


Clinical Target Volume Auto-Segmentation of Esophageal Cancer for Radiotherapy After Radical Surgery Based on Deep Learning

Technology in Cancer Research & Treatment
Volume 20: 1-11
© The Author(s) 2021
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/15330338211034284
journals.sagepub.com/home/tct


Ruifen Cao, PhD^{1,2} , Xi Pei, PhD³, Ning Ge, MD⁴, and Chunhou Zheng, PhD^{1,2}

Abstract

Radiotherapy plays an important role in controlling the local recurrence of esophageal cancer after radical surgery. Segmentation of the clinical target volume is a key step in radiotherapy treatment planning, but it is time-consuming and operator-dependent. This paper introduces a deep dilated convolutional U-network to achieve fast and accurate clinical target volume auto-segmentation of esophageal cancer after radical surgery. The deep dilated convolutional U-network, which integrates the advantages of dilated convolution and the U-network, is an end-to-end architecture that enables rapid training and testing. A dilated convolution module for extracting multiscale context features containing the original information on fine texture and boundaries is integrated into the U-network architecture to avoid information loss due to down-sampling and improve the segmentation accuracy. In addition, batch normalization is added to the deep dilated convolutional U-network for fast and stable convergence. In the present study, the training and validation loss tended to be stable after 40 training epochs. This deep dilated convolutional U-network model was able to segment the clinical target volume with an overall mean Dice similarity coefficient of 86.7% and a respective 95% Hausdorff distance of 37.4 mm, indicating reasonable volume overlap of the auto-segmented and manual contours. The mean Cohen kappa coefficient was 0.863, indicating that the deep dilated convolutional U-network was robust. Comparisons with the U-network and attention U-network showed that the overall performance of the deep dilated convolutional U-network was best for the Dice similarity coefficient, 95% Hausdorff distance, and Cohen kappa coefficient. The test time for segmentation of the clinical target volume was approximately 25 seconds per patient. This deep dilated convolutional U-network could be applied in the clinical setting to save time in delineation and improve the consistency of contouring.

Keywords

segmentation, dilated convolution, esophageal cancer, clinical target volume, radiotherapy

Abbreviations

BN, batch normalization; CNN, convolutional neural network; CT, computed tomography; CTV, clinical target volume; DDU-net, deep dilated convolutional U-network; DSC, Dice similarity coefficient; EC, esophageal cancer; GT, standard ground truth; HD, Hausdorff distance; KAP, Cohen kappa coefficient; OAR, organ at risk; U-Net, U-network.

Received: September 15, 2020; Revised: April 26, 2021; Accepted: June 11, 2021.

¹ College of Computer Science and Technology, Anhui University, Hefei, Anhui, China

² Engineering Research Center of Big Data Application in Private Health Medicine, Fujian Province University, Putian, Fujian, China

³ University of Science and Technology of China, Hefei, Anhui, China

⁴ The First Affiliated Hospital of USTC West District, Anhui Provincial Cancer Hospital, Hefei, Anhui, China

Corresponding Author:

Ruifen Cao, PhD, College of Computer Science and Technology, Anhui University, 111 Jiulong Road, Hefei, Anhui 230601, China.

Email: rfcao@ahu.edu.cn



Introduction

Esophageal cancer (EC) is a very aggressive malignant tumor, and its incidence rate is increasing worldwide, especially in China.¹ At present, the 5-year survival rate is only 15% to 25%.^{2,3} Surgical resection is the first-choice treatment for EC,^{4,5} but the recurrence rate after radical resection is still high. Local recurrence is the main cause of treatment failure,^{6,7} and postoperative radiotherapy is the main treatment method used to control local recurrence and prolong survival.^{8,9} The most critical step of radiotherapy planning is to define and segment the clinical target volume (CTV) and organs at risk (OARs).^{10,11} This task is usually carried out manually by radiation oncologists based on recommended guidelines using a treatment planning system. However, the manual segmentation process is time-consuming and operator-dependent. The accuracy of segmentation is highly dependent on the knowledge, experience, and preferences of radiation oncologists.^{12,13}

Previous studies have indicated that a fully automatic segmentation method for radiotherapy is helpful to relieve radiation oncologists from the labor-intensive aspects of their work and increase the accuracy, consistency, and reproducibility of region-of-interest delineation. Generally, automatic image segmentation approaches can be classified into three types based on region, edge, and classification.¹⁴ The features can be extracted from the intensity, gradient, and texture. However, it is still difficult to accurately segment the regions of interest on computed tomography (CT) images with boundary insufficiencies based on gray-level information because of the low contrast-to-noise ratio and high-density artifacts. Atlas-based segmentation, which incorporates prior knowledge into the process of segmentation, is one of the most commonly used image segmentation techniques in clinical software.¹⁵⁻¹⁷ However, the tumor target may vary greatly according to the patient's body shape and size and the cancer type and state, making it difficult to build a "universal atlas" for the tumor target. In addition, it is time-consuming because of the deformable registration process.

In the last few years, a quantum leap has been made in deep learning because of advancements in many areas. One particular area is the progression of convolutional neural network (CNN) architectures for image classification and segmentation.¹⁸⁻²² Interest in applying CNNs to radiotherapy has increased. The first work on OAR delineation with CNNs in radiotherapy was reported by Ibragimov and Xing,²³ who used CNNs for OAR segmentation in head and neck CT images. Tong *et al*²⁴ developed a CNN-based method for multi-organ segmentation in head and neck cancer radiotherapy. Men *et al*^{25,26} developed a tumor volume segmentation technique for rectal cancer, nasopharyngeal carcinoma, and breast cancer; Li *et al*²⁷ focused on tumor target segmentation of nasopharyngeal cancer in CT images based on deep learning methods; and Zhang *et al*²⁸ focused on gross target volume automatic segmentation in non-small cell lung cancer using a modified version of ResNet. Automatic segmentation of the tumor target based on deep learning has become a

research hotspot,²⁹ but few studies have explored the role of deep learning in auto-segmentation of the CTV of EC based on CT images as well as the efficiency of auto-segmentation in end-to-end clinical application. In addition, because the esophageal tumor is resected, the obvious tumor is no longer present on patients' postoperative images, increasing the difficulty of CTV delineation based on deep learning.¹⁰

The present study provides 4 novel contributions. First, to the best of our knowledge, there are no previous reports on CTV auto-segmentation in planning CT images for patients with EC after radical surgery. Thus, we developed a deep learning model to segment the CTV of radiotherapy after radical surgery. Second, a deep dilated convolutional module in the deep dilated convolutional U-network (DDU-net) was introduced to extract original context information directly from input images and compensate contextual features into U-network (U-Net) encode layers, and the results showed that better segmentation performance and good convergence were achieved using our DDU-net method. Third, batch normalization (BN) was added to the DDU-net and original U-Net, and the loss function based on the Dice similarity coefficient (DSC) was adopted for fast and stable convergence in the training of DDU-net and U-Net. Finally, our study results showed that the overall performance of the DDU-net was better than that of the U-Net, modified U-Net, and attention U-Net. Consequently, the DDU-net could rapidly delineate the CTV with high accuracy.

