Fully convolutional network-based multi-output model for automatic segmentation of organs at risk in thorax

Jie Zhang¹,²,³, Yiwei Yang¹,²,³, Kainan Shao¹,²,³, Xue Bai¹,²,³, Min Fang¹,⁴,⁵, Guoping Shan¹,²,³ and Ming Chen¹,⁴,⁵

¹Institute of Cancer and Medicine, Chinese Academy of Sciences, Hangzhou, China
²Department of Radiation Physics, Cancer Hospital of the University of Chinese Academy of Sciences, Hangzhou, China
³Department of Radiation Physics, Zhejiang Cancer Hospital, Hangzhou, China
⁴Department of Radiation Oncology, Cancer Hospital of the University of Chinese Academy of Sciences, Hangzhou, China
⁵Department of Radiation Oncology, Zhejiang Cancer Hospital, Hangzhou, China

Abstract

Purpose: To propose a multi-output fully convolutional network (MOFCN) to segment bilateral lung, heart and spinal cord in the planning thoracic computed tomography (CT) slices automatically and simultaneously.

Methods: The MOFCN includes two components: one main backbone and three branches. The main backbone extracts the features about lung, heart and spinal cord. The extracted features are transferred to three branches which correspond to three organs respectively. The longest branch to segment spinal cord is nine layers, including input and output layers. The MOFCN was evaluated on 19,277 CT slices from 966 patients with cancer in the thorax. In these slices, the organs at risk (OARs) were delineated and validated by experienced radiation oncologists, and served as ground truth for training and evaluation. The data from 61 randomly chosen patients were used for training and validation. The remaining 905 patients' slices were used for testing. The metric used to evaluate the similarity between the auto-segmented organs and their ground truth was Dice. Besides, we compared the MOFCN with other published models. To assess the distinct output design and the impact of layer number and dilated convolution, we compared MOFCN with a multi-label learning model and its variants. By analyzing the not good performances, we suggested possible solutions.

Corresponding author:
Ming Chen, Zhejiang Cancer Hospital, No. 1 Banshan East Road, Gongsu District, Hangzhou, Zhejiang 310000, China.
Email: chenming@zjcc.org.cn

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).
Results: MOFCN achieved Dice of $0.95 \pm 0.02$ for lung, $0.91 \pm 0.03$ for heart and $0.87 \pm 0.06$ for spinal cord. Compared to other models, MOFCN could achieve a comparable accuracy with the least time cost.

Conclusion: The results demonstrated the MOFCN’s effectiveness. It uses less parameters to delineate three OARs simultaneously and automatically, and thus shows a relatively low requirement for hardware and has potential for broad application.

Keywords
Fully convolutional network, multi-output architecture, organ at risk, automatic segmentation, thoracic radiotherapy

Introduction
An effective radiotherapy is to kill tumor using a high prescribed dose while sparing organs at risk (OARs).1 Delineating OARs in computed tomography (CT) scans is an essential step of optimizing a radiation treatment plan and evaluating clinical goals quantitatively. Several studies have shown that the delineation accuracy is highly correlated to tumor control and radiotherapy toxicities.2–4 Additionally, the inconsistencies arising from inter- and intra-observer variations have an impact on quantitative5–9 and dosimetric10–13 analysis. Therefore, the task of delineating OARs requires accuracy and less inter- and intra-observer variations.14–17

Although a lot of conventional image segmentation approaches have been used in the task of automatic or semi-automatic OARs delineation,18–20 such as those based on image grayscale information,21 statistical shape modeling22,23 and body atlas-based methods,24–26 they can’t guarantee the accuracy, consistency and reproducibility. The grayscale information-based algorithms are sensitive to imaging artifacts, low contrast-to-noise ratio (CNR), and so on. A segmentation based on a statistical shape model uses prior knowledge about a generic structure to constrain segmentation of individual organs. Such models may be challenged by any abnormalities. The body atlas-based methods are actually deformable image registrations referencing to an image with all segmented organs (i.e. the atlas). For them, the anatomical variation and vanishing volumes may be problematic. More importantly, those segmentations have no incorporation of the essential human clinical decision-making. The prior-knowledge and/or experience about what voxels are included as part of the segmented OAR can’t just come from image and usually has an inherent correlation to the treatment quality. Above all, a mankind’s work on the definition of OAR contours still serves as the gold criterion.21

As an artificial intelligence (AI) method of mimicking human beings’ activity, the fully convolutional network (FCN) is proposed to delineate OARs automatically.27–31 The first attempt of applying a convolutional neural network (CNN) on OARs’ segmentation was reported by Ibragimov and Xing.31 Their work focused on the OARs in head and neck region. Their input image was an image patch that was around voxels belonging to the interested OAR, instead of a whole CT scan. For the thoracic OARs, a 11-layer FCN was adopted to label the voxels of lung and achieved Dice of 0.96.28 Dice is the common similarity metric between two
images. According to Zhu et al., a 13-layer and 5-channel CNN delineated the spinal cord with a Dice of 0.71–0.79, and lung and heart with a Dice of 0.87–0.95 in a series of images with the size of 96 × 96 pixels. A 96 × 96 CT usually involves 106 mm × 106 mm image, it is unable to show a normal human’s whole transverse section. A large amount of methods used a single-input and single-output FCN with many layers to delineate one OAR in down-sampled or cropped CT scans. Therefore, they lack the appropriateness for clinical application and show a high computation requirement to delineate all OARs for a lot of patients. Corresponding to the high computation requirement, the high cost may limit their widespread application in most hospitals in China.

