

Automatic Segmentation of the Gross Target Volume in Non-Small Cell Lung Cancer Using a Modified Version of ResNet

Technology in Cancer Research & Treatment
 Volume 19: 1-9
 © The Author(s) 2020
 Article reuse guidelines:
sagepub.com/journals-permissions
 DOI: 10.1177/1533033820947484
journals.sagepub.com/home/tct


Fuli Zhang, PhD¹ , Qiusheng Wang, PhD², and Haipeng Li, MS²

Abstract

Radiotherapy plays an important role in the treatment of non-small cell lung cancer. Accurate segmentation of the gross target volume is very important for successful radiotherapy delivery. Deep learning techniques can obtain fast and accurate segmentation, which is independent of experts' experience and saves time compared with manual delineation. In this paper, we introduce a modified version of ResNet and apply it to segment the gross target volume in computed tomography images of patients with non-small cell lung cancer. Normalization was applied to reduce the differences among images and data augmentation techniques were employed to further enrich the data of the training set. Two different residual convolutional blocks were used to efficiently extract the deep features of the computed tomography images, and the features from all levels of the ResNet were merged into a single output. This simple design achieved a fusion of deep semantic features and shallow appearance features to generate dense pixel outputs. The test loss tended to be stable after 50 training epochs, and the segmentation took 21 ms per computed tomography image. The average evaluation metrics were: Dice similarity coefficient, 0.73; Jaccard similarity coefficient, 0.68; true positive rate, 0.71; and false positive rate, 0.0012. Those results were better than those of U-Net, which was used as a benchmark. The modified ResNet directly extracted multi-scale context features from original input images. Thus, the proposed automatic segmentation method can quickly segment the gross target volume in non-small cell lung cancer cases and be applied to improve consistency in contouring.

Keywords

deep learning, automatic segmentation, gross target volume, non-small cell lung cancer, residual convolutional block, convolutional neural network

Abbreviations

CNN, convolutional neural network; CT, computed tomography; DSC, Dice similarity coefficient; FPR, false positive rate; GTV, gross target volume; JSC, Jaccard similarity coefficient; MRI, magnetic resonance imaging; NSCLC, non-small cell lung cancer; PET, positron emission tomography; TPR, true positive rate; ResNet, residual neural network.

Received: February 16, 2020; Revised: July 3, 2020; Accepted: July 9, 2020.

Introduction

Lung cancer is a major cause of cancer-related death among both men and women, accounting for 1.6 million deaths annually worldwide.^{1,2} According to the 2017 China Urban Cancer Data Report, lung cancers are the most prevalent malignancies in terms of morbidity and mortality in urban areas, and non-small cell lung cancer (NSCLC) accounts for about 75% to 80% of all lung cancer cases. Patients with lung cancer require comprehensive treatment, and radiotherapy can be used in all stages. At least one session of radiotherapy is performed on

¹ Radiation Oncology Department, The Seventh Medical Center of Chinese PLA General Hospital, Beijing, China

² School of Automation Science and Electrical Engineering, Beihang University, Beijing, China

Corresponding Author:

Fuli Zhang, Radiation Oncology Department, The Seventh Medical Center of Chinese PLA General Hospital, Beijing 100700, China.
 Email: radiozfli@163.com



over half of patients for either curative or palliative purposes.³ Radiotherapy of NSCLC requires accurate location about the tumor. Precise, patient-specific radiotherapy plans are usually designed based on computed tomography (CT) images to deliver high irradiated doses to the target volume while sparing organs at risk (OARs) as much as possible.⁴

Accurate segmentation of the target volume is very important for successful radiotherapy delivery. Traditionally, such segmentation has been performed by manual delineation on planning CT images, with the help of magnetic resonance imaging (MRI) or positron emission tomography (PET) images when necessary. However, manual delineation is labor-intensive, time-consuming, and subjective, and it has considerable inter- and intra-observer variability.^{5,6} Thus, accurate automatic segmentation methods are highly desired and useful for pre-treatment radiotherapy planning.

In recent years, deep learning methods have gained popularity and shown outstanding capabilities in autosegmentation of tumors of regions such as the head and neck, breast, and rectum.⁷⁻¹⁸ Some studies have also reported deep learning-based automatic segmentation of lung tumors.¹⁹⁻²⁴ Bi Nan et al. focused on autosegmentation of the clinical target volume (CTV) of lung cancer using a network based on ResNet-101.¹⁹ Four recent studies addressed the application of deep learning-based autosegmentation in primary lung tumors on the basis of MR or PET/CT images.²⁰⁻²³ Jiang J et al. developed two multiple resolution residually connected networks (MRRN) for lung cancer but only evaluated the accuracy of the proposed models without clinical implications such as time gain.²⁴ Thus, few studies have explored the role of deep learning in autosegmentation of the gross target volume (GTV) of NSCLC on CT images as well as the efficiency of autosegmentation in end-to-end clinical application.

In this paper, we used a modified version of ResNet to segment the GTV in patients with medically/technically inoperable NSCLC. We adopted an encoder-decoder structure similar to U-Net, which is becoming more widely used in semantic segmentation tasks.²⁵ Compared with fully convolutional networks, which rely heavily on the use of atrous convolution to generate high-quality segmentation results, it has shown outstanding performance in terms of memory and computing power and no limitations on the type of backbone network that can be used.^{26,27} As the encoder, ResNet34 was used to fully extract deeper image features and prevent the disappearance of the gradient during the training process.²⁸ The decoder employs feature fusion and upsampling that are inspired by feature pyramid, but more lightweight.^{29,30} The features from all levels of the ResNet34 architecture were merged into a single output to achieve deep fusion of deep semantic features with shallow appearance ones, which increased the accuracy of the segmentation results. Then, the performance of the modified ResNet was compared with that of U-Net, which is commonly used in medical image processing.^{12,14,20} Finally, the efficiency of the autosegmentation in clinical work was evaluated by comparison with manual delineation in terms of contouring time.

