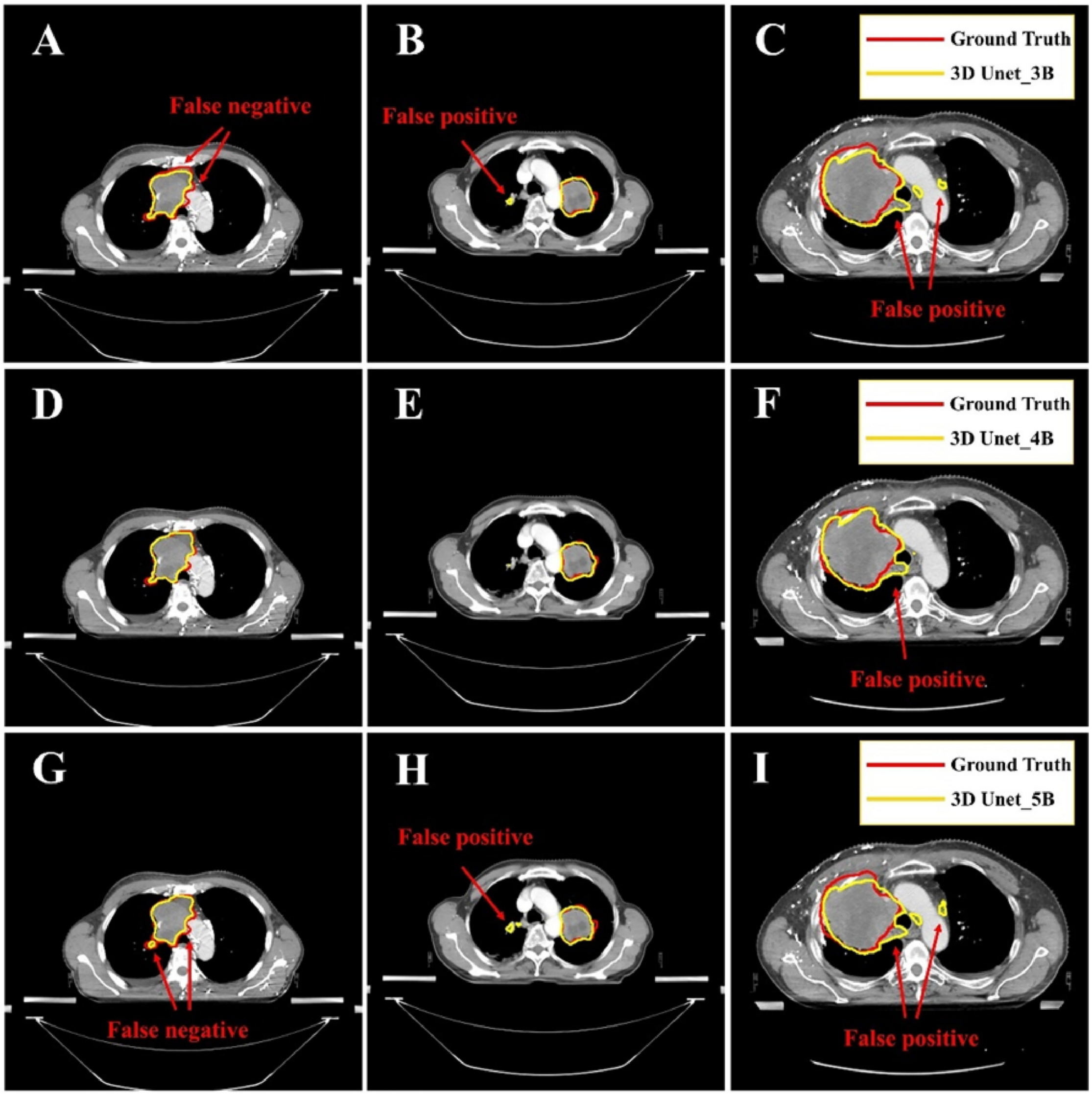

Figure 5. Comparison of segmentation results of different depths of 3D Unet. A-C. 3D Unet_3B segmentation results. D-F. 3D Unet_4B segmentation results. G-I. 3D Unet_5B segmentation results.