This paper proposes a multi-output FCN (MOFCN) and reports our experience of using it on a large testing set. Specifically, the contributions of this paper are as follows:

- We propose a network consisting of a shared main backbone and multiple distinct output branches. It segments bilateral lung, heart and spinal cord automatically and simultaneously in planning CTs that include the whole body’s transverse section. This design tries to combine the information sharing of various neighboring organs and independent feature extraction for individual one.
- The proposed model is evaluated on a large testing dataset. It is more like a clinical application scenario where a model is applied on a large data, but is trained on a limited quantity of images. The testing dataset includes abnormal cases, such as different body type (high/short, fat/thin), spongy bone and lung collapse. It is different to most public reports.
- Based on the above test results, we analyze for the failure causes and give possible solutions.

The remainder of this paper is organized as follows. Section 2 introduces the MOFCN architecture and experiment details. The experiment results are shown and compared with other models in section 3. The study on the design of MOFCN is also shown in section 3 by comparing with its variants. In section 4, we further discuss the results and the exceptive cases in which MOFCN fails and suggest the potential solutions. The study’s conclusion is presented in section 5.

**Materials and methods**

**Data acquisition**

The data (i.e. 19,277 CT scans totally) was collected from 966 patients with cancers in the thorax receiving radiotherapy from January 2015 to December 2018 in our department. All patients underwent CT scans on Light Speed (GE Healthcare, Chicago, America) or Brilliance CT Big Bore system (Philips Healthcare, Best, the Netherlands) on helical scan mode with or without contrast. All CT images were
reconstructed into matrix size of $512 \times 512$ with a thickness of 5 mm. The OARs were delineated in the planning CTs by experienced radiation oncologists and were regarded as the ground truth.

**Multi-output fully convolutional network (MOFCN) for segmentation**

The proposed MOFCN’s workflow and structure are shown in Figure 1.

**Image preprocessing.** To discard unnecessary air intensities, all images were cut into $512 \times 256$. Then their pixel intensities were linearly scaled from the tissue radiodensity range of $[-135\text{HU}, 215\text{HU}]$ to grayscale range of $[0, 255]$.

**Architecture of MOFCN.** In Figure 1(b), MOFCN’s convolution kernel size decreases with the increase of the network depth. It is based on the hypothesis that when the clinicians delineate OARs in the planning CTs, they first find the rough edges of OARs in their mind, and then determine their contour carefully. The big-size convolution kernel in MOFCN is used first to extract rough features from the big-size patches of a CT image. Then a small-size kernel is adopted to derive the subtle features.

Based on the above designing thought and limited by the computing hardware, the first network layer is a dilated convolutional one. The difference between a dilated and a conventional convolution is that they extract features from image patches with different sizes when using the same-size kernel. The dilated one’s image patch is larger than the conventional one. Each convolution layer contains an activation function of ReLU, except the last one. The function in the last convolution layer is sigmoid to produce a probability for a binary classification (i.e. background and foreground).

The MOFCN includes two components: one main backbone and three branches. The main backbone extracts the image features about 3 OARs. Then the features are transferred to three branches to generate three output images. This design is to ensure that the feature maps can involve the neighboring organs’ spatial information, because an organ’s location not only relates to its gray value, but also relates to its neighboring tissues. This point is discussed detailly in the last subsection of Discussion.

**Postprocessing for final prediction.** The three distinct outputs in MOFCN give three probability values ($p_X$) of a pixel belonging to the foreground ($X$). By comparing them, the final classification is determined according to the largest one. Next, a connected component labeling algorithm is used to correct discontinuities. Two/one largest connected components for lung/heart are kept in the final output.

If in some situations that $p_{\text{heart}}$ equals to $p_{\text{lung}}$, such as around the boundary of them, the pixel would be classified as both lung and heart. This is based on the two organs’ roles in radiotherapy. In radiation treatment, lung and heart are both dose volume-limiting organs. Their average dose and the percentages of volume receiving a certain dose attract our attention. Therefore, some pixels with dual identification have little impact on the clinical goals. This rule is also for the dual label (e.g.
Figure 1. Illustration of the: (a) workflow and (b) architecture of the proposed multi-output fully convolutional network (MOFCN). BN: batch normalization; conv: convolution; 2D: two-dimension.
a pixel, around the edge of lung and heart, is labeled as both) in ground truth, since the manual contour can’t reach pixel-level accuracy.

**Experiments**

The proposed MOFCN was implemented using Python (an open-source programming language for deep learning). The program ran on two Intel® E5-2630 processors (2.20 GHz) and four Graphics Cards (GeForce® GTX 1080, NVIDIA, Delaware, America).

