

# Automatic delineation of organs at risk in non-small cell lung cancer radiotherapy based on deep learning networks\*

Anning Yang<sup>1</sup>, Na Lu<sup>2</sup>, Huayong Jiang<sup>2</sup>, Diandian Chen<sup>2</sup>, Yanjun Yu<sup>2</sup>, Yadi Wang<sup>2</sup>, Qiusheng Wang<sup>1</sup>, Fuli Zhang<sup>2</sup> (✉)

<sup>1</sup> School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China

<sup>2</sup> Department of Radiation Oncology, the Seventh Medical Center of Chinese PLA General Hospital, Beijing 100700, China

## Abstract

**Objective** To introduce an end-to-end automatic segmentation method for organs at risk (OARs) in chest computed tomography (CT) images based on dense connection deep learning and to provide an accurate auto-segmentation model to reduce the workload on radiation oncologists.

**Methods** CT images of 36 lung cancer cases were included in this study. Of these, 27 cases were randomly selected as the training set, six cases as the validation set, and nine cases as the testing set. The left and right lungs, cord, and heart were auto-segmented, and the training time was set to approximately 5 h. The testing set was evaluated using geometric metrics including the Dice similarity coefficient (DSC), 95% Hausdorff distance (HD95), and average surface distance (ASD). Thereafter, two sets of treatment plans were optimized based on manually contoured OARs and automatically contoured OARs, respectively. Dosimetric parameters including Dmax and Vx of the OARs were obtained and compared.

**Results** The proposed model was superior to U-Net in terms of the DSC, HD95, and ASD, although there was no significant difference in the segmentation results yielded by both networks ( $P > 0.05$ ). Compared to manual segmentation, auto-segmentation significantly reduced the segmentation time by nearly 40.7% ( $P < 0.05$ ). Moreover, the differences in dose-volume parameters between the two sets of plans were not statistically significant ( $P > 0.05$ ).

**Conclusion** The bilateral lung, cord, and heart could be accurately delineated using the DenseNet-based deep learning method. Thus, feature map reuse can be a novel approach to medical image auto-segmentation.

**Key words:** non-small cell lung cancer; organs at risk; medical image segmentation; deep learning; DenseNet

Received: 15 January 2022

Revised: 11 March 2022

Accepted: 10 April 2022

In a typical clinical workflow of radiotherapy (RT), a radiation oncologist manually segments a tumor target and the organs at risk (OARs) based on the information provided by the computed tomography (CT), magnetic resonance, and/or positron emission tomography/CT images [1, 2]. However, this process is usually time-consuming and laborious, and the quality of segmentation significantly depends on the prior knowledge and/or experience of the radiation oncologist. Although it is easy

to distinguish organs with high contrast in CT images, distinguishing the boundary between a tumor tissue and surrounding normal tissue with similar contrast is difficult. Furthermore, inconsistencies in the target and OARs segmentations have been reported for both inter- and intra-observer segmentation variability [3–7]. These factors affect the accuracy and efficacy of RT. Therefore, it is imperative to improve the consistency and efficiency of image segmentation.

✉ Correspondence to: Fuli Zhang. Email: radiozfli@163.com

\* Supported by a grant from the Beijing Municipal Science and Technology Commission Foundation Programme (No. Z181100001718011).

© 2022 Huazhong University of Science and Technology

In recent years, deep learning-based automatic medical image segmentation learning has emerged as a popular topic in the field of RT<sup>[8-10]</sup>. Huang *et al.*<sup>[11]</sup> proposed the DenseNet network, which deploys the concept of feature map reuse for small training datasets in supervised learning. DenseNet connects multiple dense blocks with a transition layer and concatenates the channels of each dense block feature map in series to increase the number of feature maps and improve their utilization rate.

In this study, FC\_DenseNet, a deep learning model based on DenseNet and fully convolutional networks, was employed. The model learns the planar distribution characteristics of the OARs in CT images through the dense block module to achieve end-to-end accurate OARs delineation for non-small cell lung cancer (NSCLC) patients.