Materials and Methods

Dataset

Patients with different NSCLC staging, tumor size, and location were included. All patients were aged 40–89 years. For patients with lymph node involvement, only the primary tumor mass was assessed in this study. The full CT image dataset was divided into two parts: the training and test sets. The training set consisted of 300 patients, whose tumor sizes (but not tumor positions) had a roughly even distribution. To ensure that the test set would facilitate objective evaluation of the automatic segmentation model, the test set included 30 patients with an even spread of tumor sizes and tumor positions. The general characteristics of the patients in the training and test sets are shown in Table 1. All patients received regular 3D CT scans.

The CT images were acquired on a Philips Brilliance Big Bore simulator (Philips Medical Systems, Madison, WI) from the level of the larynx to the bottom of the lungs with 5-mm slice thickness on helical scan mode. There were 57–121 slices per patient. The study was approved by the ethics committee of the Seventh Medical Center of people's liberation army (PLA) General Hospital. All patients provided written consent for storage of their medical information in the hospital database. The GTVs of all cases (including training and test cases) were delineated by an experienced senior radiation oncologist who specializes in the thoracic region (Xu WD) and were then peer-reviewed by two other experts (Wang YD and Gao JM). These manual delineations were used to generate the ground truth in this study. Then, the CTV was defined as the GTV plus additional margins of 6 mm and 8 mm for squamous cell carcinoma and adenocarcinoma, respectively. Further additional margins were added to form the final planning target volume.

Data Preprocessing and Augmentation

The CT image values of different tissues and organs were converted into gray values and stored in Digital Imaging and Communications in Medicine format. To highlight the information about the tumor and the surrounding tissues in the CT images, it was necessary to convert the 4096-level grayscale images into 255-level ones.

The performance of convolutional neural networks (CNNs) relies heavily on the size of the training dataset used. In our study, data augmentation techniques were employed to further enrich the data of the training set. The size, shape, and location of tumors vary from patient to patient in CT images of NSCLC, but the pixel intensity of tumors after CT scanning remains relatively fixed. Therefore, the data augmentation techniques applied in GTV autosegmentation did not change the pixel intensity values in the original CT images. Applying flip, translation, scaling, and cropping operations to a CT image only made minor adjustments to the position, shape, and size of the tumor without changing the pixel intensity, which was equivalent to creating a new CT image. Thus, augmentation helped the network to learn invariance and reduced overfitting during network training. In addition, the data augmentation was

Table 1. Characteristics of Patients in Training and Test Sets.

Characteristics	Training set	Testing set
No. patients	300	30
Tumor site, right: left	136:164	13:17
Stage at diagnosis	I = 11; II = 84; III = 118; IV = 87	I = 2; II = 8; III = 11; IV = 9
Tumor volume, median (range)	85.4cc(1.6cc-678.6 cc)	75.4cc(5.6cc-349.4 cc)
Lobe location		
Upper left	112	10
Lower left	67	7
Upper right	53	6
Middle right	48	4
Lower right	20	3
Contact with chest wall	31	4
Contact with mediastinum	105	15

performed randomly: the image data and their augmentations were both randomly chosen. During each epoch, 2/3 of the training set was randomly selected for augmentation. To create each augmented image, 4 augmentation techniques were randomly combined. Therefore, each epoch's training images consisted of two groups: one comprised 4032 augmented images, which accounted for 2/3 of the training set, and the other comprised 2016 original images, which accounted for 1/3 of the training set. These two groups comprised the entire training set. The same transformations were also applied to the test set. Figure 1 shows an example of a CT image and its ground truth labeling before and after data augmentations.

Proposed Deep Learning Model

Convolutional layers, pooling layers, and the activation function are the basic components of a CNN.³¹ Convolutional layers convolve an image using convolutional kernels to obtain feature maps in which each element is connected to the previous layer by the corresponding weights of the convolutional kernels.¹³ Pooling layers fuse the spatial features of adjacent pixels into feature maps, making the images' feature representations more compact.²⁵ The activation function is responsible for increasing the network's nonlinear expression.

The configuration of the modified ResNet is shown in Figure 2. An encoder-decoder was used to increase feature resolution. The encoding path used a ResNet34 backbone to fully extract the deep features of the CT images while avoiding any performance degradation caused by deepening the network. A lightweight dense-prediction branch was applied in the decoding path. A simple design that merges the features from all levels of the ResNet34 into a single output was proposed: deep semantic features at multiple spatial resolutions were concatenated in the channel dimension and then were merged with shallow appearance features to generate dense pixel outputs.

The ResNet34-based encoder was divided into five stages, each of which generated feature map output at different scales. The ResNet34 architecture employed cross-layer connection via identity mapping, which could learn new features in addition to receive input features, effectively solving the network degradation problem because of the deep layered network structure. The structure of the residual learning block contained an identity residual block and a convolutional residual block, as shown in Figure 3.