In the collected 966 cases, the data from 61 randomly chosen patients were used for training and validation. Wherein, there were 856 CT scans (≈4% of all images) in the training set and 384 CT scans (≈2% of all images) in the validation set. Other 905 cases were used for testing. It encompassed 18,037 scans (≈94% of all images) for assessment. All images included 3 OARs. When splitting these cases, we labeled each case with a unique number from 1 to 966. Then we re-ordered the numbers using the function of random in Python. In the re-ordered sequence, the first 61 numbers corresponded to the identities (IDs) of cases in the training and validation sets. The remaining 905 numbers were the case IDs in the testing set.

The learning algorithm was Adaptive Moment Estimation (Adam).³⁷ The learning rate and epoch were $10^{-3}$ and 3000 respectively. The batch size equaled to 64. Therefore, in each iteration of one epoch, we transferred 64 CT images into the network for training. Then the network outputted three organs’ estimated segmentation images. They were compared with the ground truth for backpropagation. In one epoch, there were 14 ($\lceil \frac{856}{64} \rceil = 14$, $\lceil \cdot \rceil$ means the top integral function) iterations, because there were 856 training images. After each epoch, the loss of validation set was checked and only the model with the minimum loss would be saved.

**Loss function and performance evaluation**

The similarity between the auto-segmented OAR and its ground truth is scored using Dice similarity coefficient:

$$Dice_i = \frac{2|X \cap Y|}{|X \cup Y|}$$

in which $X$ is the model output, $Y$ is the ground truth, $i$ is the CT slice index. Dice ranges from 0 to 1. The similarity goes high with the increase of Dice.

The loss function $L$ is:

$$L = -\frac{1}{N} \sum_{i=1}^{N} (Dice_i^{\text{lung}} + Dice_i^{\text{heart}} + Dice_i^{\text{spinalcord}})$$

where $N$ is the number of CT slices, $i$ is the CT slice index. The superscript of “Dice” indicates its corresponding OAR.
We applied the proposed MOFCN on each patient’s case and assessed it using Dice in three dimensions (3D). For further analysis on under segmentation, we used the boxplot method to find outliers (i.e., those not good performances). Specifically, the \( k \)th testing case was identified as outlier, because its

\[
\text{Dice}_p^k < Q1 - 3 \times (Q3 - Q1)
\]

\( p \) indicates different organs. Q1 and Q3 are the first and third quartiles of testing results.

### Results

#### Testing results

The statistical results on the testing set are listed in Table 1. MOFCN achieves Dice of 0.95 ± 0.02 for lung, 0.91 ± 0.03 for heart and 0.87 ± 0.06 for spinal cord. The number of outlier cases are 1, 15, and 13 for lung, heart and spinal cord respectively.

Among the 23 outlier cases listed in Table 1, 4 cases (No. 114, 287, 363, and 710) show abnormality. Specially, No. 114 displays osteoplastic nodule. No. 287, 363, and 710 shows lung collapse. We infer that the abnormality is the reason of under segmentation in the four cases. The remaining 19 cases are further investigated in the Discussion section.

#### Comparison with other models

In this section, we provide a comparison between our network and other published models in the first subsection for an extensive evaluation. We compare the proposed model with its variants to study our design points: the dilated convolution (in the second subsection) and the distinct multiple outputs (in the third subsection). The effect of layer number on the segmentation performance is also shown in the second subsection.

---

**Table 1.** The segmentation performance of the proposed model on the testing set.

| Dice ± σ | Lung       | Heart      | Spinal cord |
|----------|------------|------------|-------------|
| \( \mu \) | 0.95 ± 0.02| 0.91 ± 0.03| 0.87 ± 0.06 |
| Q1       | 0.94       | 0.90       | 0.85        |
| Q3       | 0.96       | 0.93       | 0.90        |

Number of outliers

| Outlier case No. | 1  | 15           |
|------------------|----|--------------|
|                  | 287| 71, 75, 104, 111, 200, 222, 363, 439, 525, 587, 613, 698, 710, 761, 805 |

Outlier No. 111, 114, 176, 193, 200, 248, 363, 402, 417, 525, 688, 710, 761

- \( \mu \): average; \( \sigma \): standard deviation; Q1: the first quartile; Q3: the third quartile.
- Excluding the low-Dice cases caused by inter-observer variations.
Comparison with other published methods. In this subsection, we compare the MOFCN with other models, and list them in Table 2. The MOFCN was re-trained on the 36 training cases from 2017 AAPM Thoracic Auto-segmentation Challenges. All methods were assessed on the 12 online testing cases.

Table 2 illustrates that MOFCN achieves a comparable segmentation accuracy with other models, but the least time cost.

Comparison with MOFCN variants. In this subsection, we investigate the effect of the MOFCN’s layer number and dilated convolution on the segmentation performance. By removing different layers in MOFCN or replacing the dilated convolution with a conventional one, we constructed multiple variants. These variants were also re-trained on the 36 training cases and tested on the 12 online testing cases from the same challenge in the first comparison. The results are listed in Table 3.

