Original paper

Segmentation of organs-at-risk in cervical cancer CT images with a convolutional neural network

Zhikai Liu^a, Xia Liu^{a,1}, Bin Xiao^b, Shaobin Wang^b, Zheng Miao^a, Yuliang Sun^a, Fuquan Zhang^{a,*}^a Department of Radiation Oncology, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100730, China^b MedMind Technology Co., Ltd., Beijing 100080, China

ARTICLE INFO

Keywords:

Cervical cancer radiotherapy
Automatic segmentation
Organs-at-risk
Convolutional neural network

ABSTRACT

Purpose: We introduced and evaluated an end-to-end organs-at-risk (OARs) segmentation model that can provide accurate and consistent OARs segmentation results in much less time.

Methods: We collected 105 patients' Computed Tomography (CT) scans that diagnosed locally advanced cervical cancer and treated with radiotherapy in one hospital. Seven organs, including the bladder, bone marrow, left femoral head, right femoral head, rectum, small intestine and spinal cord were defined as OARs. The annotated contours of the OARs previously delineated manually by the patient's radiotherapy oncologist and confirmed by the professional committee consisted of eight experienced oncologists before the radiotherapy were used as the ground truth masks. A multi-class segmentation model based on U-Net was designed to fulfil the OARs segmentation task. The Dice Similarity Coefficient (DSC) and 95th Hausdorff Distance (HD) are used as quantitative evaluation metrics to evaluate the proposed method.

Results: The mean DSC values of the proposed method are 0.924, 0.854, 0.906, 0.900, 0.791, 0.833 and 0.827 for the bladder, bone marrow, femoral head left, femoral head right, rectum, small intestine, and spinal cord, respectively. The mean HD values are 5.098, 1.993, 1.390, 1.435, 5.949, 5.281 and 3.269 for the above OARs respectively.

Conclusions: Our proposed method can help reduce the inter-observer and intra-observer variability of manual OARs delineation and lessen oncologists' efforts. The experimental results demonstrate that our model outperforms the benchmark U-Net model and the oncologists' evaluations show that the segmentation results are highly acceptable to be used in radiation therapy planning.

1. Introduction

Cervical cancer is one of the most commonly diagnosed diseases for women worldwide [1]. Radiation therapy is a major clinical treatment for cervical cancer [2]. Accurate segmentation of organs-at-risk (OARs) is critical for minimising radiation toxicities to these normal structures during irradiation. Manual delineation of the OARs regions is considered the gold standard in current clinical practice. However, OARs delineation is time-consuming and labour-intensive work for radiation oncologists. It is estimated that an oncologist needs 90–120 min to delineate the OARs in a cervical cancer patient [3]. Moreover, due to the different levels of expertise and preferences of radiation oncologists, large inter-and intra-observer variation has been reported [4–10]. The variation may even introduce more differences than planning and setup errors [4,11–13].

Automatic OARs segmentation tools can alleviate oncologists'

efforts and reduce variability. Related algorithms and models have been widely discussed. The poor contrast and high variability in size, location and shape of cervical cancer OARs restrict the use of intensity-based and model-based segmentation methods [14,15]. The methods that have been commonly used in the existing studies during last few years can be categorized into two groups: Atlas-based methods and Convolutional Neural Network (CNN) based methods.

Previously, auto-segmentation techniques have been clustered into 3 generations of algorithms, with multi-atlas based and hybrid techniques (third generation) being considered the state-of-the-art [16]. Atlas-based methods are straightforward, and they commonly use an atlas to provide the training images to match with the testing images [17–19]. The performance of atlas-based methods heavily relies on the deformable image registration accuracy and the selected atlases. Atlas-based methods have achieved acceptable results for segmentation of OARs especially for head and neck cancer patients [20–23]. The registration

* Correspondence author.

¹ Xia Liu contribute equally to this article.

algorithms tend to work well for the structure in head and neck with relatively fixed locations and shapes. But most results still require significant manual editing as atlas-based methods encountered difficulties with patient variability in the regions with small areas, regions with larger motion, deformable organs or organs near tumour regions [23–25]. For instance, Ayadi et al. evaluated the performance of an atlas-base method in through a multi-centre study in the case of prostate cancer, the mean Dice Similarity Coefficients (DSC) $\pm$ standard deviations were 0.62 ± 0.09 , 0.80 ± 0.19 , 0.66 ± 0.09 and 0.69 ± 0.1 for prostate, bladder, rectum and hips respectively [26]. Although multi-atlas was introduced to improve the robustness of segmentation as compared to single-atlas, they are prone to topological errors, potentially affecting subsequent treatment planning and require time-consuming manual editing [27,28].