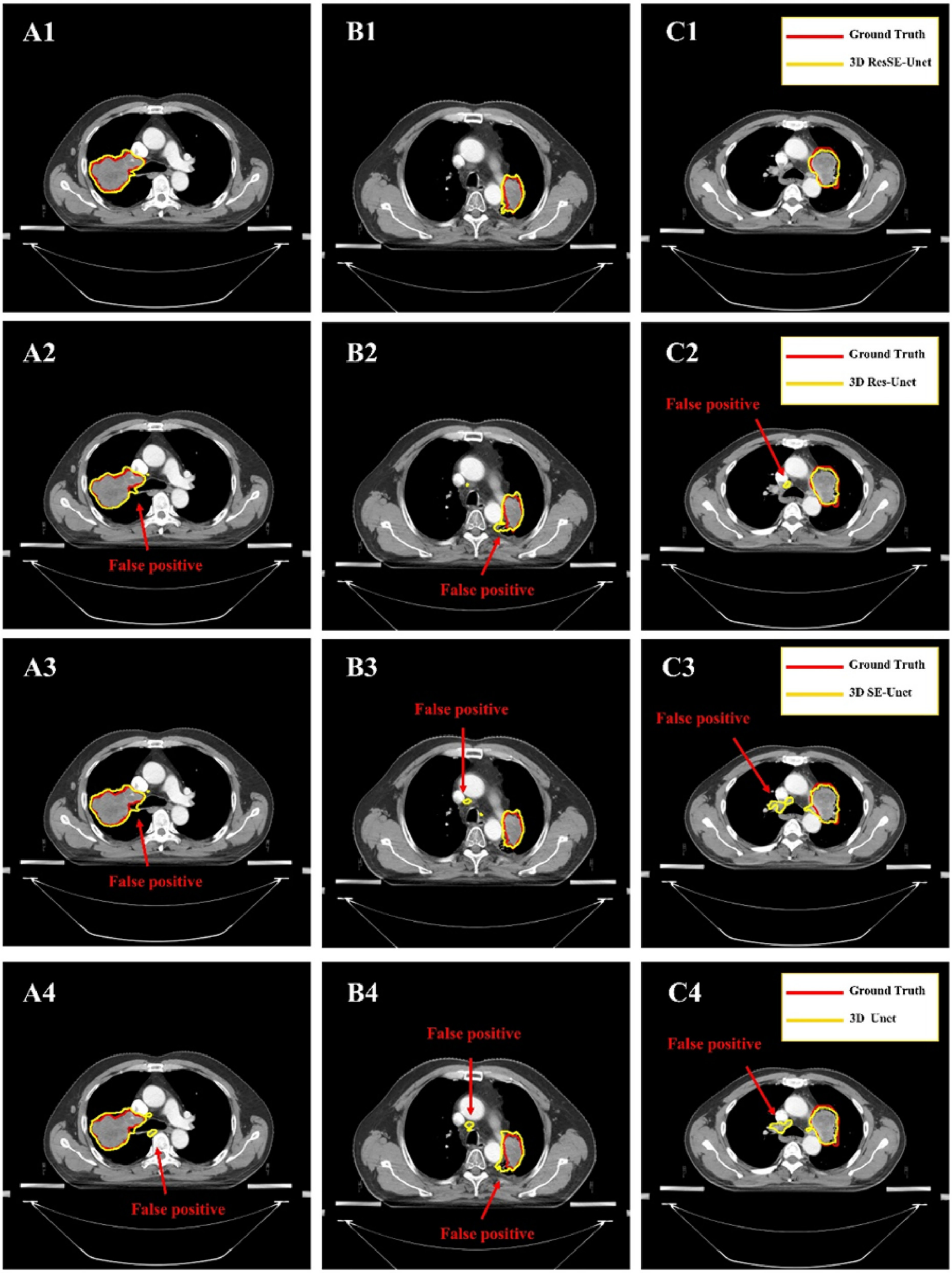

Figure 6. Comparison of segmentation results of 3D ResSE-Unet, 3D Res-Unet, 3D SE-Unet, and 3D Unet. A1-C1. 3D ResSE-Unet segmentation results. A2-C2. 3D Res-Unet segmentation results. A3-C3. 3D SE-Unet segmentation results. A4-C4. 3D Unet segmentation results.

residual connection and the introduction of channel attention mechanism can better improve the segmentation results of 3D Unet. The introduction of residual connection increases the DSC score of 3D Unet from 0.7090 to 0.7247 and reduces HD_{95} from 33.89mm to 21.64mm, and MSD from 7.030mm to 5.121mm. The introduction of the channel attention mechanism increases the DSC score of 3D Unet from 0.7090 to 0.7222 and reduces HD_{95} from 33.89mm to 23.46mm, and MSD from 7.030mm to 5.487mm. In addition, the introduction of residual connection and channel attention mechanisms at the same time can achieve better segmentation results, and the average values of DSC, HD, and MSD can reach 0.7367, 21.39mm, 4.962mm, respectively.