## Materials and methods

### Data acquisition and preprocessing

CT images of 36 NSCLC patients from the Seventh Medical Center of the PLA General Hospital were acquired. The obtained four-dimensional-CT images were scanned using a Philips Brilliance Big Bore simulator (Philips Medical Systems, Madison, WI, USA) from the level of the larynx to the bottom of the lungs with a 3-mm slice thickness in the helical scan mode. This study was approved by the Ethics Committee of the Seventh Medical Center at the Chinese PLA General Hospital. All the patients provided written consent towards recording their medical information in the hospital database. By analyzing the DICOM file, the grey value of the original CT image was mapped to a range of 0–255, and the window width and level were set to 400 and 40, respectively. Different OARs were filled with different grayscale values to generate masked images as training labels, as shown in Fig. 1.

The training and validation datasets comprised 3803 CT images recorded from 27 patients and 650 images recorded from six patients, respectively. The testing set

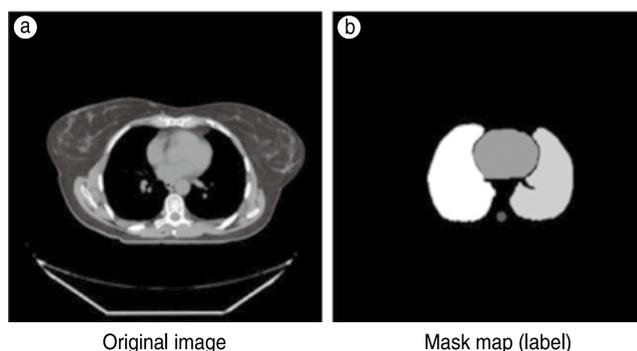


Fig. 1 Original image and mask map (label)

included 567 images recorded from nine patients. After data cleaning and enhancement, the images were sent to FC\_DenseNet. All the training, validation, and testing tasks were performed on an 11-GB NVIDIA GeForce GTX 1080 Ti GPU. The start and end times of the manual and auto-segmentation operations for each patient in the testing set were recorded.

### FC\_DenseNet model for segmentation

In this study, FC\_DenseNet was trained to auto-segment four types of OARs for diagnosing or monitoring NSCLC. The specific architecture of the model is illustrated in Fig. 2. The segmentation process was primarily divided into two parts: (a) The left half, called the analysis path, was composed of a density block module and transition down module connected by a short cut layer to extract image features. (b) The right half, called the synthesis path, was upsampled by the transition-up transposition convolution module to recover the size of the feature image layer. To improve the accuracy of the reconstructed image and accelerate the convergence of the network parameters, feature maps of the same size in the analysis path were connected in series as the input to the next layer of the