Table 3 shows the trend of decreasing accuracy and time cost when reducing layer number. The layer numbers of 11, 10, and 9 don’t include removing layer in the lung branch. Hence, the lung Dice doesn’t show too much change when comparing with other two organs.

The first and last rows in Table 3 displays that the lung and heart Dice achieved by the no-dilation-convolution MOFCN are significantly lower than the proposed one. It may be caused by the size decrease of receptive filed because of no dilated convolution. As a result, there is no enough knowledge to help the MOFCN variant for classification and it needs more time for correcting discontinuities.

Comparison with multi-label learning-based single-output model. To evaluate the distinct-output design in this subsection, we compare the proposed MOFCN with a single-output model which is based on multi-label learning (abbr. MLFCN). The architecture of MLFCN is shown in Figure 2(a). It is totally same as the spinal cord branch of MOFCN to reduce the impact of network depth on segmentation. The activation function in the last layer of MLFCN is softmax.

Table 2. Segmentation performance of seven methods. The results are expressed as average ± standard deviation for Dice. Testing time is the average segmentation time for one patient.

| Approach        | Dice Lung    | Dice Heart   | Dice Spinal cord | Testing time |
|-----------------|--------------|--------------|------------------|--------------|
| Method 1        | 0.97 ± 0.02  | 0.93 ± 0.02  | 0.88 ± 0.04      | 30 s         |
| Method 2        | 0.98 ± 0.02  | 0.92 ± 0.02  | 0.89 ± 0.04      | 10 s         |
| Method 3        | 0.98 ± 0.02  | 0.91 ± 0.02  | 0.87 ± 0.11      | 6 min        |
| Method 4        | 0.96 ± 0.04  | 0.92 ± 0.02  | 0.85 ± 0.04      | 2 min        |
| Method 5        | 0.96 ± 0.03  | 0.85 ± 0.04  | 0.83 ± 0.08      | 2 min        |
| 3D Unet-3D Unet | 0.98 ± 0.01  | 0.93 ± 0.02  | 0.89 ± 0.04      | 1 min        |
| MOFCN           | 0.93 ± 0.03  | 0.90 ± 0.03  | 0.82 ± 0.07      | 3.16 s       |

*Method 1–5 are the deep learning methods reported in a published work and their lung Dices were obtained by averaging left and right lung Dices. MOFCN is our model.
Softmax produces four probability values of a pixel belonging to four labels (i.e. background, lung, heart and spinal cord) and the largest one gives the final classification. MLFCN was trained using the same data and optimization settings as MOFCN, but its loss function was categorical cross entropy. The MLFCN was also assessed on the same testing data. Its results are compared with MOFCN in Figure 2(b).

Figure 2(b) suggests that (a) MOFCN gets a better spinal cord segmentation than MLFCN and (b) MLFCN and MOFCN achieve equally well results when segmenting lung and heart. For (a), we infer that the enhancement is beneficial from the distinct output, because the design concentrates the last several layers’ learning capability on segmenting spinal cord, rather than distributing to other tasks. This is also why we employ multiple outputs. It is to combine the information sharing of various neighboring organs and independent feature extraction for individual one. For (b), it may be caused by the small contribution of a small-volume organ (e.g. spinal cord) on recognizing large-volume organs (e.g. lung and heart). Besides, even with deeper network in MLFCN, the delineation of lung and heart don’t show too much improvement.

Discussion

We analyze the testing result of spinal cord, further investigate the 19 outlier cases, and take spinal cord for example to explain why we adopt the structure of multiple outputs.

Table 3. Segmentation performance of MOFCN variants. The results are expressed as average ± standard deviation for Dice (the results with three decimal places are printed in parenthesis for meticulous comparison). Testing time is the average segmentation time for one patient.

| Layer number | Lung    | Heart   | Spinal cord | Testing time |
|--------------|---------|---------|-------------|--------------|
|              | (Dice)  | (Dice)  | (Dice)      |              |
| 12*          | 0.93 ± 0.03 | 0.90 ± 0.03 | 0.82 ± 0.07 | 3.16 s       |
|              | (0.932 ± 0.029) | (0.902 ± 0.031) | (0.824 ± 0.069) |              |
| 11§          | 0.93 ± 0.03 | 0.90 ± 0.03 | 0.83 ± 0.04 | 3.19 s       |
|              | (0.926 ± 0.034) | (0.898 ± 0.032) | (0.831 ± 0.041) |              |
| 10§          | 0.92 ± 0.05 | 0.89 ± 0.03 | 0.80 ± 0.06 | 3.15 s       |
|              | (0.915 ± 0.052) | (0.888 ± 0.031) | (0.795 ± 0.057) |              |
| 9§           | 0.92 ± 0.03 | 0.85 ± 0.05 | 0.79 ± 0.06 | 2.99 s       |
|              | (0.922 ± 0.031) | (0.852 ± 0.045) | (0.791 ± 0.055) |              |
| 8§           | 0.72 ± 0.09 | 0.73 ± 0.07 | 0.74 ± 0.07 | 2.77 s       |
|              | (0.718 ± 0.093) | (0.726 ± 0.074) | (0.737 ± 0.075) |              |
| 12 no dilation† | 0.70 ± 0.12 | 0.69 ± 0.11 | 0.87 ± 0.04 | 3.55 s       |
|              | (0.700 ± 0.116) | (0.690 ± 0.111) | (0.869 ± 0.037) |              |