CNN based models in deep learning methods have become dominant solutions for natural image semantic segmentation problems. Popular models include the FCN [29], SegNet [30], U-Net [31], FC-DenseNet [32], PSPNet [33], DeepLabv3+ [34] and so on. In the field of OARs delineation in radiation therapy, deep learning based auto-segmentation techniques have shown to provide significant improvements over more traditional approaches suggesting we have entered the fourth generation of auto-segmentation algorithm development [16]. Recently published studies showed the effectiveness of CNN-based models in head and neck cancer OARs [35–37], thoracic OARs [38–40], abdomen OARs [41], and pelvis OARs [42,43]. Studies taken in [44] demonstrate deep learning approach often performed better than the multi-atlas method in radiotherapy automatic segmentation field. Among all the CNN based segmentation models, U-Net was commonly used, including for the multiple pelvis organs segmentation task. Samaneh K. et al. used the U-Net architecture to delineate the prostate, bladder, and rectum in male pelvic CT images [45]. Anjali B. et al. proposed a 3D U-Net with encoding arm modified with aggregated residual networks (ResNeXt) for volumetric segmentation of prostate, bladder, rectum, and femoral heads [46]. A multi-pass U-Net was proposed to automatic segment the pelvis in [47].

U-Net is a well-known CNN architecture for medical image segmentation. It follows the encoding and decoding path to create dense outputs as FCN, and also includes skip-connections, which concatenate features from the encoding to the decoding layers, to provide higher resolution features to better localize and learn representations from the input image. As an efficient end-to-end training model, U-Net does not require a pre-trained network, and could be trained from scratch to achieve accurate segmentation results with very little labelled training data [31]. Unlike natural images, it is hard to collect and label medical images, even for experienced radiologists and oncologists. Furthermore, due to privacy issues, it is often not easy to provide large amounts of data to the public research community. Hence, it is necessary to design a segmentation model that can be trained from scratch with relatively less medical data. In this study, we proposed a new segmentation network based on the 2D U-Net model to segment multiple OARs in cervical cancer CT images slice by slice. We trained the model from scratch using the amount of data we have and achieved satisfactory performance. To the best of our knowledge, this is the first work on cervical cancer OARs segmentation including the bladder, bone marrow, femoral head, rectum, small intestine and spinal cord. In this model, we proposed a Context Aggregation Block to extract the features in a range of receptive fields and used a Squeeze-Extract Block [48] to reweight the organ mask importance. Comprehensive experiments were conducted and the results demonstrate the efficiency and effectiveness of our proposed model.

2. methods and materials

2.1. Data

The anonymized data set consists of 105 patients' CT scans, collected at Peking Union Medical College Hospital, Beijing, China. All

these patients were diagnosed as having locally advanced cervical cancer. The average age $\pm$ standard deviation of these patients were 51.81 ± 10.14 years old. All the CT scans followed the DICOM protocol and were scanned using a Philips Brilliance Big Bore CT scanner from November 2012 to January 2015. The pixel spacing of the data was $1.1543 \text{ mm} \times 1.1543 \text{ mm}$ with a 512×512 matrix, and the thickness was 5 mm with slice number ranged from 71 to 119 (the average number was 96 slices per patient). Seven organs, including the bladder, bone marrow, femoral head left, femoral head right, rectum, small intestine and spinal cord, were defined as cervical cancer OARs. OARs contours delineated manually by trained radiation oncologists before radiotherapy in the clinical practice were used as the segmentation ground truth. To ensure the delineation quality in this medical centre, all the delineated contours was reviewed and modified (only when needed) together by a professional radiation oncologist committee consists of eight oncologists with more than 10 years' experience on radiotherapy of pelvic tumours, which is the clinical routine in the radiotherapy department of this hospital.