In the identity residual block, the input X was passed directly to the output as the initial result. The network's learning goal was changed from the desired output $H(X)$ to the difference between the desired output and the input, which was called the residual $F(X)=H(X) - X$. The identity mapping would have been achieved if $F(X) = 0$, and therefore, $H(X)=X$. However, the residual was not actually equal to 0. The convolutional residual block added the convolution values to the appropriate branch of the identity map to change the dimension of the feature map. In the experiment, the identity residual block and the convolutional residual block were used interchangeably.

The decoding network's feature fusion and upsampling were inspired by the feature pyramid network, but the present network's advantage is the fusion of deep semantic features with shallow appearance ones. We proposed a simple architecture that merges the features from all levels of the ResNet34 into a single output. The feature maps of stages 3, 4, and 5 were upsampled by bilinear upsampling and convolution until they reached 1/4 of the input image's scale. These deep semantic features (which included different levels of global information) were concatenated in the channel dimension to form "thicker" features.³² The feature maps were passed through a stack of convolutional layers to fuse different features and reduce the channel dimension until the number of channels was the same as the number of feature maps in stage 2. Then, the feature values were summed to increase the expression in a single channel. Finally, the feature maps were upsampled by bilinear upsampling and convolution again until they reached the scale of the input image, and then pixel-level classification was performed.

Model Training

The proposed models were implemented in the Python development environment on Windows operating system using a custom version of the Keras framework based on Tensorflow. All training and test experiments were run on an NVIDIA GeForce GTX 1080 Ti GPU with 11 GB memory. The dimensions of the training data and the minibatch size played a significant role in the computational burden of the proposed autosegmentation method. A minibatch of size 8 was used because of the GPU's limited memory.

All input images were single-channel grayscale ones of size 512×512 . Normalization was performed using population-level data to prevent gradient convergence from slowing because of differences in CT images' magnitudes. The

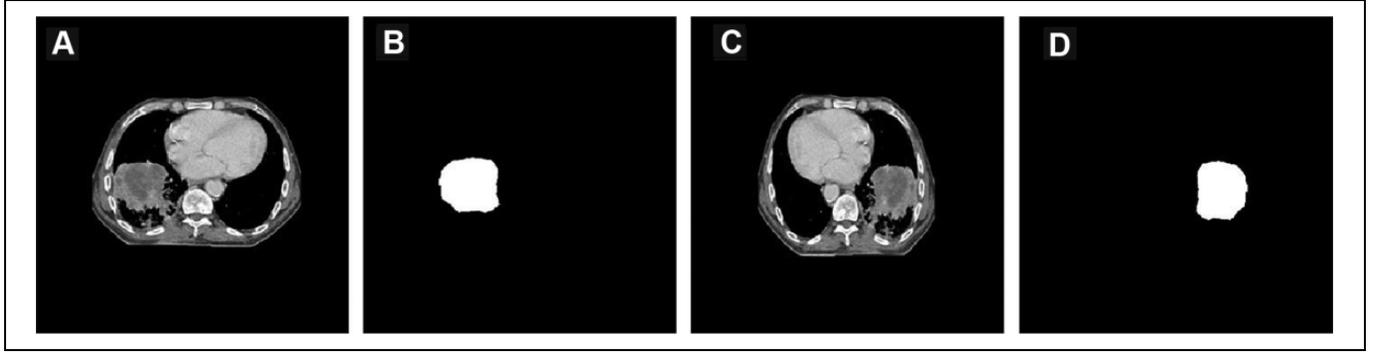


Figure 1. CT images and corresponding labels before and after data augmentation. A. Original CT image. B. Original ground truth. C. Augmented CT image. D. Augmented ground truth.