*The MOFCN proposed in this work.
§The variants constructed by removing hidden convolution layers in two orders (bottom-up and spinal cord-heart-main backbone).
†The variant constructed by replacing the dilated convolution with a conventional one.
Table 1 suggests that the spinal cord, corresponding to the longest network branch in MOFCN, shows the lowest Dice among the three organs. We infer that the main reason is the inter-observer variations. Figure 3 gives some representative slices in which the spinal cord segmentation reaches a Dice of 0.48–0.63, but they are clinically accepted according to the recommended atlas of spinal cord.36

Detailed investigation of 19 exceptional cases

In this section, we give a detailed investigation on the above 19 exceptions. By reviewing each case, we categorize the under-segmentation into misrecognition and missed recognition (as shown in Table 4). The former one means that the non-
object pixels are recognized as part of object. The latter one represents that the object pixels are not recognized as part of object.

Analysis on the failure cause of misrecognition. For a CNN, the misrecognition directly results from the similar features of the non-object pixel, compared to an object one.

Table 4. Under-segmentation category for 19 outlier testing cases.

| Category                   | Under segmentation performance | Outlier case No. |
|----------------------------|--------------------------------|-----------------|
| Misrecognition             | Heart                          | 75, 104, 439    |
| Missed recognition         | Heart                          | 71, 222, 587, 613, 698, 805 |
|                            | Spinal cord                    | 176, 193, 248, 402, 417, 688 |
|                            | Heart and spinal cord          | 111, 200, 525, 761 |

Figure 3. Illustration of spinal cord segmentation with a Dice of <0.7, but clinically accepted (a-f) are 6 such cases Denotation: No., patient number; s., slice number.
Figure 4 illustrates the misrecognition images of No. 75, 104, and 439. Compared to the reference scan in Figure 4(d), there are non-zero neighboring pixels around heart (indicated by the yellow thick arrows in Figure 4) in the three failure cases. Accordingly, we infer that the misrecognition reason is the lack or less of such samples in training data, hence the model didn’t learn how to distinguish heart from its uncommon neighbors.

Analysis on the failure cause of missed recognition. The segmentation task of a CNN is a pixel-wise classification task. The classification is only based on the knowledge of the pixel’s reception field. Thus, any variations in the reception field, such as geometric and grayscale variations, are high likely to cause an under segmentation. To validate our conjecture on the missed recognition causes, we conducted the geometric or grayscale adjustments on the original input images for improving performance and display them in Figure 5.

Figure 5(a–c) show that most cases receive improvement, when we adjust the input image according to the reference (as shown in Figure 5(d)). The adjustments
include: (a) resizing the body to have the same size; (b) rotating the body until its left-right direction approximates horizontal and (c) rescaling the mapping between tissue radiodensity and gray value to increase the number of 255-grayvalue pixels around the spinal cord.

Figure 5. Bar graph of segmentation accuracy before and after adjustment for the missed recognition cases. (a–c) relate to the heart, spinal cord (abbr. sp) and both respectively. (d) is the adjustment reference (No. 10 slice for No. 1 patient). The body size = 300 × 207. The left-right direction = horizon. The gray value (v) around the spinal cord ≈ 255.

Zhang et al.
Future work. Above all, there are two factors impacting the proposed MOFCN. They are (a) special radiodensity of tissues and (b) geometric change, such as body size change and rotation. The factor (a) may be caused by the spongy bone, pulmonary atelectasis and other special cases that are not or less included in the training data.

In our future work, to overcome factor (a-b), we would include more special radiodensity and geometry samples in training to improve the model’s anti-interference performance or adopt an adaptive grayscale transformation. 3D models are also potential in leading to better performances. By expanding the convolution kernel to 3D and hence involving inter-slice information, it is likely to learn more about the comprehensively spatial relationship among different voxels and survive from multiple disturbance. To consider both performance and cost for a widespread clinical application, we would investigate the correlation of segmentation improvement to various number of 3D model parameters. Further modification would be a tradeoff between performance and efficiency (e.g. cost on hardware, running time).

Another possible solution for factor (b) is to develop a geometric correction in the image preprocessing to guarantee desirable segmentations.

The three OARs segmented by our model are the dose-limited organs for conventional radiation treatment. For stereotactic body radiotherapy (SBRT), more than three organs are involved in concerns. In such scenario, we can attach other more branches to the proposed MOFCN to build a modified one, and use its parameters as part of the initial values to train the modified network. The pre-trained parameters may accelerate the optimization.

Auto-segmentation of spinal cord using a multi-output structure

There are two advantages of MOFCN on auto-segmenting spinal cord: (a) one is to use a relatively large kernel at the first convolutional layer to involve neighboring organs’ information about spinal cord; (b) the second one is to use the multi-output network structure to avoid over-fitting.