The representative comparison results of four models are shown in Figure 6. As shown, there is the problem of false positives in the segmentation results of 3D Unet. However, the introduction of residual connection and channel attention mechanism can better solve this problem. Intuitively, compared with 3D Unet, the problem of false positives has been improved in the results of 3D Res-Unet, 3D SE-Unet, and 3D ResSE-Unet. And it can be found that 3D ResSE-Unet realizes the best results.

In addition, we also compared the network parameters and average segmentation time of the four models, as shown in Table 5. It can be seen that the introduction of the channel attention mechanism hardly increases the number of model parameters and does not reduce the segmentation efficiency. But the introduction of residual connections will increase the number of model parameters and slightly reduce the efficiency of segmentation. And it can be found that compared with 3D Unet, 3D ResSE-Unet parameters increase from 21.54MB to 44.66MB, but the average segmentation time only increases by 1 second.

Discussion

Radiotherapy is one of the main treatments for stage III NSCLC. Accurately delineating GTV is essential to achieve precise radiotherapy. Radiologist manual delineation is time-consuming and has inter-and intra-observer variability. However, these problems can be solved by automatic segmentation methods based on CNNs. At present, the research on the automatic delineation of GTV for NSCLC radiotherapy mainly uses 2D CNNs and ignores spatial features of tumors from CT data. In this work, we chose 3D Unet as the base model and used two different methods to improve 3D Unet. We designed a model named 3D ResSE-Unet and achieved the automatic segmentation of GTV of stage III NSCLC radiotherapy. The segmentation results of different depths of 3D Unet are shown in Table 3 and Figure 5. From the perspective of the 3D Unet structure, the deeper the network, the more feature scales that can be extracted, and the better the segmentation results will be obtained. This can explain why the segmentation results of 3D Unet_4B are better than 3D Unet_3B. However, the deeper the network, the more spatial information is lost through max pooling operations, which is not suitable for segmenting small targets. In the training set, some tumors were too small, and the minimum GTV was 3.884 cm^3. And the overlap cropping technique in the preprocessing will cause the 3D data block to be trained only to contain part of the GTV. Therefore, the segmentation result of 3D Unet_5B is not as good as 3D Unet_4B.

Compared with other depths, 3D Unet with four downsamplings is the more suitable structure for our work, but the segmentation results still have the problem of false positives. Two methods were adopted to solve this problem in this article. To solve the problem of vanishing gradients and strengthen the transmission of features, the residual connection mechanism is introduced into 3D Unet. And the channel attention mechanism also has been introduced into 3D Unet to strengthen the useful channel features and suppress the useless channel features. As shown in Table 4, 5, and Figure 6, compared with 3D Unet, the introduction of residual connection and channel attention mechanism can solve the problem of false positives and improve the segmentation performance. Especially, 3D ResSE-Unet realizes the best results. Although the introduction of channel attention mechanism and residual connection will slightly reduce the segmentation efficiency, it only takes 10s to segment one case and still can meet the needs of clinical applications.

The comparison between the proposed approach in this article and three lung tumor delineation methods developed in previous papers has been summarized in Table 6. Compared with 2D CNNs,[21,23] the proposed model can obtain the same segmentation accuracy while using fewer cases. This is due to the overlapping technique used in preprocessing. In this way, each case is fully utilized as much as possible and 3D data blocks to be trained in the training set have been expanded. And the segmentation model can make full use of the z-axis information of the CT image. With the same number of cases, our method is more likely to obtain better segmentation performance. Compared with the research of Cui Y et al,[25] our segmentation results are average. The dense connections and V-net used in their segmentation model provide new ideas for our follow-up research. However, the influence of each convolution channel feature on the prediction result is ignored in their study. Their segmentation performance may be further improved by introducing the channel attention mechanism.

Although we have achieved automatic segmentation of GTV for stage III NSCLC, our experiment still has the following limitations. Firstly, only 214 cases of stage III NSCLC have been collected for our experiment. This number is relatively small

Table 5. Comparison of network parameters and average segmentation time.