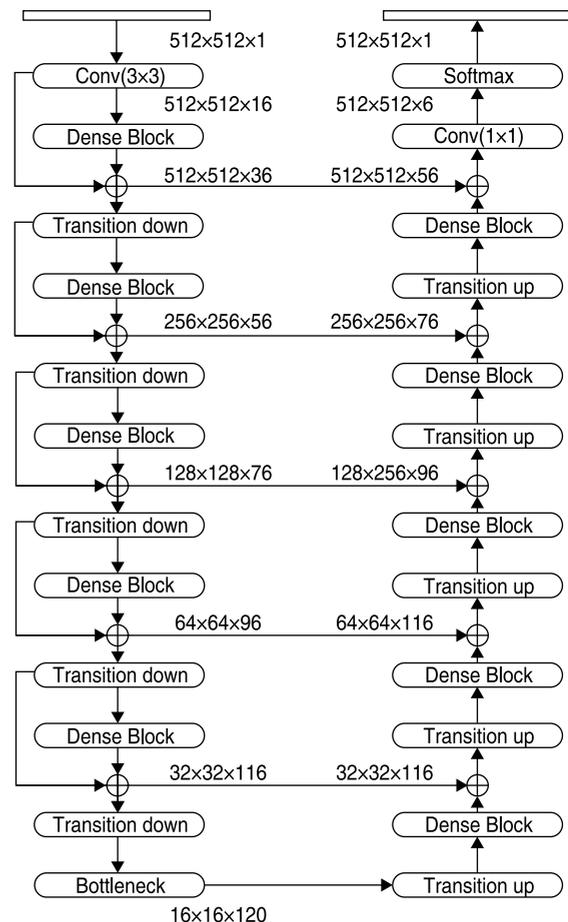
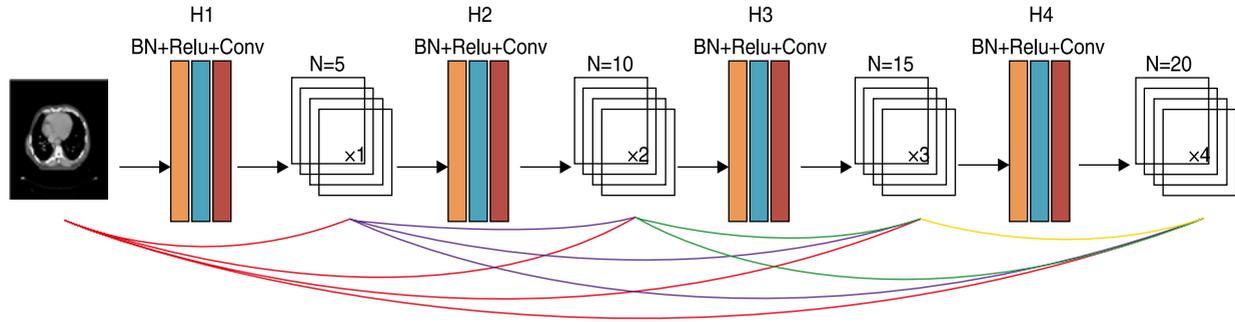


Fig. 2 Scheme of FC\_DenseNet



**Fig. 3** Scheme of Dense Block

density block.

The input to each layer of the dense block was composed of all the outputs of the front layer after a dense connection (as shown in Fig. 3). The output of each layer possessed the following corresponding functional relationship with the output of the other front layers:

$$X_{i+1} = H(X_1, X_2, \dots, X_i) \quad (1)$$

where  $H$  (\*) is a nonlinear function, denoting a series of operations, including batch normalization, ReLU activation, pooling, and convolution, that were used to extract features, adjust the size of the feature map, and reduce the channel dimension. A bottleneck architecture was set in each network, as the operation of dense connections could induce a surge in the number of channels and increase the training difficulty. The bottleneck architecture uses a  $1 \times 1$  convolution kernel to realize cross-channel feature fusion and enhance the feature extraction ability of the network.

**FC\_DenseNet training**

Following cleaning and augmentation, the data were sent to the FC\_DenseNet for training. The weight and bias of the network were updated using the cross-entropy loss function, as follows:

$$L_s = -\sum_{i=1}^k \left( y \log \hat{y} + (1-y) \log (1-\hat{y}) \right) \quad (2)$$

$$\hat{y} = (1 + e^{-w\hat{x}+b})^{-1} \quad (3)$$

where  $\chi$  is the input of the network,  $\hat{y}$  is the posterior probability output after network regression, and  $\kappa$  is the number of categories.

In this study, an early stop module was incorporated to detect the network accuracy and loss function value with an increase in the iterative epoch, and the network architecture of DensNet56 in the 30th epoch was selected. During the network training process, the initial learning rate was set to  $1e-3$ , which decreased with an increase in the epoch. This ensured that the network could converge quickly in the initial stage of training while preventing poor feature generalization arising from network overfitting. The average segmentation time of the training

set and that of a single  $512 \times 512$  CT image were set to approximately 12.58 min/epoch and 0.17s, respectively. Approximately 13.4 s were required to delineate all the CT images of a patient.

**Accuracy evaluation**

The auto-segmentation performance was evaluated based on geometric evaluation indices, including the Dice coefficient (DSC), 95% Hausdorff distance (HD95), and average surface distance (ASD). The OAR segmentation performance of the FC\_DenseNet network was compared with that of U-Net. Thereafter, two sets of RT plans were optimized with manually contoured OARs and automatically contoured OARs, respectively. Dosimetric parameters including Dmax (i.e., the dose received by 2% of the volume) and  $V_x$  (i.e., the volume receiving more than x Gy dose as a percentage of the total volume) were obtained and compared.