In the thorax, the image features about the spinal cord are supposed to include two points: (i) it appears in gray and is surrounded by a white annular region and (ii) it lies between two large-area black connected regions. The white annular region is the spinal canal and shows white in a CT. The two large-area black connected regions are the left and right lungs. Between them lies the spinal canal.

Without a relatively large-size convolutional kernel, it is hard to distinguish the spinal cord and the humerus marrow. As shown in Figure 6, \( r_A \) and \( r_B \) encompass no lung pixels and look highly similar. They are both in gray and are surrounded by a white annular region. Only based on \( r_A \) and \( r_B \), it is hard for a CNN to learn the above point (ii) and hence a larger reception filed is needed. But with a large-size convolutional kernel and a single-output network, the optimization is prone to over-fitting. This can be avoided by a multi-output structure, because of sharing parameters in the main backbone. The shared parameters not only relate to the
spinal cord, but also relate to the lung and heart. Additionally, the shared feature maps involve the spatial knowledge of lung. It contributes to learning the above point (ii) and accelerating the optimization. In a supervised learning of a single-output network structure, the label image only involves the location of spinal cord. During optimization, it lacks the guidance information about where the lung is.

**Conclusion**

In this article, we propose a MOFCN to segment bilateral lung, heart and spinal cord in the thorax automatically and simultaneously. According to the evaluation results in 18,037 CT images and comparison with other models, the MOFCN is proved effective. Its multi-output structure can perform multiple object recognition and segmentation, which shows a relatively low requirement for computing and thus has potential for broad application. We also suggest potential solutions to further improve its performance on some “unseen” cases, including geometric change, adaptive grayscale transformation and so on.

**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.

**Funding**

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by National Key Research and Development Program (2017YFC0113201), Zhejiang Provincial Natural Science Foundation of China (LQ20H180016, LQ17H180003), Zhejiang Key Research and Development Program (2019C03003), Youth Talent Foundation of Zhejiang Medical and Health Project (2019RC023), Appropriate Technology Cultivation and Promotion of
Zhejiang Medical and Health Project (2019ZH018), Postdoctoral Program of Zhejiang Province, Chinese Postdoctoral Fund (520000-X91601) and National Natural Science Foundation of China (81230031/H18, 82001928).

**Ethics statement**

There is no human or animal participants involved in this study. All computed tomography images in our experiment were obtained for the conventional radiation treatment. We just collected and used them to train and validate the proposed model.

**ORCID iDs**

Jie Zhang https://orcid.org/0000-0002-0817-0464
Yiwei Yang https://orcid.org/0000-0001-9930-0561
Xue Bai https://orcid.org/0000-0002-9056-0882

**References**

1. De Ruysscher D, Niedermann G, Burnet NG, et al. Radiotherapy toxicity. *Nat Rev Dis Primers* 2019; 5(1): 13.
2. Mukesh M, Benson R, Jena R, et al. Interobserver variation in clinical target volume and organs at risk segmentation in post-parotidectomy radiotherapy: can segmentation protocols help? *Br J Radiol* 2012; 85(1016): e530–e536.
3. Walker GV, Awan M, Tao R, et al. Prospective randomized double-blind study of atlas-based organ-at-risk auto-segmentation-assisted radiation planning in head and neck cancer. *Radiother Oncol* 2014; 112(3): 321–325.
4. Gregoire V, Mackie T and Neve W. Prescribing, recording, and reporting photon-beam intensity-modulated radiation therapy (IMRT). *J ICRU* 2010; 10(1): 1–112.
5. Owens CA, Peterson CB, Tang C, et al. Lung tumor segmentation methods: impact on the uncertainty of radiomics features for non-small cell lung cancer. *PLoS One* 2018; 13(10): e0205003.
6. Parmar C, Velazquez ER, Leijenaar R, et al. Robust radiomics feature quantification using semiautomatic volumetric segmentation. *PLoS One* 2014; 9(7): e102107.
7. Balagurunathan Y, Gu Y, Wang H, et al. Reproducibility and prognosis of quantitative features extracted from CT images. *Transl Oncol* 2014; 7(1): 72–87.
8. Lee M, Woo B, Kuo MD, et al. Quality of radiomic features in glioblastoma multiforme: impact of semi-automated tumor segmentation software. *Korean J Radiol* 2017; 18(3): 498–509.
9. Kalpathy-Cramer J, Mamomov A, Zhao B, et al. Radiomics of lung nodules: a multi-institutional study of robustness and agreement of quantitative imaging features. *Tomography* 2016; 2(4): 430–437.
10. Hong TS, Tomé WA and Harari PM. Heterogeneity in head and neck IMRT target design and clinical practice. *Radiother Oncol* 2012; 103(1): 92–98.
11. Rasch C, Steenbakkers R and van Herk M. Target definition in prostate, head, and neck. *Semin Radiat Oncol* 2005; 15(3): 136–145.
12. Weiss E and Hess CF. The impact of gross tumor volume (GTV) and clinical target volume (CTV) definition on the total accuracy in radiotherapy. *Strahlenther Onkol* 2003; 179(1): 21–30.
13. Saarnak AE, Boersma M, van Bunningen BN, et al. Inter-observer variation in delineation of bladder and rectum contours for brachytherapy of cervical cancer. *Radiother Oncol* 2000; 56(1): 37–42.