Materials and Methods

No ethical approval was required or obtained because this study only involved the evaluation of processed images; it did not involve patient information, a prospective evaluation, human body experimentation, or highly invasive procedures. All patients provided written informed consent prior to enrollment in the study.

In this study, we introduced a deep learning model to realize automatic CTV segmentation in radiotherapy for patients with EC after radical surgery. Figure 1 is a flowchart of the study, which was an end-to-end segmentation framework that could predict pixelwise class labels in CT images. The training dataset was used to optimize the parameters of the deep learning model to achieve good CTV segmentation for radiotherapy. The testing dataset was used to assess the performance of the model.

Data Acquisition and Preprocessing

Ninety-one patients diagnosed with stage I or II upper and middle EC from January 2015 to December 2019 at Anhui Provincial Cancer Hospital were included in our study. All patients received radiotherapy after surgery. The upper bound of the CTV was the inferior edge of the cricoid cartilage, and the lower bound was 3 cm below the tracheal eminence, including the esophageal tumor bed, anastomotic stoma, and lymph nodes in regions 2, 4, 5, and 7 of the chest. In some patients,

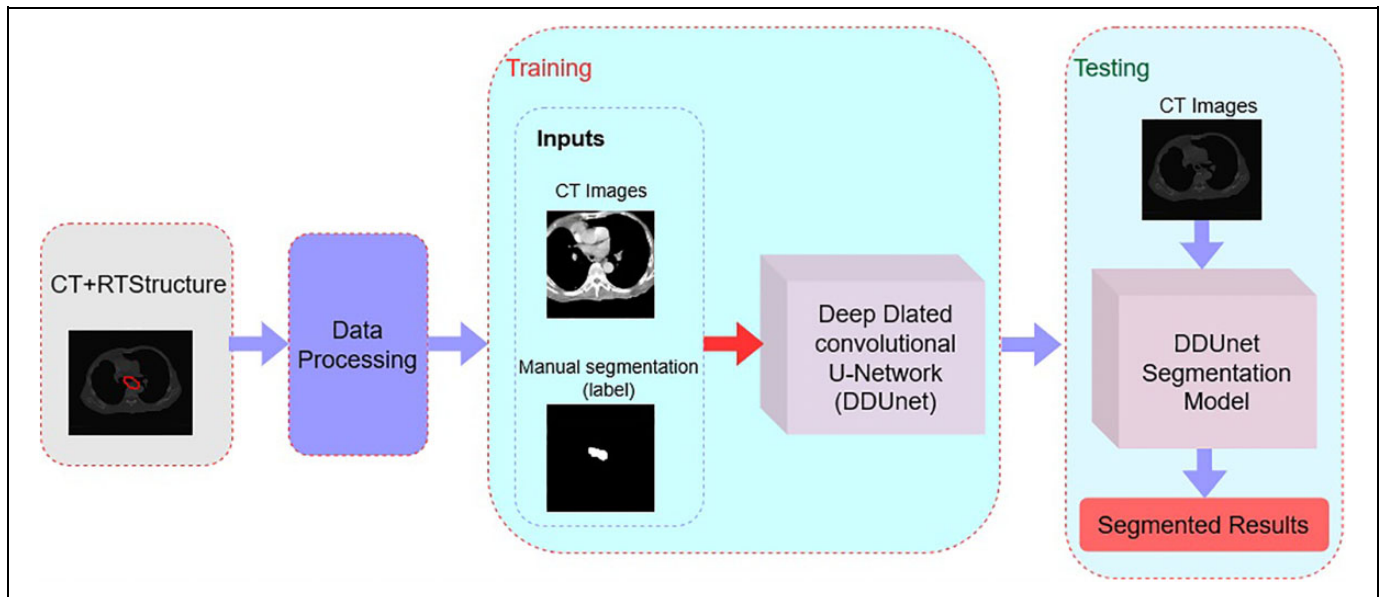


Figure 1. Flowchart of CTV segmentation based on deep learning.

however, the lower boundary of the CTV moved further downward according to the location of the resected tumor. Only patients with a CTV within the scope mentioned above were included.

All patients were immobilized with a vacuum cushion and thermoplastic mask for the neck and shoulders in the supine position. CT data were acquired on a Somatom Definition AS 40-slice CT system (Siemens Healthineers, Erlangen, Germany) or Brilliance CT Big Bore system (Philips Healthcare, Best, the Netherlands) in helical scan mode with contrast enhancement. CT images were reconstructed using a matrix size of 512×512 and thickness of 2.5 mm. Radiation oncologists contoured the CTV and OARs in the planning CT scan using a Pinnacle treatment planning system (Philips Radiation Oncology Systems, Fitchburg, WI, USA) system. Each CTV contour that was used as “standard ground truth (GT)” was delineated by an experienced oncologist, modified and verified by a senior radiation oncologist, and finally reviewed and approved by another senior radiation oncologist.

All the voxels belonging to the GT segmentation of the CTV were extracted and labeled based on the patient’s radiotherapy structures and CT data exported from the treatment planning system. Only the CT images containing the CTV were included as training and test data. We marked the CTV part as label 1 and the other part as background according to the patient’s CT images and radiotherapy structures. We preprocessed both kinds of CT images to the standard Hounsfield unit value and truncated the CT image intensity values of all scans to the range of $[-150, 200]$ Hounsfield units to remove the irrelevant details. Each CT slice and annotation image were cropped to a matrix size of 256×256 centered on the patient’s body center of the current CT slice.

DDUnet Model Architecture

The U-Net,²¹ which focuses on biomedical image segmentation, was proposed in 2016 and outperformed state-of-the-art techniques. The U-Net uses a series of down-samplings to reduce image size and increase the field of perception; it then uses up-sampling to expand the image size. Some information loss usually occurs in the process of reducing and then increasing the image size. To solve the problem of information loss in the process of down-sampling, the DDUnet integrates a multi-path dilated convolution module into the U-Net framework to extract original context information directly from the input images and compensate contextual features into high-level convolutional layers.

In the present study, the dilated convolutional module included 3 dilation convolutions with dilation factors of 1, 2, and 4 and thus had receptive fields with a size of 3×3 , 5×5 , and 9×9 pixels. The rectified linear unit function and a stride size of 1 pixel were adopted by the dilation convolution. The same padding was added to the dilation convolution process to maintain the feature size. The 3 dilation convolutions containing 64, 128, and 256 filters with a size of 3×3 respectively generated 64, 128, and 256 feature maps from the original input with a size of 256×256 . To compensate low-level context information to a high-level convolutional layer, the features output by the dilated convolution branches were used to connect with the U-Net encode layers. However, the size of those multiscale feature maps had to match, so the maximum pooling operation with a size of 3×3 and a stride size of 2, 4, and 8 pixels was used to match with the higher merged layer in the DDUnet.