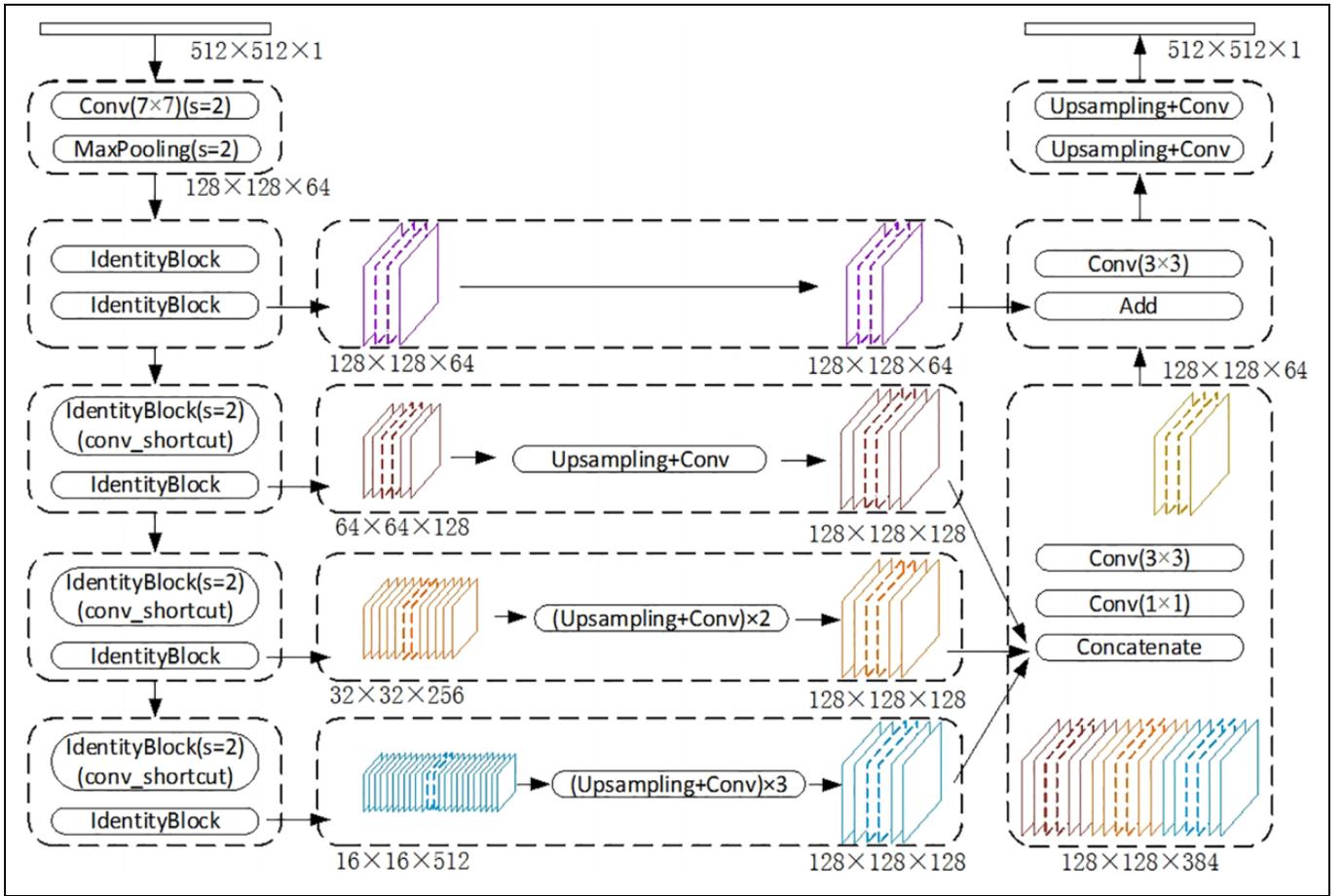


Figure 2. Proposed network structure.

encoding network performed convolution, pooling, and activation operations on the normalized input images to obtain feature maps. The size of the feature maps was continuously reduced until it reached 16×16 , and then the size of feature maps was gradually restored to 512×512 by the decoding network. Finally, the sigmoid function was used to predict the probability that each pixel belonged to each category, and cross-entropy was applied to measure the similarity between the probability distribution predicted by the model and that of real samples. Weighted cross-entropy was applied to force the

loss function to pay more attention to the foreground class,³³ as defined in Eq. 1:

$$\left[H(y) = -\frac{1}{N} \sum_{i=1}^N \left(\frac{N - \sum_i \hat{y}_i}{\sum_i \hat{y}_i} y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \right) \right] \quad (1)$$

where y_i belongs to the probability distribution of real samples, and \hat{y}_i belongs to the predicted one.

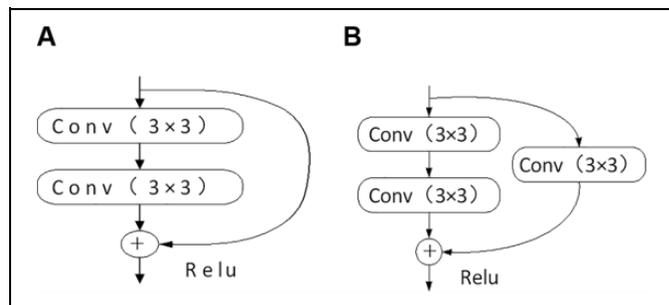


Figure 3. Structure of the residual learning block. A. identity residual block. B. convolutional residual block.

The network training duration was set as 50 epochs. Adaptive Adam optimization was applied because of the sparseness of single images' data. Adam performs better than other types of optimization, such as stochastic gradient descent, with a sparse gradient.^{34,35} The loss drops quickly when the Adam optimizer is used, so the learning rate is relatively small (set as 0.0001). Each parameter's learning rate was dynamically adjusted by the first and second moment estimations of the gradient. The exponential decay rates of the first and second moment estimations in the Adam optimization algorithm were set as 0.9 and 0.999, respectively. To prevent overfitting, a dropout layer and L2 regularization were employed in our study. Additionally, it was important to use batch normalization, which helps the inputs of each layer of the neural network maintain the same distribution, making training easier.

Evaluation Metric

The proposed model's performance was verified by comparison with the expert segmentation results, which were regarded as the ground truth. We evaluated the similarity between the autosegmentation results and ground truth by the Dice similarity coefficient (DSC), Jaccard similarity coefficient (JSC), true positive rate (TPR), and false positive rate (FPR).

Statistical Analysis

Student's *t*-test was performed to compare the autosegmentation results of the two models by using SPSS 20 software (version 20.0, SPSS Inc, Chicago, USA). Quantitative data were expressed in the form of mean \pm standard deviation ($\bar{x} \pm s$). A value of $p \leq 0.05$ was considered statistically significant.

Results

Delineation Results by the Modified ResNet

Training of the modified ResNet took approximately 12 hours. In the encoder, the size of the feature maps shrank, but the contained semantic information increased, at deeper network layers. In the decoder, the size of the feature maps expanded with increased numbers of upsampling operations. After 50 epochs of training, the cross-entropy loss of the training and test datasets were stable at 0.0009 and 0.0080, respectively.

During dense-pixel prediction, if a pixel value from the sigmoid function was greater than 0.5, the pixel was considered as part of the target. Otherwise, the pixel was considered as part of the background. The autosegmentation method based on the modified ResNet was applied to the test dataset, and the DSC, JSC, TPR, and FPR values were calculated. The proposed method achieved comparable results to the manual segmentation, especially for larger tumors. Figure 4 shows typical autosegmentation results of test set images along with their corresponding reference results.