**Statistical analysis**

SPSS 20.0, a statistical software (version 20.0, SPSS Inc., Chicago, USA), was used for the Wilcoxon signed rank test, and the difference, at a significance level of  $\alpha = 0.05$ ,  $P < 0.05$ , was found to be statistically significant.

**Table 1** Comparison of Dice parameters for both networks ( $\bar{x} \pm s$ )

	Cord	Heart	Right lung	Left lung
Densenet	0.90 ± 0.02	0.84 ± 0.10	0.93 ± 0.06	0.97 ± 0.01
U-Net	0.87 ± 0.05	0.82 ± 0.12	0.92 ± 0.07	0.96 ± 0.02
P value	0.106	0.752	0.904	0.141

**Table 2** Comparison of HD95 parameters for both networks ( $\bar{x} \pm s$ )

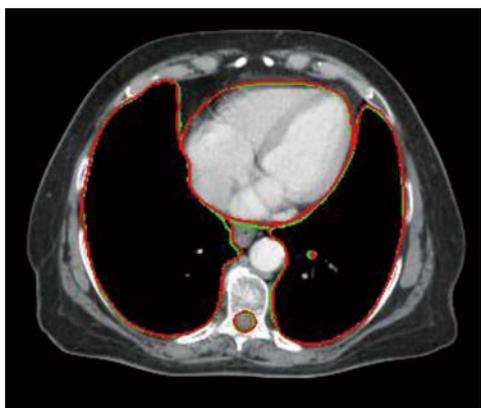
	Cord	Heart	Right lung	Left lung
Densenet	1.85 ± 0.36	15.95 ± 16.0	9.48 ± 5.50	6.97 ± 3.41
U-Net	2.42 ± 0.66	18.65 ± 15.2	13.2 ± 8.99	9.56 ± 4.62
P value	0.109	0.930	0.642	0.255

**Table 3** Comparison of ASD parameters for both networks ( $\bar{x} \pm s$ )

	Cord	Heart	Right lung	Left lung
Densenet	0.69 ± 0.13	6.98 ± 5.55	1.81 ± 1.61	1.11 ± 0.51
U-Net	0.86 ± 0.26	7.52 ± 4.65	2.55 ± 2.82	1.72 ± 0.60
P value	0.304	0.900	0.789	0.053

## Results

The geometric evaluation indices including DSC, HD95, ASD for the FC\_DenseNet and U-Net networks are listed in Tables 1–3, respectively. As can be inferred, DenseNet outperforms U-Net, although there is no significant difference between the two ( $P > 0.05$ ). A comparison of the results obtained for automatic and manual segmentation based on DenseNet for a typical patient is presented in Figure 4. The dose-volume parameters of the OARs based on manual and automatic segmentation are listed in Table 4. There are no statistically significant differences between the dosimetric parameters of the manually and automatically delineated OARs ( $P > 0.05$ ). The average time of manual segmentation by a radiation oncologist for nine patients in the testing set is 15.2 min. Whereas, auto-segmentation, which can be achieved with an average segmentation time of 9.0 min, significantly improves the delineation efficiency ( $P < 0.05$ ).



**Fig. 4** Schematic of automatic OAR segmentation using the DenseNet network (green and red lines represent manual and automatic segmentation contours, respectively)

**Table 4** Dosimetric comparison of the OARs between manual and DenseNet auto-segmentation.

Dosimetric parameters		Manual	AI	<i>P</i> value
Spinal Cord	Dmax	16.15±8.62	18.75±7.41	0.314
	V30	0.40±1.09	0.06±0.12	0.180
Heart	V40	0.19±0.56	0.03±0.07	0.180
	Mean (Gy)	1.71±1.58	1.33±1.17	0.110
	V5	27.55±6.81	28.19±6.78	0.515
Lung All	V10	15.49±4.41	14.83±4.54	0.953
	V20	9.40±3.69	9.44±3.89	0.859
	V30	6.57±3.25	6.64±3.36	0.263
	Mean (Gy)	6.49±1.94	6.56±2.01	0.173
MU		979.56±97.49	977.11±102.19	0.515