As shown in Figure 2, we constructed a 5-level hierarchical DDUnet with some innovative modifications based on the original design (U-Net) to achieve the goal of CTV segmentation

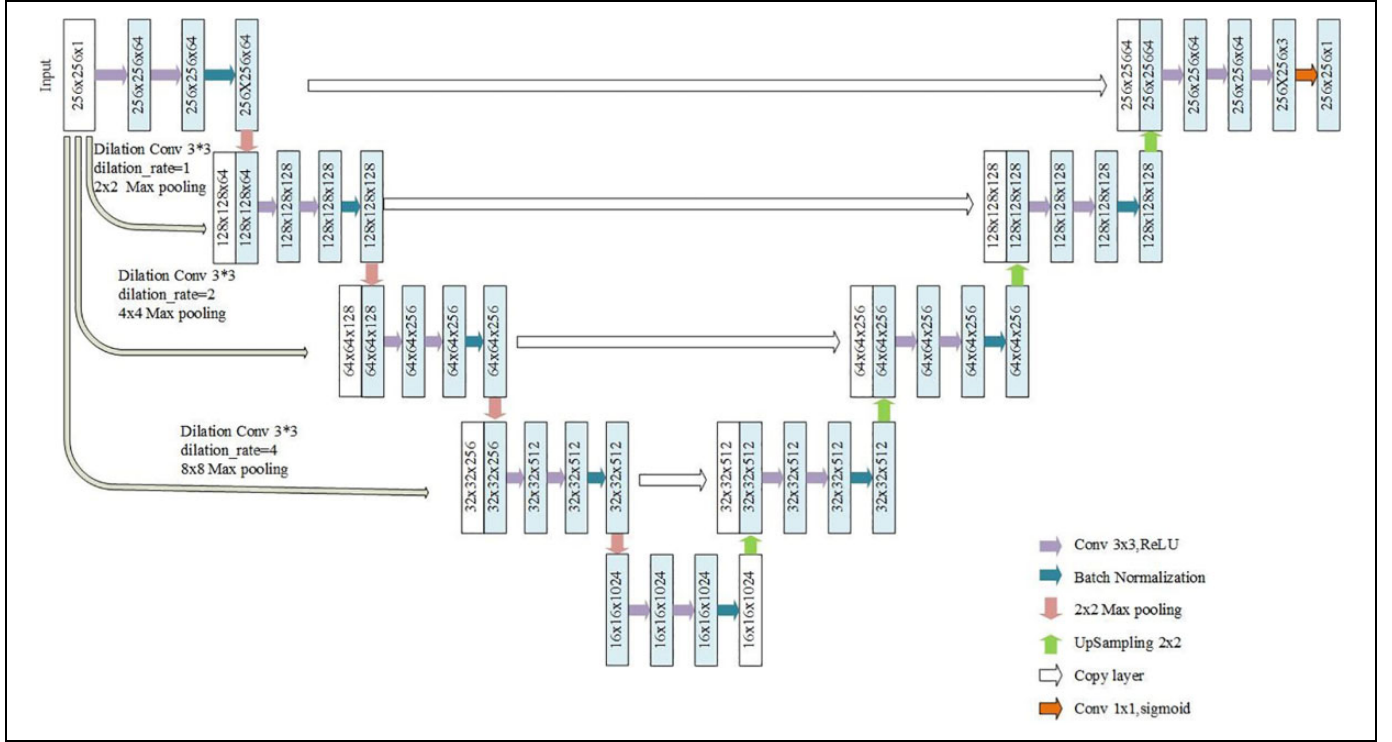


Figure 2. DDUnet architecture. The numbers in the boxes represent the outputs of the operation. The third dimension is the number of features, and the numbers of the first 2 dimensions represent the size of each 2-dimensional feature.

of EC. The detailed structure of the DDUnet model is shown Table 1. The input started with 256×256 pixel images. Five levels with 4 maximum pooling operations were chosen to reduce the feature size from 256×256 to 16×16 pixels, allowing for the 3×3 convolution with rectified linear unit operations to connect the center of the tumor to the edge of the body for all patients. The same padding was added to the convolution process to maintain the feature size. In the DDUnet, BN was added after the 2 convolutions for every level, which allowed for a more equal updating of the weights throughout the network and led to faster convergence. After the BN operation, the features were down-sampled after maximum pooling to the next level and connected with the output of the dilated module by the concatenate function of the Keras package. Our experiments showed that the number of BN operations should not be too high; only 1 BN per level was enough. The pooling options reduced the spatial size of the feature map, which needed to be recovered to the original spatial size for the segmentation task. Therefore, the decoder part deployed a deep neural network, which took pooling layer 5 as the input and a series of up-sampling layers. All layers used 3×3 convolution with the same padding set. Thus, we could carry out pixel-level classification for the segmentation task. The final outputs generated the predicted label for each pixel.

In deep learning, the loss function is the “baton” of the whole network model, which guides the network parameter learning through the error back-propagation caused by the

marking of prediction samples and real samples. To make the model converge quickly, we used a loss function based on the DSC that can express the segmentation results directly and address the class imbalance problem present in the target volume segmentation data. The loss function is shown as

$$\text{Loss} = 1 - \frac{2 \times (CTV_{\text{predict}} \cap CTV_{\text{GT}})}{CTV_{\text{predict}} \cup CTV_{\text{GT}}}, \quad (1)$$

where CTV_{predict} represents the auto-segmented CTV and CTV_{GT} is the GT segmentations of CTV .

Model Training

To assess the overall performance of the model, 19 patients with 1104 CT slices were randomly selected as a test set, and a 5-fold cross-validation procedure was then performed on the remaining 72 patients with 3482 CT slices. Each of the 5 models divided the 3482 CT slices into 80% in the training set and 20% in the validation set. Five separate models were initialized, trained, and validated in a unique combination of training and validation. Each model predicted a pixel classification label from the CT image. From these 5 trained models, we took the best-performing model based on its validation and training DSC and evaluated this model in the test set.

The Adam algorithm was chosen as the optimizer to minimize the loss function. We used a learning rate of 1×10^{-4} and the default Adam parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$, and

Table 1. Detailed Model of DDUnet.