Figure 4 shows that there was a large overlap between the autosegmentation results and the ground truth. The average DSC value of the entire test set was 0.73, and the DSC of the best single test image reached 0.96.

Comparison of the Modified ResNet and U-Net

For further comparison, we also segmented the GTV of NSCLC using U-Net, which is commonly applied as a benchmark in medical image segmentation. The U-Net was trained from scratch on this dataset, and the hyperparameters were consistent with those used by the modified ResNet. The settings were as follows: single-channel grayscale images of size 512×512 and Adam optimization algorithm with base learning rate 0.0001. A sufficient number of training epoch iterations (50) were performed to ensure that the network was well-trained. The average values of each metric used in the two different networks are listed in Table 2. The DSC, JSC, and TPR values of the modified ResNet were higher than those of U-Net, indicating that the modified ResNet performed better than U-Net at GTV autosegmentation of NSCLC images ($P < 0.05$). Figure 5 shows the autosegmentation results of the test set images along with the reference results obtained using two different networks.

Time Gain

The average time for autosegmentation of the GTV by the proposed model was about 21 ms per slice, but the results still required additional manual slice-by-slice modification before the final clinical implementation. The deep learning-assisted delineation took an average of about 10 min per patient, in contrast to manual delineation taking an average of about 15 min per patient. Thus, the proposed technique significantly improved the efficiency of segmentation in our routine clinical work.

Discussion

The contouring of target volumes is an important aspect of treatment planning in radiotherapy but is usually time-consuming, and the quality of the contours relies on the operator's skill level. In recent years, deep learning-based automatic segmentation has become very popular.

Some studies have reported automatic segmentation of lung tumors using deep learning methods. Bi Nan et al. applied a 2D

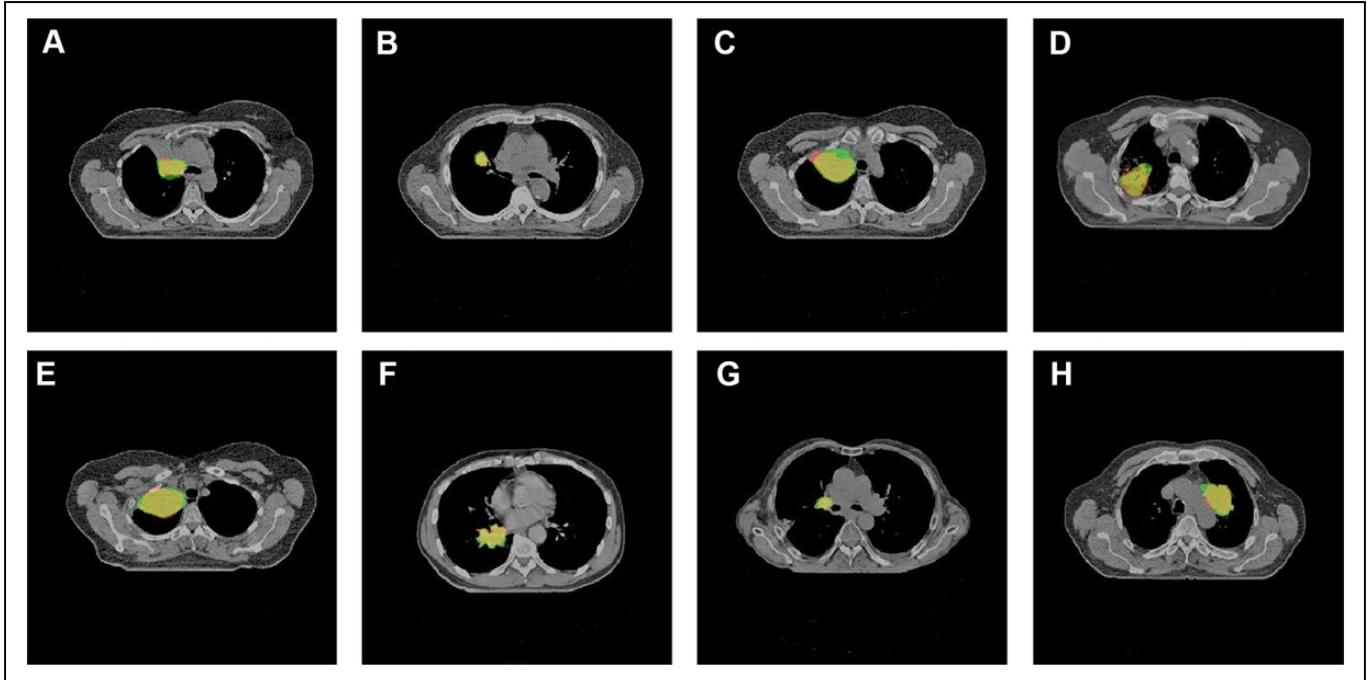


Figure 4. Segmentation results of GTV. The green part indicates the ground truth. The red part indicates the auto-segmentation results. The yellow part indicates the intersection between the two.

Table 2. Quantitative Evaluation Metrics of the Modified ResNet and U-Net ($\bar{x} \pm s$).