	Parameters(MB)	Average segmentation time per patient(s)
3D Unet	21.54	9.58
3D SE-Unet	21.75	9.60
3D Res-Unet	44.24	10.58
3D ResSE-Unet	44.66	10.61

Table 6. Comparison of 3D ResSE-Unet and other methods of tumor delineation for NSCLC radiotherapy

	CNN	Number of patients	Modality	Results
Bi N et al[21]	DD-ResNet (2D)	269(NSCLC) Training:200 Validation:50 Testing:19	CT	DSC:0.75
Zhang F et al[23]	modified ResNet (2D)	330(NSCLC) Training:300 Testing:30	CT	DSC:0.73
Cui Y et al[25]	DVNs (3D)	192(NSCLC SBRT) 10-fold cross validation (training:174-175,test:19-20)	CT	DSC:0.832 HD:4.57mm
This article	3D ResSE-Unet	214(NSCLC) Training:148 Validation:30 Testing:36	CT	DSC:0.7367 HD_{95}:21.39mm

and needs to be further increased. The tumor location, shape, and size of different patients will be very different. Increasing the number of cases used for training may further improve the generalization ability and prediction accuracy of the segmentation model. Secondly, we have only realized the automatic segmentation of stage III NSCLC. The segmentation effect of this model on GTV of stages I, II, and IV NSCLC needs further study. Thirdly, compared with other cancers, lung tumors vary greatly in size, shape, and location. The relationship between these features and segmentation accuracy has not been further analyzed. Fourth, there is no further comparison between deep learning-assisted delineation and manual delineation in terms of efficiency and inter-and intra-observer variability. Fifth, we only performed a joint assessment of the primary gross tumor volume and the lymph node gross tumor volume and did not analyze their segmentation results separately. Sixth, our department did not adopt respiratory motion management until 2018, and in order to obtain enough cases, we collected patients from 2017 to 2020, so our experiments were carried out with free breathing.

In the future, we can make some new attempts to achieve better segmentation performance. Firstly, compared to the residual connection, a more extreme connection pattern has been developed, which is called the dense connection.[30] In this pattern, each layer receives the output features of all previous layers as input and passes its feature maps to all subsequent layers. And the dense connection also can alleviate the vanishing gradient problem and encourage feature reuse. In future work, the residual connection may be replaced with dense connections. Secondly, the channel attention mechanism only pays attention to the difference of different channel information but ignores local information in each channel. However, the local spatial attention mechanism[31–33] can solve this problem by calculating the feature importance of each pixel in the space domain. Thus, combining the advantages of the two attention mechanisms to improve the segmentation effect is the next work that can be studied. Thirdly, our research is only based on CT images, which can provide high-resolution anatomical details. Currently, PET/CT and magnetic resonance images(MRI) have been widely used in the diagnosis and treatment of cancer. PET images can provide quantitative metabolic information. MRI can provide clear soft tissue contrast and help to distinguish the tumor from the surrounding normal tissues. Integrating multi-modal images can obtain richer tumor feature information and may improve the accuracy of tumor segmentation. And some scholars have carried out researches based on multi-modal images.[34–38]

Conclusion

In this article, a 3D CNN named 3D ResSE-Unet is proposed for GTV segmentation of stage III NSCLC. This model can fully excavate the three-dimensional spatial information of tumors and realize accurate and rapid segmentation of GTV. 3D ResSE-Unet is based on 3D Unet and combines the advantages of residual connection and channel attention mechanism. Compared with 3D Unet, 3D ResSE-Unet segmentation can achieve more accurate segmentation and can solve the problem of over-segmentation. This model provides a new tool for realizing the automatic delineation of GTV for lung cancer radiotherapy. But the current segmentation results still need to be adjusted manually before clinical application. In the future, the proposed method may be further improved to improve segmentation accuracy and efficiency and assist to achieve accurate and effective radiotherapy.

Acknowledgements

The authors thank the radiation oncologists in Department of radiation oncology, Chongqing University Cancer Hospital, for the target delineation.

Declaration of Conflicting Interests

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

Ethics statement

The study was approved by the ethics committee of Chongqing University Cancer Hospital (No. CZLS2021231-A, Date:13-Sep-2021).

Funding

The work was supported by National Natural Science Foundation of China(81972857 and 11805025) and Innovative Project of Shapingba Diatrict Chongqing(Jcd202135).