## Discussion

As can be observed from the results produced by both networks, it is easier to segment the lungs than it is to segment the spinal cord and heart. In the original CT image, there are obvious boundaries between the left and right lungs; further, the deep learning network makes it easier to extract the edge features. Compared to the lungs, although the spinal cord possesses a bone structure as a support and obvious texture and edge differentiation, it accounts for less area in the image. The negative samples in the background image significantly outnumber the positive samples in the spinal cord, thereby producing an imbalance that degrades the segmentation accuracy. The heart is at the center of the slice and is surrounded by other organs such as the larynx and esophagus. The image features at the center are not strong, which makes the segmentation result slightly worse than that of the lungs.

Compared to U-Net, the average DSC of FC\_DenseNet is slightly higher, and the variance is small, which indicates that the auto-segmentation performance of FC\_DenseNet is more stable, and the generalization of the model is better. The HD95 score is an index used to measure the maximum distortion of the segmentation results, and it is influenced by the number of outliers. FC\_DenseNet had better continuity and produced fewer outliers. The number of CT layers for each patient was not the same; therefore, the segmentation time for each patient was different. According to the evaluation benchmark<sup>[12]</sup> provided by the report on the thoracic organs auto-segmentation challenge organized by the American Association of Medical Physicists annual meeting in 2017, the organ with the highest DSC is the lung, with an average value between 0.95 and 0.98. The results of our study are relatively consistent with this value. Zhang *et al.*<sup>[13]</sup> developed a two-dimensional-AS-CNN based on the ResNet101 network using a dataset of 250 lung cancer patients. The average DSC scores for the left lung, heart, right lung, spinal cord, and esophagus were 0.94, 0.89, 0.94, 0.82, and 0.73, respectively. The DSC score of the spinal cord obtained by the proposed model was 0.89, which was significantly better than that of the AS-CNN. FC\_DenseNet, used in this study, is a lightweight model with a more concise architecture.

Owing to the difference in the training datasets, it was difficult to compare the advantages and disadvantages of the proposed method and those of the extant method. However, although significantly fewer training cases were used in this study, FC\_DenseNet exhibited a strong feature extraction ability in the training of small samples, and the segmentation results were similar to those of the training model with large datasets.

Zhu *et al.*<sup>[14]</sup> proposed an auto-segmentation model based on depth convolution to segment the CT images

of patients with lung cancer. In this model, a U-shaped network with a three-dimensional convolution kernel was used. The HD95 score was between 7.96 mm and 8.74 mm, and the ASD was between 1.81 mm and 2.92 mm. The resulting segmentation performance was significantly better than that of DenseNet. This may be because DenseNet, used in this study, is a two-dimensional model, and the extracted features are different owing to the poor continuity of the feature sequence in space.

Currently, there exist three primary development directions for deep learning networks in medical image segmentation. The first direction is to deepen the network level and depth, extract deeper semantic features to obtain a stronger expression ability, or widen the network to increase the number of channels to obtain additional information in the same layer such as the texture features of different frequencies and boundary features in different directions. The second direction is to achieve a more effective spatial feature extraction ability by learning the sequence association properties of multiple CT levels of a given case, represented by three-dimensional U-Net and several other derivative networks. The third direction, represented by DenseNet, is to improve the utilization rate of feature maps by sharing them layer by layer to enhance the feature expression ability of the image and improve the generalization performance of the network<sup>[15]</sup>.

Herein, the results demonstrated that FC\_DenseNet outperforms U-Net with regard to the segmentation of OARs; even when the training set contained fewer images, FC\_DenseNet still effectively prevented overfitting. Simultaneously, it prevented gradient disappearance during the training process by repeatedly using different levels of feature maps. Thus, this study provides a new approach for medical-image segmentation.

### Acknowledgments

Not applicable.

### Funding

Supported by a grant from the Beijing Municipal Science and Technology Commission Programme (No. Z181100001718011).

### Conflicts of interest

The authors indicated no potential conflicts of interest.