	Modified ResNet	U-Net	P
DSC	0.73 \pm 0.07	0.64 \pm 0.09	0.000
JSC	0.68 \pm 0.09	0.52 \pm 0.12	0.000
TPR	0.74 \pm 0.07	0.61 \pm 0.10	0.000
FPR	0.0012 \pm 0.0014	0.0008 \pm 0.0004	0.099
Segmentation time/slice(ms)	21 \pm 7	28 \pm 8	0.000

ResNet101 to achieve effective auto-segmentation of the CTV in postoperative lung cancer. Deep learning-assisted contouring by 11 junior physicians achieved an average DSC of 0.75.¹⁹ Wang C et al. developed a patient-specific adaptive CNN called A-net to simulate the workflow of adaptive radiotherapy and used past weekly MRI data and target volumes to segment lung tumors on the current weekly MRI. The patient-specific A-net can segment tumors with an average DSC of 0.82. Thus, it outperformed the population-based A-net, which had an average DSC of 0.64, as well as the population-based U-net, which had an average DSC of 0.59.²⁰ Zhong ZS used two coupled 3D U-Nets to realize auto-segmentation of NSCLC tumors in PET-CT images, and the average DSCs on CT and PET were 0.86 and 0.83, respectively. Subsequent cosegmentation by the deep learning method using both PET and CT data outperformed the results using either PET or CT alone.²¹ Zhao XM et al proposed a novel multi-modality segmentation method based on a 3D FCN for lung tumor auto-segmentation on PET-CT images. The results demonstrated that the proposed network was effective,

fast, and robust and achieved significant performance gain over CNN-based methods and traditional methods using PET or CT data only (average DSC: 0.85).²² Jiang J, Hu YC, et al. proposed an adversarial domain adaptation-based deep learning method for automatic NSCLC tumor segmentation on T2-weighted MRI images. The proposed method used a U-Net trained with a limited number of original MRIs and some synthesized ones. The method produced a DSC of 0.74 when trained with only synthesized MRIs, and the best DSC (0.80) was achieved on the test set when the model was trained in a semi-supervised setting.²³ Jiang J, Hu YC, et al. developed two multiple resolution residually connected network (MRRN) formulations that simultaneously combine features across multiple image resolution and feature levels through residual connection to detect and segment lung tumors. The method achieved average DSCs of 0.74, 0.75, and 0.68 for 3 different datasets.²⁴

In this study, we proposed a modified ResNet model for auto-segmentation of the GTV of NSCLC, which achieved rapid and fairly complete end-to-end tumor auto-segmentation. We compared the proposed modified ResNet with U-Net. The proposed modified ResNet performed better than U-Net and improved the consistency of contouring, which could help to streamline radiotherapy workflows. Two possible reasons could be proffered for this: 1) The ResNet34 backbone network of the encoder introduced more nonlinear expressions and extracted deeper, more semantically advanced features. 2) The decoder's simple design that merged the features from all levels of the ResNet34 into a single output to generate dense-pixel prediction was effective. Deep semantic features at multiple spatial resolutions were concatenated in the channel dimension

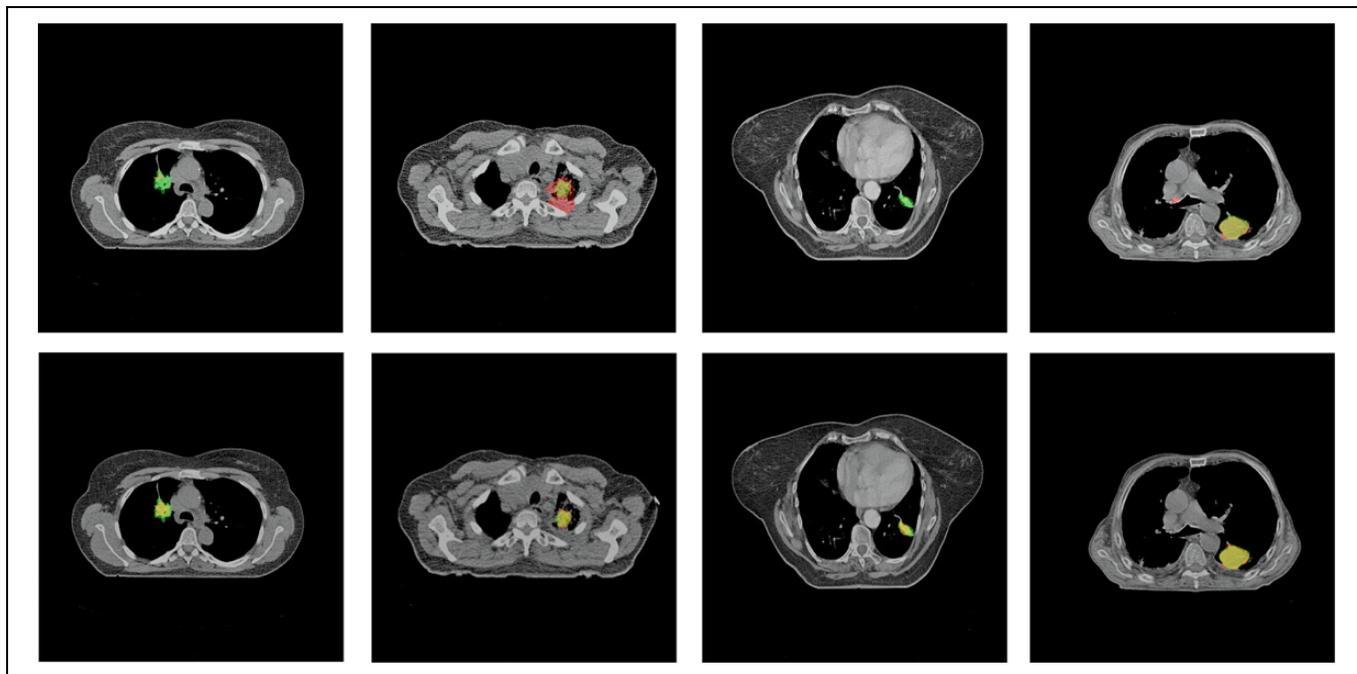


Figure 5. Segmentation results of the GTV by the proposed network (bottom figures) and U-Net (top figures). The green part indicates the ground truth. The red part indicates the auto-segmentation results. The yellow part indicates the intersection between the two.

and then summed with shallow appearance features to increase the expression in a single channel. Compared with U-Net, these two improvements ensure that the modified ResNet can directly extract multi-scale context features from the original input images. Thus, the modified ResNet performed better at this task than U-Net.