2.2. Proposed method

Cervical cancer OARs segmentation is a challenging task since these organs have various sizes, shapes, and locations and some of them may have several isolated regions and unclear boundaries. Following the design of popular deep learning-based segmentation models, we also formulated this OARs segmentation problem as a binary pixel-level classification problem. The predicted masks and ground truth masks were one-hot encoded, which means each mask contained 6 channels in a fixed order with one channel representing one organ. Note that we treated the femoral head left and femoral head right as one organ in the model but evaluated them as two individual organs to meet clinical requirements. Hence, we have 6 channels instead of 7 channels. Despite the simple architecture of U-Net, many studies have demonstrated the effectiveness of U-Net in medical image segmentation [49–52]. Benefitting from U-Net, we designed a new segmentation model that followed the U-Net architecture. Our model contains an encoder part and a decoder part, while the convolutional layers in the U-Net are replaced by Context Aggregation Blocks. Both the encoder and decoder consist of five Context Aggregation Blocks, and the feature maps in the encoder part are concatenated to the corresponding feature map in the decoder part. Fig. 1 shows the overall architecture of our network.

The Context Aggregation Block, which is shown in Fig. 2, uses dilated convolutions with different dilated rates and normal convolution layers with different kernel sizes. This convolution block is similar to the method described in [42,53], but we also utilize common convolution layers with larger kernel sizes [54]. The outputs of these convolution layers are concatenated as the final output feature map. In almost all CNN models, the feature maps are gradually down sampled to capture the semantic contextual information in different image scales and gain a sufficiently large receptive field [55]. The features in a larger receptive field can help to improve the localization capabilities and reduce the false positive predictions, which means that the model predicted isolated regions that should not be segmented [56]. In addition, the dilation convolution can acquire the features in a larger receptive field without increasing the number of parameters. Although the dilation convolution has achieved success in many studies, it is also reported that the dilation convolution can cause “gridding artefacts”, which can degrade the model's performance [57]. Therefore, we combine dilation convolution layers with dilation rate of 2 and 3 and normal convolution layers with large kernel sizes to aggregate the features.

After the convolution layers, we use the Squeeze-Extract block to assign different weights to each channel, thus to reweight each organ mask importance. The Squeeze-Extract block is illustrated in Fig. 3. It contains a global average pooling layer, two fully connected layers and a sigmoid layer, which scale the output of each channel to a value

Fig. 1. Overall architecture of the proposed model.

between 0 and 1. Then, the original feature map is multiplied by the value that we obtain from the sigmoid.

Our model is a multi-class classification model, and class balance is a key issue when training deep learning models. The dataset is naturally imbalanced because of organs' different sizes and appearances. More specifically, the small intestine, bladder and bone marrow are much larger, which means that they have more pixels than the rectum, spinal cord and femoral head. To overcome the class imbalance issue, we use the class weighted cross-entropy loss function. For prediction scores y

and training targets x , the weighted cross-entropy loss L is defined as follows:

$$\mathcal{L}(x, y) = L = \text{mean}(\{l_1, l_2, \dots, l_n\}) \quad (1)$$

$$l_n = -w_n [y_n \log x_n + (1 - y_n) \log(1 - x_n)] \quad (2)$$

the loss L is the mean loss of n classes, and the loss of each class l_n is the product of the corresponding weight (w_n) and cross entropy loss, where the weights are set to fix values according to the size of organs by experience.

Fig. 2. Context Aggregation Block used in the proposed model.

Fig. 3. Squeeze-Extract Block used in the proposed model.

2.3. Implementation details

We implemented our model using Python 3.6 and PyTorch 1.0 [58]. The dataset was split into three separate parts: the training set, the validation set and the testing set. The training set contained 77 patients, the validation set contained 14 patients and the testing set also contained 14 patients. We clamped the input image intensity to -1024 HU and 2048 HU. Zero-mean normalization was used for preprocessing. We used the Adam [59] optimization algorithm for the optimization with a 0.001 learning rate. The model was trained over 50 epochs and the best model was selected, which is the one that had the lowest validation loss score. The validation loss score of the selected model was 0.017 and the corresponding training loss score was 0.012 . The convolutional layers are initialized using Xavier Uniform Initialization [60]. All these convolution layers were followed by a batch normalization layer [61] and a ReLU layer [62]. We trained and tested our model using a GTX 1080 GPU.