Layer name	Type	Stride	Dilation	Output
Dilation conv	3 × 3	1	1	256 × 256 × 64
	3 × 3	1	2	256 × 256 × 128
	3 × 3	1	4	256 × 256 × 256
Maxpool1	3 × 3	2	None	128 × 128 × 64
	3 × 3	4	None	64 × 64 × 128
	3 × 3	8	None	32 × 32 × 256
Conv1 (×2)	3 × 3	1	None	256 × 256 × 64
BN				256 × 256 × 64
Maxpool2	2 × 2	2	None	128 × 128 × 64
Conv2 (×2)	3 × 3	1	None	128 × 128 × 128
BN				128 × 128 × 128
Maxpool3	3 × 3	2	None	64 × 64 × 128
Conv3 (×2)	3 × 3	1	None	64 × 64 × 256
BN				64 × 64 × 256
Maxpool4	2 × 2	2	None	32 × 32 × 256
Conv4 (×2)	3 × 3	1	None	32 × 32 × 512
BN				32 × 32 × 512
Maxpool5	2 × 2	2	None	16 × 16 × 512
Conv5 (×2)	3 × 3	1	None	16 × 16 × 1024
BN				16 × 16 × 1024
UpSampling	2 × 2	2	None	32 × 32 × 512
Conv6 (×2)	3 × 3	1	None	32 × 32 × 512
BN				32 × 32 × 512
UpSampling	2 × 2	2		64 × 64 × 256
Conv7 (×2)	3 × 3	1	None	64 × 64 × 256
BN				64 × 64 × 256
UpSampling	2 × 2	2		128 × 128 × 128
Conv8 (×2)	3 × 3	1	None	128 × 128 × 128
BN				128 × 128 × 128
UpSampling	2 × 2	2		256 × 256 × 64
Conv9 (×2)	3 × 3	1	None	256 × 256 × 64
Conv10	3 × 3	1	None	256 × 256 × 3
Conv11	1 × 1	1	None	256 × 256 × 1

decay = 0. Because of the fast convergence of the improved U-Net and DDUnet, the evolution stopped at approximately 40 epochs. Thus, we chose 40 epochs when training the model. The deep network architecture was implemented in Keras2.1.6 with TensorFlow1.5 as the backend. One NVIDIA TITAN V GPU with 12-GB memory was used for training and testing.

The model was trained on a single slice of the patient's CT images. The output was pixel-level classification for that patient slice. The training batch size was 6 slices.

Performance Evaluation

When the training process was finished, the performance of the model was evaluated with the other 19 patients with 869 CT slices. During the testing phase, all CT slices of the 19 test patients were tested one by one. The input was the 2-dimensional CT images and the final output was the pixel-level classification, which was the most likely classification label. All voxels that belonged to GT were extracted and labeled. The DSC, Hausdorff distance (HD), and Cohen kappa coefficient (KAP)³⁰ were used to evaluate the performance of the models.

DSC: The Dice metric measures volumetric overlap between segmentation results and GT annotations. The DSC was computed as shown in equation (2):

$$\text{DSC} = \frac{2 \times (A \cap B)}{A + B}, \quad (2)$$

where A is the set of voxels in the GT and B is the corresponding set of voxels in the segmentation results. Therefore, the closer the DSC value is to 1, the closer the result of auto-segmentation is to the GT.

The HD is the maximum distance from one set to the nearest point in the other set. More formally, the HD from set A to set B is a maximin function, defined as:

$$\begin{aligned} d_H(A, B) &= \max(d(A, B), d(B, A)) \\ &= \max\{\max_{a \in A} \min_{b \in B} d(a, b), \max_{b \in B} \min_{a \in A} d(a, b)\} \end{aligned} \quad (3)$$

The 95% HD is similar to the maximum HD. However, it is based on the calculation of the 95th percentile of the distances between the boundary points in A and B. The purpose of using this metric is to eliminate the impact of a very small subset of outliers. In this study, the 95% HD was chosen.

The KAP is a measure of agreement between 2 samples.³⁰ As an advantage over other measures, the KAP takes into account the agreement caused by chance, which makes it more robust. The KAP is calculated as follows:

$$\text{KAP} = \frac{f_a - f_c}{N - f_c}, \quad (4)$$

$$f_a = TP + TN, \quad (5)$$

$$f_c = \frac{(TN + FN)(TN + FP) + (TP + TP)(FN + TP)}{N}, \quad (6)$$

where N is the total number of observations (in this study, N is the number of voxels that need to be segmented) and f_a and f_c are calculated according to the so-called confusion matrix, containing the true positives, false positives, true negatives, and false negatives.

Results

To evaluate the efficiency of the automatic segmentation, the performance of the DDUnet was compared with that of the original U-Net, modified U-Net (U-Net + BN), and attention U-Net, which integrates an attention gate into the U-Net and improves the performance for segmentation tasks in medical images.³¹ To ensure that the experiments were carried out fairly in the training stage, the same training configuration was used for all the models.

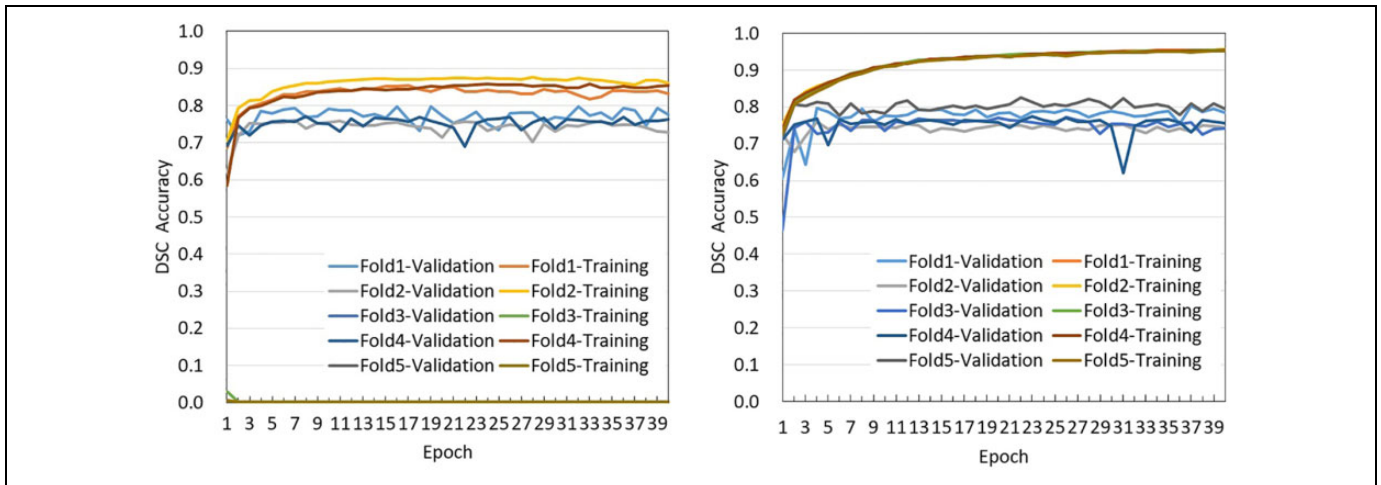


Figure 3. Plot of training and validation DSC with epochs of U-Net (left) and U-Net + BN (right).