ORCID iD

Xinhao Yu https://orcid.org/0000-0003-1270-864X

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA-Cancer J Clin*. Nov-Dec 2018;68(6):394-424. doi:10.3322/caac.21492
2. Miller KD, Nogueira L, Mariotto AB, et al. Cancer treatment and survivorship statistics, 2019. *CA Cancer J Clin*. 2019;69(5):363-385. doi:10.3322/caac.21565
3. Vorwerk H, Zink K, Schiller R, et al. Protection of quality and innovation in radiation oncology: the prospective multicenter trial the German society of radiation oncology (DEGRO-QUIRO study) evaluation of time, attendance of medical staff, and resources during radiotherapy with IMRT. Article. *Strahlenther Onkol*. Apr 2014;190(5):433-443. doi:10.1007/s00066-014-0634-0
4. Van de Steene J, Linthout N, de Mey J, et al. Definition of gross tumor volume in lung cancer: inter-observer variability. Article. *Radiother Oncol*. Jan 2002;62(1):37-49. doi:10.1016/s0167-8140(01)00453-4
5. Louie AV, Rodrigues G, Olsthoorn J, et al. Inter-observer and intra-observer reliability for lung cancer target volume delineation in the 4D-CT era. Article. *Radiother Oncol*. May 2010;95(2):166-171. doi:10.1016/j.radonc.2009.12.028
6. Giraud P, Elles S, Helfre S, et al. Conformal radiotherapy for lung cancer: different delineation of the gross tumor volume (GTV) by radiologists and radiation oncologists. Article. *Radiother Oncol*. Jan 2002;62(1):27-36. doi:10.1016/s0167-8140(01)00444-3
7. Steenbakkers R, Duppen JC, Fitton I, et al. Observer variation in target volume delineation of lung cancer related to radiation oncotogist-computer interaction: a 'Big Brother' evaluation. Article. *Radiother Oncol*. Nov 2005;77(2):182-190. doi:10.1016/j.radonc.2005.09.017
8. Meyer P, Noblet V, Mazzara C, Lallement A. Survey on deep learning for radiotherapy. Article. *Comput Biol Med*. Jul 2018;98:126-146. doi:10.1016/j.compbiomed.2018.05.018
9. Sahiner B, Pezeshk A, Hadjiiski LM, et al. Deep learning in medical imaging and radiation therapy. Review. *Med Phys*. Jan 2019;46(1):e1-e36. doi:10.1002/mp.13264
10. Maier A, Syben C, Lasser T, Riess C. A gentle introduction to deep learning in medical image processing. Review. *Z Med Phys*. 2019;29(2):86-101. doi:10.1016/j.zemedi.2018.12.003
11. Cardenas CE, Yang JZ, Anderson BM, Court LE, Brock KB. Advances in auto-segmentation. Article. *Semin Radiat Oncol*. Jul 2019;29(3):185-197. doi:10.1016/j.semradonc.2019.02.001
12. Hesamian MH, Jia W, He XJ, Kennedy P. Deep learning techniques for medical image segmentation: achievements and challenges. Article. *J Digit Imaging*. Aug 2019;32(4):582-596. doi:10.1007/s10278-019-00227-x
13. Wang Y, Zu C, Hu G, et al. Automatic tumor segmentation with deep convolutional neural networks for radiotherapy applications. *Neural Process Lett*. 2018;48(3):1323-1334. doi:10.1007/s11063-017-9759-3
14. Li SH, Xiao JH, He L, Peng XC, Yuan XD. The tumor target segmentation of nasopharyngeal cancer in CT images based on deep learning methods. Article. *Technol Cancer Res Treat*. Nov 2019;18:8. doi:10.1177/1533033819884561
15. Men K, Zhang T, Chen XY, et al. Fully automatic and robust segmentation of the clinical target volume for radiotherapy of breast cancer using big data and deep learning. *Phys Medica*. Jun 2018;50:13-19. doi:10.1016/j.ejmp.2018.05.006
16. Guo Z, Guo N, Gong K, Zhong SA, Li QZ. Gross tumor volume segmentation for head and neck cancer radiotherapy using deep dense multi-modality network. Article. *Phys Med Biol*. Oct 2019;64(20):14. doi:10.1088/1361-6560/ab440d
17. Cardenas CE, Anderson BM, Aristophanous M, et al. Auto-delineation of oropharyngeal clinical target volumes using 3D convolutional neural networks. Article. *Phys Med Biol*. Nov 2018;63(21):12. doi:10.1088/1361-6560/aae8a9
18. Men K, Chen X, Zhang Y, et al. Deep deconvolutional neural network for target segmentation of nasopharyngeal cancer in planning computed tomography images. *Front Oncol*. 2017:7315. doi:10.3389/fonc.2017.00315
19. Cai M, Wang J, Yang Q, et al. Combining images and T-staging information to improve the automatic segmentation of nasopharyngeal carcinoma tumors in MR images. *IEEE Access*. 2021;9:21323-21331. doi:10.1109/access.2021.3056130
20. Men K, Dai JR, Li YX. Automatic segmentation of the clinical target volume and organs at risk in the planning CT for rectal cancer using deep dilated convolutional neural networks. *Med Phys*. Dec 2017;44(12):6377-6389. doi:10.1002/mp.12602
21. Bi N, Wang JB, Zhang T, et al. Deep learning improved clinical target volume contouring quality and efficiency for post-operative radiation therapy in non-small cell lung cancer. Article. *Front Oncol*. Nov 2019;9(8):1192. doi:10.3389/fonc.2019.01192
22. Wang C, Tyagi N, Rimner A, et al. Segmenting lung tumors on longitudinal imaging studies via a patient-specific adaptive convolutional neural network. *Radiother Oncol*. 2019;131:101-107. doi:10.1016/j.radonc.2018.10.037
23. Zhang F, Wang Q, Li H. Automatic segmentation of the gross target volume in non-small cell lung cancer using a modified version of ResNet. *Technol Cancer Res Treat*. 2020;19:1915330338209474 84. doi:10.1177/1533033820947484
24. Jiang J, Hu YC, Liu CJ, et al. Multiple resolution residually connected feature streams for automatic lung tumor segmentation from CT images. Article. *IEEE Trans Med Imaging*. Jan 2019;38(1):134-144. doi:10.1109/tmi.2018.2857800