The developed approach appears promising, but some aspects of the study have limitations. Firstly, the training set did not contain many different cases because of the limited number of available image sets, whereas the GTV auto-segmentation results were affected by variations in tumor position (e.g., some tumors were peripherally located, centrally located, or had broad chest wall contact), shape, size, and respiratory and cardiac motion, so correlations between size/shape/location and the evaluation metrics were not discussed. Secondly, distance metrics like Hausdorff distance or distance to agreement were not applied to measure the auto-contour's degree of spatial conformity. Thirdly, inter-observer and intra-observer variability was not examined in this study. The lack of such performance assessment between the two methods used in this study results in a limited ability to evaluate geometric discrepancy.

In addition, if the tumors were small or attached to the mediastinum closely, accurate segmentation would have been difficult, as the 2D network implementation ignored the relationship between different CT slices of the same patient. Thus, employing a 2.5D or 3D network should be considered to improve segmentation of the GTV of NSCLC. A 3D network could obtain more accurate segmentation results on CT sequences by using information about the previous slice to guide the segmentation of the next slice. A 2.5D network may

also provide 3D-like context. However, because of the problem of matching the amount of data and model parameters, a 3D model would need a larger training dataset to avoid overfitting. Additionally, attention mechanisms, which can control the importance of features at different spatial locations through a gating signal, are a potential solution. As the size of the training dataset increases, the difference between the weights of the attention map in the target and background areas should increase, which would improve the accuracy of auto-segmentation.

To achieve better label balancing of the loss function, the network performs weighted learning of difficult samples adaptively. For example, the focal loss function proposed for the target detection task focuses on samples with incorrect classifications during the training process.³⁶ Combining sensitivity and specificity as the network's loss function is also a popular trend in medical segmentation tasks, and it has achieved accurate segmentation for small tumors.

Conclusion

A modified version of ResNet was able to perform automatic segmentation quickly and accurately. The automatically generated contours offered a good starting point for segmentation of primary tumors, but the results still require some manual modification before final clinical implementation. Moreover, the modified ResNet was efficient and conducive to reducing oncologists' labor intensity. Compared with U-Net, the modified ResNet was notably more accurate based on overlap and receiver operating characteristic curves. In addition, because the shape, size, and spatial orientation of the lung tumors in the

dataset varied greatly, data augmentation was adopted. The performance of the proposed method, including accuracy and efficiency, can be further improved in our future work.

Acknowledgments

The authors thank the radiation oncologists in Radiation Oncology Department, 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 the seventh Medical Center of PLA General Hospital (No. 2018-24).

Funding

The study was supported by Beijing Municipal Science and Technology Commission (No.Z181100001718011). The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Fuli Zhang  <https://orcid.org/0000-0002-7213-9031>