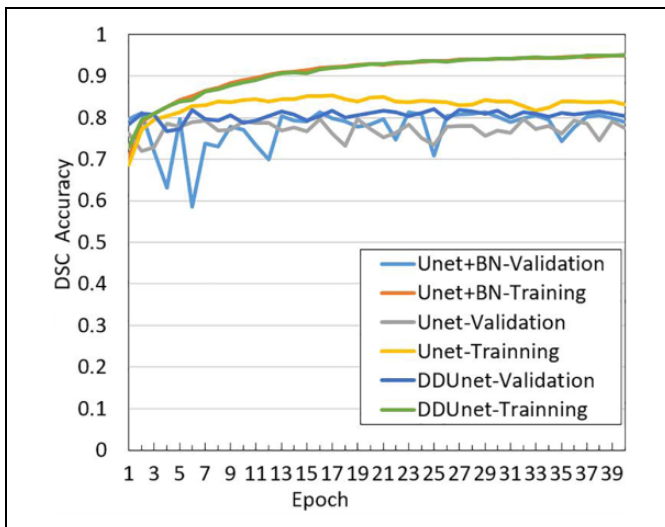


Figure 4. DSC accuracy plot of training and validation with epochs of the U-Net, U-Net + BN, and DDU-Net.

Results of Model Training

The stability of the U-Net was poor when used directly. Figure 3 (left) shows the DSC values of training and validation for each epoch when 5-fold cross-validation of the U-Net was performed. Folds 3 and 5 of the 5 folds were very slow and useless. To solve this problem, BN was added after 2 convolution layers in each level of the U-Net. Figure 3 (right) shows the DSC values of training and validation of the modified U-Net (i.e., U-Net + BN). All 5 folds were useful, indicating that the stability and convergence were better than those of the U-Net overall.

Figure 4 shows the training and validation DSC values as a function of epochs from the best-performing fold of the DDU-Net, U-Net, and U-Net + BN models. Both the trained DSC values of the DDU-Net and U-Net + BN gradually converged to 0.95, and the DSC values in the validation set tended to be stable when the epoch neared 40. For the U-Net, both the

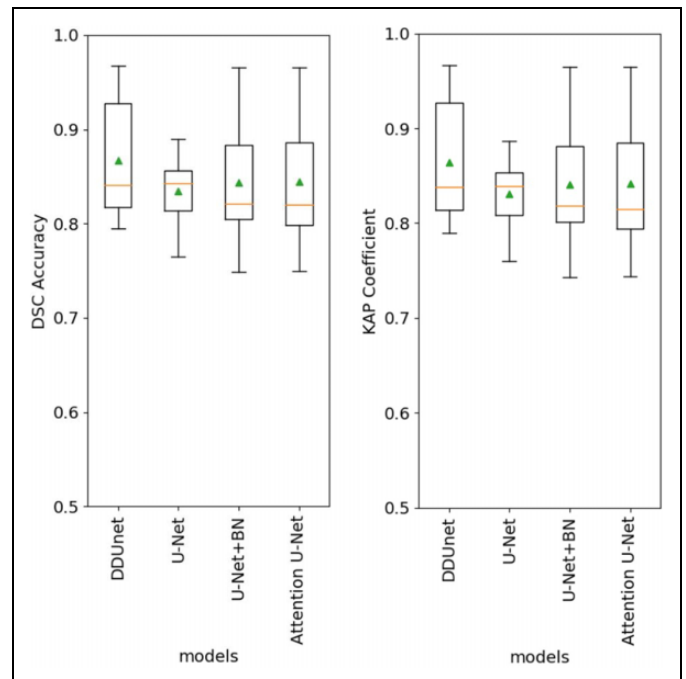


Figure 5. Boxplots obtained from DSC and KAP analyses.

training and validation DSC values were worse than those of the U-Net + BN and DDU-Net. All 3 models obtained after training were used to evaluate the segmentation performance on the test set of patients.

Results of Model Testing

The DSC, 95% HD, and KAP results for CTV segmentation in the test set are summarized in Figures 5 to 7 and Table 2. The proposed DDU-Net method outperformed the U-Net, U-Net + BN, and attention U-Net. Both the average and minimum DSC values of the DDU-Net were the best among all 4 models. The maximum DSC value was the same as that of the U-Net + BN and attention U-Net but higher than that of the U-Net. Table 2

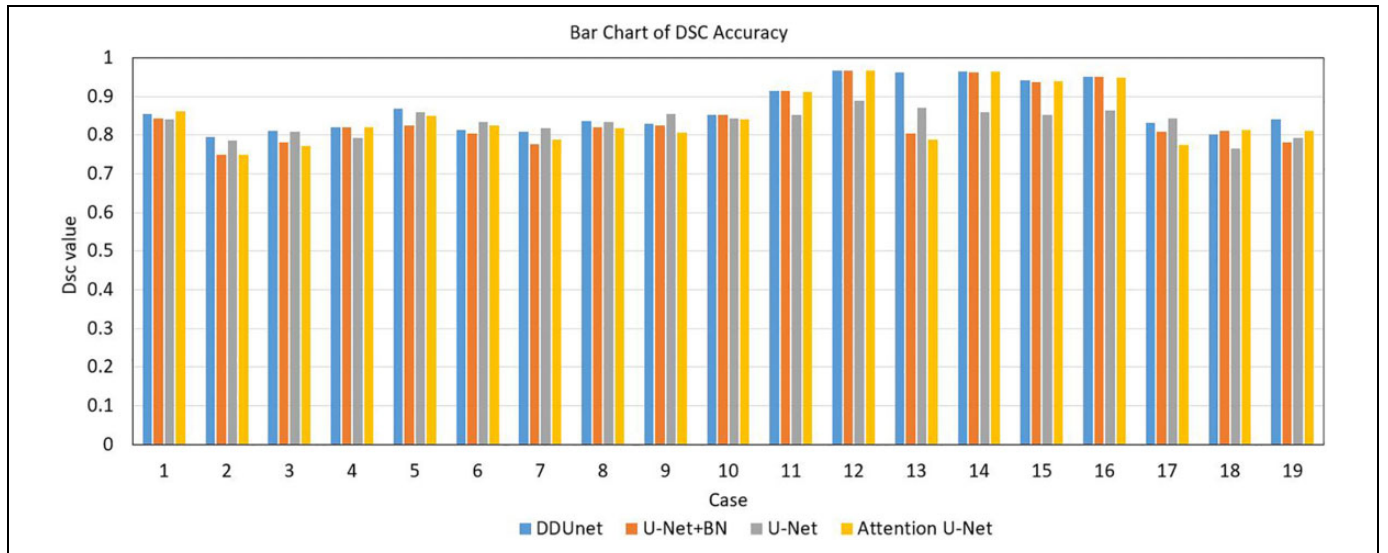


Figure 6. Bar chart of DSC values for different patients using different models.