### Author contributions

Fuli Zhang and Qiusheng Wang contributed conception and design of the study. Fuli Zhang and Anning Yang trained the deep learning models, Anning Yang drafted the manuscript. Na Lu, Huayong Jiang, Diandian Chen, Yanjun Yu and Yadi Wang helped to collect the data and evaluate radiotherapy planning. All authors read, discussed and approved the final manuscript.

### Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

### Ethical approval

Not applicable.

### References

- Eaton BR, Pugh SL, Bradley JD, et al. Institutional Enrollment and Survival Among NSCLC Patients Receiving Chemoradiation: NRG Oncology Radiation Therapy Oncology Group (RTOG) 0617. *J Natl Cancer Inst.* 2016;108(9):djw034.
- Wang EH, Rutter CE, Corso CD, et al. Patients selected for definitive concurrent chemoradiation at high-volume facilities achieve improved survival in stage III non-small-cell lung cancer. *J Thorac Oncol.* 2015;10(6):937-943.
- Martin S, Johnson C, Brophy M, et al. Impact of target volume segmentation accuracy and variability on treatment planning for 4D-CT-based non-small cell lung cancer radiotherapy. *Acta Oncol.* 2015;54(3):322-332.
- Speight R, Sykes J, Lindsay R, et al. The evaluation of a deformable image registration segmentation technique for semi-automating internal target volume (ITV) production from 4DCT images of lung stereotactic body radiotherapy (SBRT) patients. *Radiother Oncol.* 2011;98(2):277-283.
- van Dam IE, van Sornsens de Koste JR, Hanna GG, et al. Improving target delineation on 4-dimensional CT scans in stage I NSCLC using a deformable registration tool. *Radiother Oncol.* 2010;96(1):67-72.
- Jameson MG, Holloway LC, Vial PJ, et al. A review of methods of analysis in contouring studies for radiation oncology. *J Med Imaging Radiat Oncol.* 2010;54(5):401-410.
- Brouwer CL, Steenbakkers RJ, van den Heuvel E, et al. 3D Variation in delineation of head and neck organs at risk. *Radiat Oncol.* 2012;7:32.
- Wong J, Fong A, McVicar N, et al. Comparing deep learning-based auto-segmentation of organs at risk and clinical target volumes to expert inter-observer variability in radiotherapy planning. *Radiother Oncol.* 2020;144:152-158.
- Wang Z, Chang Y, Peng Z, et al. Evaluation of deep learning-based auto-segmentation algorithms for delineating clinical target volume and organs at risk involving data for 125 cervical cancer patients. *J Appl Clin Med Phys.* 2020;21(12):272-279.
- 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.
- Huang G, Liu Z, Van Der Maaten L, et al. Densely Connected Convolutional Networks. *IEEE Conference on Computer Vision and Pattern Recognition*; 2017. Available at: <https://doi.org/10.48550/arXiv.1608.06993>
- Yang J, Veeraraghavan H, Armato SG 3rd, et al. Autosegmentation for thoracic radiation treatment planning: A grand challenge at AAPM 2017. *Med Phys.* 2018;45(10):4568-4581.
- Zhang T, Yang Y, Wang J, et al. Comparison between atlas and convolutional neural network based automatic segmentation of multiple organs at risk in non-small cell lung cancer. *Medicine (Baltimore).* 2020;99(34):e21800.
- Zhu J, Zhang J, Qiu B, et al. Comparison of the automatic

segmentation of multiple organs at risk in CT images of lung cancer between deep convolutional neural network-based and atlas-based techniques. *Acta Oncol.* 2019;58(2):257-264.

15. Ke L, Deng Y, Xia W, et al. Development of a self-constrained 3D DenseNet model in automatic detection and segmentation of nasopharyngeal carcinoma using magnetic resonance images. *Oral Oncol.* 2020;110:104862.

**DOI 10.1007/s10330-022-0553-3**

**Cite this article as:** Yang AN, Lu N, Jiang HY, et al. Automatic delineation of organs at risk (OARs) in NSCLC radiotherapy based on deep learning network. *Oncol Transl Med.* 2022;8(2):83–88.