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin.* 2018;68(4):7-30. doi:10.3322/caac.21442
2. Ferlay J, Soerjomataram I, Dikshit R, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer.* 2015;136(5):E359-E386. doi:10.1002/ijc.29210
3. Tang W, Peng Q, Lyu Y, et al. Risk prediction models for lung cancer: perspectives and dissemination. *Chin J Cancer Res.* 2019; 31(2):316-328. doi:10.21147/j.issn.1000-9604.2019.02.06
4. Liu HH, Balter P, Tutt T, et al. Assessing respiration-induced tumor motion and internal target volume using four-dimensional computed tomography for radiotherapy of lung cancer. *Int J Radiat Oncol Biol Phys.* 2007;68(2):531-540. doi:10.1016/j.ijrobp.2006.12.066
5. van Mourik AM, Elkhuisen PH, Minkema D, et al. Multiinstitutional study on target volume delineation variation in breast radiotherapy in the presence of guidelines. *Radiother Oncol.* 2010; 94(3):286-291. doi:10.1016/j.radonc.2010.01.009
6. Barkati M, Simard D, Taussky D, Delouya G. Magnetic resonance imaging for prostate bed radiotherapy planning: an inter- and intra-observer variability study. *J Med Imaging Radiat Oncol.* 2016;60(2):255-259. doi:10.1111/1754-9485.12416
7. Iqbal S, Ghani MU, Saba T, Rehman A. Brain tumor segmentation in multi-spectral MRI using convolutional neural networks (CNN). *Microsc Res Tech.* 2018;81(4):419-427. doi:10.1002/jemt.22994
8. Janardhanprabhu S, Malathi V. Brain tumor detection using depth-first search tree segmentation. *J Med Syst.* 2019;43(8): 254. doi:10.1007/s10916-019-1366-6
9. Liu Y, Stojadinovic S, Hrycushko B, et al. A deep convolutional neural network-based automatic delineation strategy for multiple brain metastases stereotactic radiosurgery. *PLoS One.* 2017; 12(10):e0185844. doi:10.1371/journal.pone.0185844
10. Razzak MI, Imran M, Xu G. Efficient brain tumor segmentation with multiscale two-pathway-group conventional neural networks. *IEEE J Biomed Health Informat.* 2019;23(5):1911-1919. doi:10.1109/JBHI.2018.2874033
11. Thillaikkarasi R, Saravanan S. An enhancement of deep learning algorithm for brain tumor segmentation using kernel based CNN with M-SVM. *J Med Syst.* 2019;43(4):84. doi:10.1007/s10916-019-1223-7
12. Men K, Dai J, Li Y. 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.* 2017;44(12):6377-6389. doi:10.1002/mp.12602
13. Trebeschi S, van Griethuysen JJM, Lambregts DMJ, et al. Deep learning for fully-automated localization and segmentation of rectal cancer on multiparametric MR. *Sci Rep.* 2017;7(1):5301. doi:10.1038/s41598-017-05728-9
14. Jian J, Xiong F, Xia W, et al. Fully convolutional networks (FCNs)-based segmentation method for colorectal tumors on T2-weighted magnetic resonance images. *Aust Phys Eng Sci Med.* 2018;41(2):393-401. doi:10.1007/s13246-018-0636-9
15. Bauer S, Nolte LP, Reyes M. Fully automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field regularization. *Med Image Comp Compassist Interv.* 2011;14(pt 3): 354-361. doi:10.1007/978-3-642-23626-6_44
16. Men K, Zhang T, Chen X, et al. Fully automatic and robust segmentation of the clinical target volume for radiotherapy of breast cancer using big data and deep learning. *Phys Med.* 2018;50:13-19. doi:10.1016/j.ejmp.2018.05.006
17. Men K, Boimel P, Naylor JJ, et al. Cascaded atrous convolution and spatial pyramid pooling for more accurate tumor target segmentation for rectal cancer radiotherapy. *Phys Med Biol.* 2018; 63(18):185016. doi:10.1088/1361-6560/aada6c
18. Lin L, Dou Q, Jin YM, et al. Deep learning for automated contouring of primary tumor volumes by mri for nasopharyngeal carcinoma. *Radiology.* 2019;291(3):677-686. doi:10.1148/radiol.2019182012
19. Bi N, Wang J, Zhang T, et al. Deep learning improved clinical target volume contouring quality and efficiency for postoperative radiation therapy in non-small cell lung cancer. *Front Oncol.* 2019;9:1192. doi:10.3389/fonc.2019.01192
20. 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
21. Zhong Z, Kim Y, Plichta K, et al. Simultaneous cosegmentation of tumors in PET-CT images using deep fully convolutional networks. *Med Phys.* 2019;46(2):619-633. doi:10.1002/mp.13331
22. Zhao X, Li L, Lu W, Tan S. Tumor co-segmentation in PET/CT using multi-modality fully convolutional neural network. *Phys Med Biol.* 2018;64(1):015011. doi:10.1088/1361-6560/aaf44b

23. Jiang J, Hu YC, Tyagi N, et al. Tumor-aware, Adversarial domain adaptation from CT to MRI for lung cancer segmentation. *Med Image Comput Comp Assist Interv.* 2018;11071:777-785. doi:10.1007/978-3-030-00934-2_86
24. Jiang J, Hu YC, Liu CJ, et al. Multiple resolution residually connected feature streams for automatic lung tumor segmentation from CT Images. *IEEE Trans Med Imaging.* 2019;38(1):134-144. doi:10.1109/TMI.2018.2857800
25. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. *Med Image Comput Comput-Assist Interv.* 2015;9351:234-241. doi:10.1007/978-3-319-24574-4_28
26. Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions. *ICLR.* 2016.
27. Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation. *IEEE Trans Pattern Anal Mach Intell.* 2017;39(4):640-651. doi:10.1109/TPAMI.2016.2572683
28. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *IEEE Conference on Computer Vision and Pattern Recognition.* 2016:770-778. doi:10.1109/CVPR.2016.90
29. Nakai H, Nishio M, Yamashita R, et al. Quantitative and qualitative evaluation of convolutional neural networks with a deeper u-net for sparse-view computed tomography reconstruction. *Acad Radiol.* 2019;67(4):563-574. doi:10.1016/j.acra.2019.05.016
30. Lin TY, Dollár Piotr, Girshick R, et al. Feature pyramid networks for object detection. *IEEE Conference on Computer Vision and Pattern Recognition.* 2017:936-944. doi:10.1109/CVPR.2017.106
31. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM.* 2017;60(6):84-90. doi:10.1145/3065386
32. Kirillov A, Girshick R, He K, Dollár P. Panoptic feature pyramid networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 2019:6399-6408.
33. Sudre CH, Li W, Vercauteren T, Ourselin S, Cardoso MJ. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support.* Cham; 2017;240-248. doi:10.1007/978-3-319-67558-9_28
34. Kingma D, Ba J. Adam: A method for stochastic optimization. *Published as a conference paper at the 3rd International Conference for Learning Representations.* 2014.
35. Bottou L. *Stochastic Gradient Descent Tricks.* Springer; 2012:421-436. doi:10.1007/978-3-642-35289-8_25
36. Lin TY, Goyal P, Girshick R, Kaiming HE, Dollar P. Focal loss for dense object detection. *IEEE Trans Pattern Anal Mach Intell.* 2020;42(2):318-327. doi:10.1109/TPAMI.2018.2858826