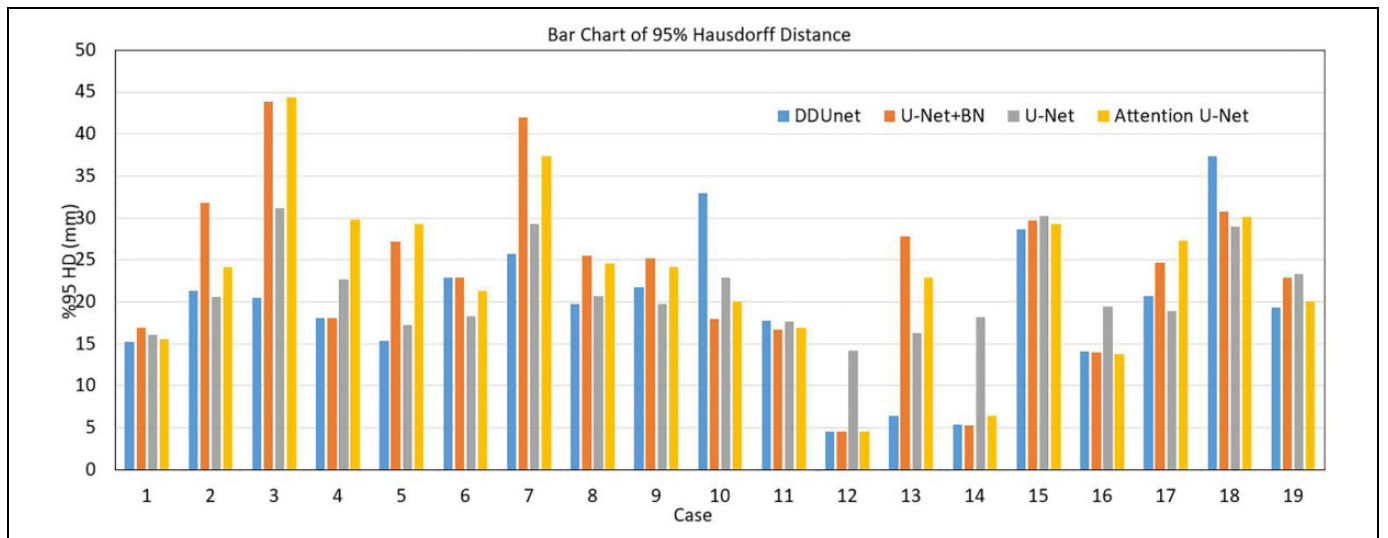


Figure 7. Bar chart of 95% HD for different patients using different models.

Table 2. Comparison of Performance of CTV Segmentation for Patients With EC After Surgery.

Models	DSC (%)			95% HD (mm)			KAP		
	Max	Avg	Min	Max	Avg	Min	Max	Avg	Min
U-Net	89.0	83.5	76.5	31.3	21.4	14.2	88.7	83.1	76
U-Net + BN	96.7	84.4	74.9	43.8	23.6	4.57	96.5	84	74.3
DDUnet	96.7	86.7	79.5	37.4	19.4	4.57	96.7	86.3	79
Attention U-Net	96.7	84.5	75	44.4	22.3	4.57	96.5	84.2	74.4

Note: The boldface values indicate the best values among all models.

and Figure 5 show that the KAP followed the same trend as the DSC. Meanwhile, the KAP for all test cases was >0.79 . Figure 6 shows the DSC accuracy of all 4 models for 19 test cases. The

DSC results of most test cases were the highest among all models; only a few were slightly lower. Figure 7 shows the 95% HD results of the test cases for all the models. The 95% HD of the DDUnet for 11 of the 19 test cases was lower than that of the other 3 models. The maximum, average, and minimum 95% HD of the DDUnet were 37.4 mm, 19.4 mm, and 4.57 mm, respectively. These values showed reasonable overlap of the auto-segmented contours with those manually delineated by senior radiation oncologists.

Figure 8 shows the CTV auto-segmentation results from the transverse, coronal, and sagittal planes. In these examples, the contours auto-segmented by the DDUnet were very close to the GT contours, although inconsistencies were present as shown in Figure 8G and J-L. Only a few corrections were needed to confirm the results of automatic segmentation.