14. Geets X, Daisne J-F, Arcangeli S, et al. Inter-observer variability in the delineation of pharyngo-laryngeal tumor, parotid glands and cervical spinal cord: comparison between CT-scan and MRI. *Radiother Oncol* 2005; 77(1): 25–31.

15. Kepka L, Bujko K, Garmol D, et al. Delineation variation of lymph node stations for treatment planning in lung cancer radiotherapy. *Radiother Oncol* 2007; 85(3): 450–455.

16. Jensen NK, Mulder D, Lock M, et al. Dynamic contrast enhanced CT aiding gross tumor volume delineation of liver tumors: an interobserver variability study. *Radiother Oncol* 2014; 111(1): 153–157.

17. Steenbergen P, Haustermans K, Lerut E, et al. Prostate tumor delineation using multiparametric magnetic resonance imaging: inter-observer variability and pathology validation. *Radiother Oncol* 2015; 115(2): 186–190.

18. Fritscher KD, Peroni M, Zaffino P, et al. Automatic segmentation of head and neck CT images for radiotherapy treatment planning using multiple atlases, statistical appearance models, and geodesic active contours. *Med Phys* 2014; 41(5): 51910.

19. Wachinger C, Fritscher K, Sharp G, et al. Contour-driven atlas-based segmentation. *IEEE Trans Med Imaging* 2015; 34(12): 2492–2505.

20. Gomathi V and Karthikeyan S. A Proposed Hybrid Medoid Shift with K-means (HMSK) segmentation algorithm to detect tumor and organs for effective radiotherapy. In: Prasath R and Kathirvalavakumar T (eds) *Mining intelligence and knowledge exploration*. Vol. 8284. Cham: Springer, 2013, pp.139–147.

21. Whitfield GA, Price P, Price GJ, et al. Automated delineation of radiotherapy volumes: are we going in the right direction? *Br J Radiol* 2013; 86(1021): 20110718–20110718.

22. Shahedi M, Halicek M, Guo R, et al. A semiautomatic segmentation method for prostate in CT images using local texture classification and statistical shape modeling. *Med Phys* 2018; 45(6): 2527–2541.

23. Pekar V, McNutt TR and Kaus MR. Automated model-based organ delineation for radiotherapy planning in prostatic region. *Int J Radiat Oncol Biol Phys* 2004; 60(3): 973–980.

24. Ciardo D, Gerardi MA, Vigorito S, et al. Atlas-based segmentation in breast cancer radiotherapy: evaluation of specific and generic-purpose atlases. *Breast* 2017; 32: 44–52.

25. Van Elmpt W, Van der Stoep J, Van Soest J, et al. EP-1225: atlas-based segmentation reduces interobserver variation and delineation time for OAR in NSCLC. *Radiother Oncol* 2017; 123: S661.

26. Acosta O, Mylona E, Le Dain M, et al. Multi-atlas-based segmentation of prostatic urethra from planning CT imaging to quantify dose distribution in prostate cancer radiotherapy. *Radiother Oncol* 2017; 125(3): 492–499.

27. Men K, Dai J and 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.

28. Mao K-p, Xie S-p and Shao W-z. Automatic segmentation of thorax CT images with fully convolutional networks. In: Wang SX (ed.) *Current trends in computer science and mechanical automation*. Vol. 1. De Gruyter Oldenbourg, Germany: De Gruyter Open, 2017, pp.402–412.
29. 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.
30. Humpire-Mamani GE, Setio AAA, van Ginneken B, et al. Efficient organ localization using multi-label convolutional neural networks in thorax-abdomen CT scans. *Phys Med Biol* 2018; 63(8): 0850031–0850012.
31. Ibragimov B and Xing L. Segmentation of organs-at-risks in head and neck CT images using convolutional neural networks. *Med Phys* 2017; 44(2): 547–557.
32. Yang J, Veeraraghavan H, Armato SG III, et al. Autosegmentation for thoracic radiation treatment planning: a grand challenge at AAPM 2017. *Med Phys* 2018; 45(10): 4568–4581.
33. Yu F and Koltun V. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:151107122*, 2015.
34. Gu J, Wang Z, Kuen J, et al. Recent advances in convolutional neural networks. *Pattern Recognit* 2018; 77: 354–377.
35. Wu K, Otoo E, Shoshani A, et al. Optimizing connected component labeling algorithms. *Paper presented at: Medical Imaging: Image Processing, San Diego, CA, 2005.*
36. Kong FM, Ritter T, Quint DJ, et al. Consideration of dose limits for organs at risk of thoracic radiotherapy: atlas for lung, proximal bronchial tree, esophagus, spinal cord, ribs, and brachial plexus. *Int J Radiat Oncol Biol Phys* 2011; 81(5): 1442–1457.
37. Kingma DP and Ba J. Adam: a method for stochastic optimization. *arXiv preprint arXiv:14126980*, 2014.
38. Tukey JW. *Exploratory data analysis*. Vol. 2. Reading, MA: Addison-Wesley, 1977.
39. Feng X, Qing K, Tustison NJ, et al. Deep convolutional neural network for segmentation of thoracic organs-at-risk using cropped 3D images. *Med Phys* 2019; 46(5): 2169–2180.
40. Wollmann T, Gunkel M, Chung I, et al. GRUU-Net: integrated convolutional and gated recurrent neural network for cell segmentation. *Med Image Anal* 2019; 56: 68–79.