25. Cui Y, Arimura H, Nakano R, Yoshitake T, Shioyama Y, Yabuuchi H. Automated approach for segmenting gross tumor volumes for lung cancer stereotactic body radiation therapy using CT-based dense V-networks. *J Radiat Res*. Mar 2021;62-(2):346-355. doi:10.1093/jrr/rraa132
26. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. *Springer International Publishing*. 2015:234-241.
27. Cicek O, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-net: learning dense volumetric segmentation from sparse annotation. *Springer Verlag*. 2016:424-432.
28. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *IEEE Comput Soc.* 2016:770-778.
29. Hu J, Shen L, Albanie S, Sun G, Wu E. Squeeze-and-Excitation networks. *IEEE Trans Pattern Anal Mach Intell.* 2020;42-(8):2011-2023. doi:10.1109/TPAMI.2019.2913372
30. Huang G, Liu Z, Pleiss G, Van Der Maaten L, Weinberger K. Convolutional networks with dense connectivity. *IEEE Trans Pattern Anal Mach Intell.* 2019:1-1. doi:10.1109/tpami.2019.2918284
31. Jaderberg M, Simonyan K, Zisserman A, Kavukcuoglu K. Spatial transformer networks. *Neural Information Processing Systems Foundation*. 2015:2017-2025.
32. Oktay O, Schlemper J, Le Folgoc L, et al. Attention U-net: learning where to look for the pancreas. *2018:ArXiv:1804.03999.* Accessed April 01, 2018; https://ui.adsabs.harvard.edu/abs/2018arXiv180403999O.
33. Kaul C, Manandhar S, Pears N. Focusnet: an attention-based fully convolutional network for medical image segmentation. 2019:arXiv:1902.03091. https://ui.adsabs.harvard.edu/abs/2019arXiv190203091K
34. Jiang J, Alam SR, Chen I, et al. Deep cross-modality (MR-CT) educed distillation learning for cone beam CT lung tumor segmentation. *Med Phys.* 2021. doi:10.1002/mp.14902
35. Zhong ZS, Kim YS, Plichta K, et al. Simultaneous cosegmentation of tumors in PET-CT images using deep fully convolutional networks. Article. *Med Phys.* Feb 2019;46(2):619-633. doi:10.1002/mp.13331
36. Jiang J, Hu YC, Tyagi N, et al. Cross-modality (CT-MRI) prior augmented deep learning for robust lung tumor segmentation from small MR datasets. Article. *Med Phys.* Oct 2019;46-(10):4392-4404. doi:10.1002/mp.13695
37. Fu X, Bi L, Kumar A, Fulham M, Kim J. Multimodal spatial attention module for targeting multimodal PET-CT lung tumor segmentation. *IEEE J Biomed Health Inform.* 2021;25:3507-3516. doi:10.1109/jbhi.2021.3059453
38. Zhao XM, Li LQ, Lu W, Tan S. Tumor co-segmentation in PET/CT using multi-modality fully convolutional neural network. Article. *Phys Med Biol*. Jan 2019;64(1):015011. doi:10.1088/1361-6560/aaf44b