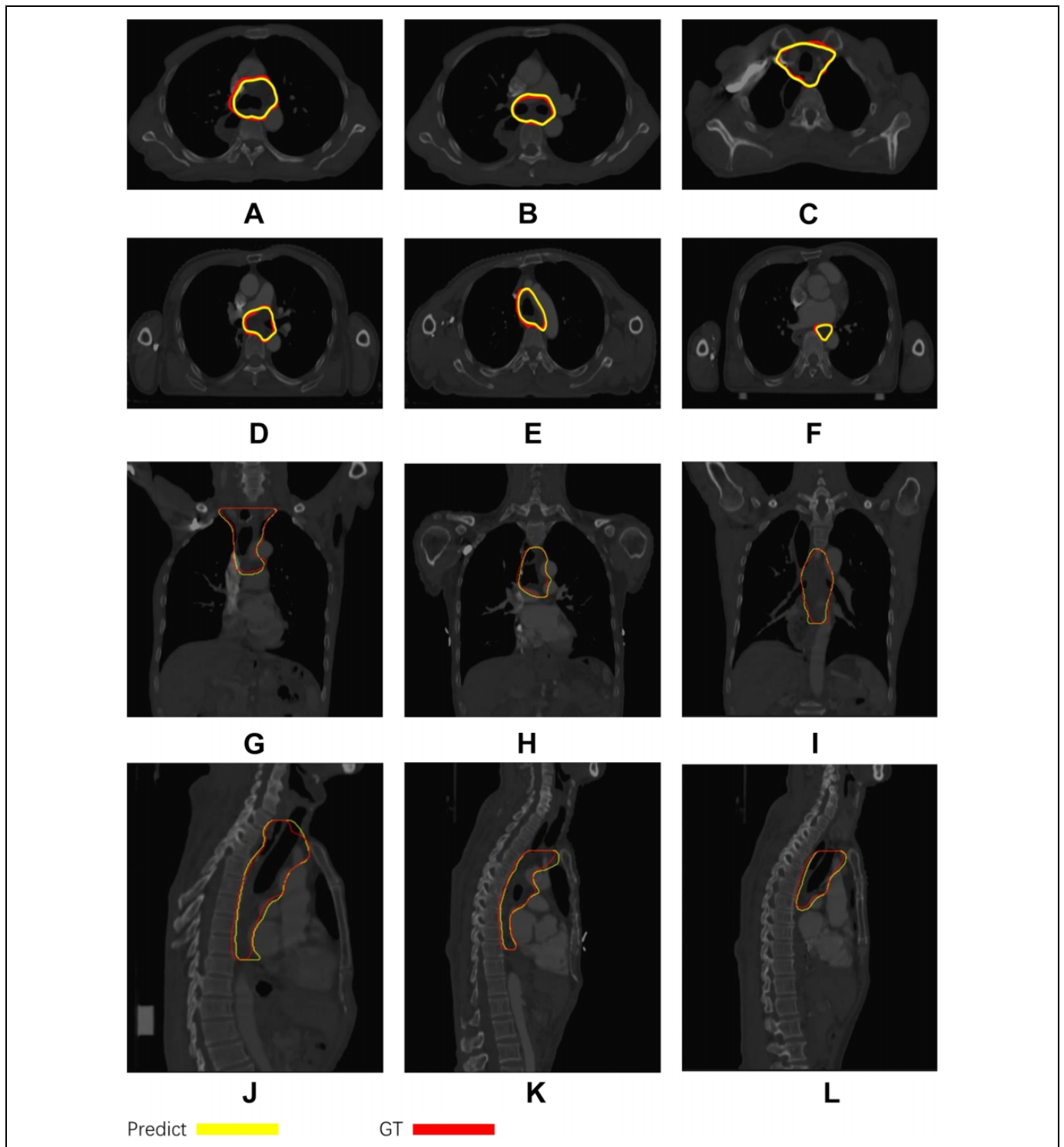


Figure 8. Segmentation results of (A-F) transverse, (G-I) coronal, and (J-L) sagittal CT slices for different patients.

The time required to perform auto-segmentation of all CTVs with the DDUnet was about 25 seconds per patient using a personal computer with an Intel Core i7-870K processor (3.7 GHz) and a NVIDIA TITAN V GPU with 12-GB memory.

Discussion

We have designed an automatic segmentation method based on deep learning for the CTV of patients with EC receiving radiotherapy after radical surgery. Our results suggest that the proposed DDUnet model can learn the semantic information from

CT images of patients with EC and produce high-quality segmentation of the CTV. Comparison of the proposed method with the U-Net and attention U-Net model showed that better segmentation performance and good convergence were achieved using our DDUnet method. The proposed DDUnet method deployed a multipath dilated convolution to extract original context information directly from input images, compensated contextual features into high-level convolutional layers, and thus improved the segmentation accuracy.

Consistency of target segmentation is important for improving radiotherapy outcomes. Interobserver and intraobserver variation is considerable.³² Automatic segmentation with guaranteed accuracy is an efficient way to reduce variability of contours among radiation oncologists. Currently, the DSC value is usually used to evaluate automatic segmentation performance. In the present study, our method showed good performance compared with the commonly used U-Net and attention U-Net on the biomedical image segmentation area. Regarding the target, the comparison is difficult because the N-stage (most often N0) and selected levels are quite different among previous studies. For the CTV, different studies have shown mean DSC values of 60%,³³ 68%/70%,³⁴ 77%,³⁵ 78%,³⁶ 79%,³⁷ 80.2%,³⁸ and 82.6%,³⁹ whereas the DSC value of the DDUnet was 86.7%. Overall, it is reasonable to conclude that the DDUnet achieved good results according to the experiments. The proposed method that learns and predicts in an end-to-end form can rapidly segment the CTV in all of a patient's image slices in approximately 25 seconds.

Although the segmentation details at the upper and lower edges of the target were slightly worse as shown in Figure 8G and J–L, the overall segmentation results were in good agreement with GT. Each patient had different surgical resection positions, resulting in different upper and lower positions of the CTV. Consequently, fewer training data were present at the edge than in the other parts.

The present study has several limitations. First, a model was trained and assessed in patients with stage I or II EC who underwent radical surgery. Tumors at different stages differ greatly in contour, volume, and complexity, and these differences influence the performance of the automated segmentation. Second, the tumors were located in the upper and middle esophagus in our study. The CTV varies greatly in terms of tumor location among patients with EC after radical surgery. This may make it difficult for the model to achieve consistently good performance in other patients who have EC with different tumor locations. This study mainly focused on EC target segmentation after radical surgery from CT images, and the training set included 63 patients. Increasing the amount of training data could make the DDUnet model more robust and accurate.

Summary

Accurate and consistent delineation of tumor targets and OARs is particularly important in radiotherapy. Several studies have focused on segmentation of OARs or the target volume using deep learning methods. However, CTV segmentation for

radiotherapy of EC after radical surgery has not been reported in the literature to date. The present report has described a method using the DDUnet architecture to auto-segment the CTV for stage I or II EC after radical surgery using CT images. The training and testing in this study were based on the original clinical data, and the segmentation results for the test cases were very close to the manual segmentation results by experienced doctors.

The results showed that the DDUnet can accurately segment CTV contours based on CT images, and only slight revision is needed for radiotherapy treatment planning. The speed of segmentation is very fast. The model obtained in this study can greatly decrease the workload of clinicians performing manual segmentation and improve their work efficiency. Thus, the DDUnet has the potential to be used in the clinical setting for auto-segmentation of the CTV of tumors treated by radiotherapy after radical surgery. In future work, multimodality tumor delineation will be studied to improve the segmentation accuracy based on other clinical cases.

Authors' Note

The source codes and data are available at <http://github.com/Ruifen-Cao/DDUnet>.

Acknowledgment

The authors thank the radiation oncologists at Anhui Provincial Cancer Hospital for assisting with the target delineation.


Declaration of Conflicting Interests

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

Funding

The authors disclose receipt of the following financial support for the research, authorship, and/or publication of this article: This study was funded in part by the National Natural Science Foundation of China under Grant 61873001, in part by the Natural Science Foundation of Anhui Province under Grant 1908085MA27, and in part by the Open Foundation of Engineering Research Center of Big Data Application in Private Health Medicine, Fujian Province University under Grant KF2020008.

ORCID iD

Ruifen Cao, PhD  <https://orcid.org/0000-0002-4223-3422>

References

1. Chen W, Zheng R, Baade PD, et al. Cancer statistics in China, 2015. *CA Cancer J Clin.* 2016;66(2):115-132.
2. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin.* 2013;65(2):87-108.
3. Kumagai K, Rouvelas I, Tsai JA, et al. Meta-analysis of post-operative morbidity and perioperative mortality in patients receiving neoadjuvant chemotherapy or chemoradiotherapy for resectable oesophageal and gastro-oesophageal junctional cancers. *Br J Surg.* 2014;101(4):321-338.