**Author biographies**

**Jie Zhang** received the B.S. degree in communication engineering and Ph.D. degree in electronic science and technology from Nanjing University. Since 2018, she has been a radiation physicist with radiation physics department, Zhejiang Cancer Hospital (Cancer Hospital of the University of Chinese Academy of Sciences). She is the author of more than 10 articles and 2 inventions. Her research interests include tumor motion tracking and the medical application of artificial intelligence.

**Yiwei Yang** received the B.S. and M.S. degrees in biomedical engineering from the Southern Medical University in 2010 and 2013, respectively. Since 2013, he has been a radiation physicist with radiation physics department, Zhejiang Cancer Hospital (Cancer Hospital of the University of Chinese Academy of Sciences). From 2014 to 2016, he was an engineer assistant. Since 2017, he has been an engineer. He is the author of more than 10 articles and 3 inventions. His research interests include medical physics, radiation dosimetry and artificial intelligence in radiotherapy.

**Kainan Shao** received the B.S. in physics from Zhejiang University in 2007 and the Ph.D. degree in theoretical physics from Zhejiang University in 2012. From 2012 to 2014, he was an engineer with the Wireless Network Laboratory, Huawei Technology Company,
Shanghai, China. Since 2014, he has been a medical physicist with the radiation physics department, Zhejiang Cancer Hospital (Cancer Hospital of the University of Chinese Academy of Sciences). He is the author of 10 articles. His research interests include treatment planning and quality assurance of linear accelerator in radiation oncology. He won the 11th place in Radiation Knowledge Plan Competition 2018.

Xue Bai received the B.S. degree in physics from the Zhejiang University, in 2008 and M.S. degree in nuclear energy and nuclear technology engineering from the Sichuan University, in 2017. From 2008 to 2013, she was an engineer assistant with the department of radiation physics in Zhejiang Cancer Hospital (Cancer Hospital of the University of Chinese Academy of Sciences). Since 2014, she has been an engineer with the same department. She is the author of more than 10 articles and more than 5 inventions. Her research interest includes the dosimetry study in radiotherapy, medical physics, the deep learning used in radiotherapy, and automatic planning.

Min Fang received the M.D. degree in clinical oncology from Wuhan University. Since 2015, she has been a clinical oncologist in the department of radiation oncology, Zhejiang Cancer Hospital (Cancer Hospital of the University of Chinese Academy of Sciences). She has published more than 10 papers and obtained 1 granted patent. Her research interests include artificial intelligence and the application in tumor microenvironment.

Guoping Shan received the B.S. degree in technical physics from the University of South China, in 1997 and M.S. degree in Biomedical Engineering from the Tsinghua University, in 2015. From 1997 to 2012, he was a senior staff radiation oncology physicist with the department of radiotherapy in Taizhou Hospital. From 2008 to 2009, he was also a radiation oncology physics resident with the department of radiation oncology in Fox Chase Cancer Center, Philadelphia, PA. From 2013 to 2014, he worked as the chief physicist in Mali National Radiotherapy Center, Bamako, Mali in Africa. Since 2012, he has been the director of radiation physics department in Zhejiang Cancer Hospital (Cancer Hospital of the University of Chinese Academy of Sciences). He has published more than 30 articles. His main research direction is medical physics. His awards and honors include Mali “bee” National Medal (Mali Africa 2014) and the Best Medical Workers (Zhejiang Province 2015).

Ming Chen received his bachelor degree in Medical College of Soochow University, the master degree in Shanghai Medical College of Fudan University and M.D. degree in Wuhan University. From 2004 to 2005, he was a postdoc at University of Michigan. He has been a professor at Sun Yat-Sen University, Zhejiang University and Soochow University. From 2012 to 2021, he is with Zhejiang Cancer Hospital (Cancer Hospital of Chinese Academy of Sciences), and from 2019 to 2021, he was the executive vice president of the hospital. Since 2021, he is the director of Medical Physics Department at Sun Yat-Sen University Cancer Center. He is the Standing Committee Member and General Secretary of Chinese Society of Radiation Oncology, Chinese Medical Association, the Standing Committee Member of the Society of Radiation Therapy, China Anti-Cancer Association, and the Vice Chairman of the Experts Committee on Radiation Therapy, Chinese Society of Clinical Oncology. He has been a principal investigator of over 20 national or ministry projects, and editors of 9 academic journals and 11 professional associations. He has published over 100 papers and was licensed 11 patents. His research interest is radiation oncology.