4. Ren Y, Su C, Zhou Y, Zhao X, Yang CL, Liu YY. Effect of bilateral supraclavicular postoperative radiotherapy in middle and lower thoracic esophageal carcinoma. *World J Gastroenterol*. 2014;20(47):17970-17975.
5. Almhanna K, Shridhar R, Meredith KL. Neoadjuvant or adjuvant therapy for resectable esophageal cancer: is there a standard of care? *Cancer Control*. 2013;20(2):89-96.
6. Nakagawa S, Kanda T, Kosugi S, Ohashi M, Suzuki M, Hatakeyama K. Recurrence pattern of squamous cell carcinoma of the thoracic esophagus after extended radical esophagostomy with three-field lymphadenectomy. *J Am Coll Surg*. 2004;198(2):205-211.
7. Mariette C, Balon JM, Piessen G, Fabre S, Van Seuningen JJ, Triboulet P. Pattern of recurrence following complete resection of esophageal carcinoma and factors predictive of recurrent disease. *Cancer*. 2010;97(7):1616-1623.
8. Cai WJ, Xin PL. Pattern of relapse in surgical treated patients with thoracic esophageal squamous cell carcinoma and its possible impact on target delineation for postoperative radiotherapy. *Radiother Oncol*. 2020;96(1):104-107.
9. Li CL, Zhang FL, Wang YDC, et al. Characteristics of recurrence after radical esophagectomy with two-field lymph node dissection for thoracic esophageal cancer. *Oncol Lett*. 2013;5(1):355-359.
10. Zhu Y, Li M, Kong L, Yu J. Postoperative radiation in esophageal squamous cell carcinoma and target volume delineation. *Oncotargets Ther*. 2016;9:4187-4196. doi:10.2147/OTT.S104221
11. Pereira GC, Traughber M, Muzic RF. The role of imaging in radiation therapy planning: past, present, and future. *Biomed Res Int*. 2014;2014:231090. doi:10.1155/2014/231090
12. Vinod SK, Myo M, Michael GJ, Lois CH. A review of interventions to reduce inter-observer variability in volume delineation in radiation oncology. *J Med Imaging Radiat Oncol*. 2016;60(3):393-406.
13. Yamazaki H, Hiroya S, Takuji T, Naohiro K. Quantitative assessment of inter-observer variability in target volume delineation on stereotactic radiotherapy treatment for pituitary adenoma and meningioma near optic tract. *Radiat Oncol*. 2011;6(1):10.
14. Ashok M, Gupta A. A systematic review of the techniques for the automatic segmentation of organs-at-risk in thoracic computed tomography images. *Arch Computat Method Eng*. 2020;28:3245-3267. doi:10.1007/s11831-020-09497-z
15. Iglesias J, Eand M, Sabuncu R. Multi-atlas segmentation of biomedical images: a survey. *Med Image Anal*. 2015;24(1):205-219.
16. Cabezas M, Oliver A, Lladó X, Freixenet J, Cuadra MB. A review of atlas-based segmentation for magnetic resonance brain images. *Comput Methods Programs Biomed*. 2011;104(3):158-177.
17. Langerak TR, Berendsen FF, Van der Heide UA, Kotte AN, Pluim JP. Multiatlas-based segmentation with preregistration atlas selection. *Med Phys*. 2013;40(9):091701. doi:10.1118/1.4816654.
18. Krizhevsky A, Sutskever I, Hinton G. ImageNet classification with deep convolutional neural networks. *NIPS*. 2012;25(2):1097-1105.
19. Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, June 24-27, 2014:580-587. IEEE.
20. Yu F, Koltun V. *Multi-Scale Context Aggregation by Dilated Convolutions*. ICLR; 2016.
21. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Munich, Germany, October 5-9, 2015:234-241. Springer.
22. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, June 24-27, 2014;39:640-651. IEEE. doi:10.1118/1.4816654
23. Ibragimov B, Xing L. Segmentation of organs-at-risks in head and neck CT images using convolutional neural networks. *Med Phys*. 2017;44(2):547-557.
24. Tong N, Gou S, Yang S, Ruan D, Sheng K. Fully automatic multi-organ segmentation for head and neck cancer radiotherapy using shape representation model constrained fully convolutional neural networks. *Med Phys*. 2018;45(10):4558-4567.
25. Men K, Chen XY, Zhang T, et al. Deep deconvolutional neural network for target segmentation of nasopharyngeal cancer in planning computed tomography images. *Front Oncol*. 2017;7(315):1-9.
26. 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 network. *Med Phys*. 2017;44(12):6377-6389.
27. Li S, Xiao J, He L, Peng X, Yuan X. The tumor target segmentation of nasopharyngeal cancer in CT images based on deep learning method. *Technol Cancer Res T*. 2019;18:1-8. doi:10.1177/1533033819884561
28. Zhang F, Wang Q, Li H. Automatic segmentation of the gross target volume in non-small cell lung cancer using a modified version of resNet. *Technol Cancer Res T*. 2020;19:1-9. doi:10.1177/1533033820947484
29. Sahiner B, Pezeshk A, Hadjiiski ML, et al. Deep learning in medical imaging and radiation therapy. *Med Phys*. 2018;46(1):1-36. doi:10.1002/mp.13264
30. Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Med Imaging*. 2015;15(1):29. doi:10.1186/s12880-015-0068-x
31. Oktay O, Schlemper J, Le Folgoc L, et al. Attention u-net: learning where to look for the pancreas. *MIDL*. 2018:1-10. doi:arxiv.org/abs/1804.03999
32. Leunens G, Menten J, Weltens C, Verstraete J, Schueren EVD. Quality assessment of medical decision making in radiation oncology: variability in target volume delineation for brain tumours. *Radiother Oncol*. 1993;29(2):169-175.
33. Jean-François D, Andreas B. Atlas-based automatic segmentation of head and neck organs at risk and nodal target volumes: a clinical validation. *Radiat Oncol*. 2013;8(154):1-11. doi:10.1186/1748-717X-8-154
34. 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

35. Qazi AA, Pekar V, Kim J, Xie J, Breen SL, Jaffray DA. Auto-segmentation of normal and target structures in head and neck CT images: a feature-driven model-based approach. *Med Phys*. 2011; 38(11):6160-6170. doi:10.1118/1.3654160
36. Tsuji SY, Hwang A, Weinberg V, Yom SS, Quivey JM, Xia P. Dosimetric evaluation of automatic segmentation for adaptive IMRT for head-and-neck cancer. *Int J Radiat Oncol Biol Phys*. 2010;77(3):707-714.
37. Stapleford LJ, Lawson JD, Perkins C, et al. Evaluation of automatic atlas-based lymph node segmentation for head-and neck cancer. *Int J Radiat Oncol Biol Phys*. 2010;77(3): 959-966.
38. Gorthi S, Duay V, Houhou N, Cuadra MB, Schick U, Becker M. Segmentation of head and neck lymph node regions for radiotherapy planning using active contour-based atlas registration. *IEEE J Sel Topics Signal Process*. 2010;3(1):135-147.
39. 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.