3. Results

3.1. Evaluation metrics

We used the Dice Similarity Coefficient (DSC) [63] and 95th percentile Hausdorff Distance (HD) [64] to evaluate our model. The DSC is defined as follows:

$$DSC = \frac{2 |X \cap Y|}{|X| + |Y|} \quad (3)$$

where $X = \{x_1, \dots, x_n\}$ and $Y = \{y_1, \dots, y_n\}$ are two finite point sets. X is the predicted mask, and Y is the ground truth. $|X \cap Y|$ represents the intersection of X and Y and $|X| + |Y|$ represents the union of X and Y . The 95HD is defined as follows:

$$95HD(X, Y) = \text{percentile}(h(X, Y) \cup h(Y, X), 95th) \quad (4)$$

$$h(X, Y) = \max_{x \in X, y \in Y} (\min(\|x - y\|)) \quad (5)$$

$\|\cdot\|$ is the Euclidean norm of the points of A and B . When the 95HD decreases, the overlap between X and Y increases. Both the DSC value and 95HD value are calculated as the mean values by averaging all the values obtained from each 2D slice. We used U-Net as the benchmark model since it is the state-of-the-art model in many segmentation problems and our method is proposed based on U-Net architecture. The results are summarized in Fig. 4 and Table 1.

3.2. Oncologist evaluation

An oncologist with more than 15 years of clinical experience evaluated the segmentation result of each slice and graded the results on four levels: Acceptable, Minor Revision, Major Revision and Not Acceptable. The test data contained 14 patients and 1253 slices. Fig. 5 shows some segmentation samples that are produced by the proposed models. Table 2 shows the evaluation results. The values in the table are the numbers of slices and the corresponding percentage. Note that different OARs have different numbers of slices.

3.3. Processing time

OARs' delineation is time-consuming work for oncologists. Averagely, it takes more than 30 s to delineate the OARs on one slice by an experienced oncologist, while our proposed method can process one slice in 0.06 s when running on a GTX 1080 GPU.

4. Discussion

OARs delineation is a key step before radiotherapy planning. Precise treatment delivery relies heavily on accurate OARs delineation. We trained and evaluated a convolution neural network for OARs segmentation. To the best of our knowledge, this is the first study on cervical cancer OARs segmentation.

The success of the mentioned models such as DeepLabv3+ and PSPNet that have been applied to natural image datasets heavily relies on the pre-trained backbone models and large amounts of labelled data. However, the amount of medical datasets is significant less compared with natural image datasets. It is shown that, for X-ray images and retinal fundus photographs, transfer learning only offers limited performance benefits and that smaller models with significantly worse ImageNet accuracy can perform equally well on medical tasks [65]. Therefore, we did not use any pre-trained model or transfer learning methods and trained our model from scratch.

We use U-Net as the benchmark model and improve U-Net to get more accurate result for OARs delineation. To capture the semantic contextual information, we used the dilation convolution and large kernels to aggregate contextual features in the Context Aggregation Blocks, enabling the model to learn multiple-scale features and extract structured relationship features effectively. In addition, to explore the differences between all the target organs, we used the Squeeze-Extract

Fig. 4. Boxplots. (a) DSC from U-Net, (b) 95HD from U-Net, (c) DSC from proposed model, (d) 95HD from proposed model. 1-bladder, 2-bone marrow, 3-femoral head left, 4-femoral head right, 5-rectum, 6-small intestine and 7-spinal cord.

Table 1
DSC and 95HD for U-net and our model.

Anatomy	U-Net		Proposed Model	
	DSC $\pm$ STD	95HD $\pm$ STD (mm)	DSC $\pm$ STD	95HD $\pm$ STD (mm)
Bladder	0.902 $\pm$ 0.110	2.406 $\pm$ 3.526	0.924 $\pm$ 0.046	5.098 $\pm$ 11.844
Bone Marrow	0.846 $\pm$ 0.047	1.957 $\pm$ 0.885	0.854 $\pm$ 0.054	1.993 $\pm$ 1.153
Femoral Head Left	0.893 $\pm$ 0.026	1.462 $\pm$ 0.496	0.906 $\pm$ 0.031	1.390 $\pm$ 0.380
Femoral Head Right	0.897 $\pm$ 0.022	1.557 $\pm$ 0.436	0.900 $\pm$ 0.023	1.435 $\pm$ 0.413
Rectum	0.778 $\pm$ 0.034	7.159 $\pm$ 5.259	0.791 $\pm$ 0.032	5.949 $\pm$ 3.781
Small Intestine	0.813 $\pm$ 0.035	6.274 $\pm$ 2.274	0.833 $\pm$ 0.030	5.281 $\pm$ 2.294
Spinal Cord	0.823 $\pm$ 0.057	3.362 $\pm$ 3.413	0.827 $\pm$ 0.063	3.269 $\pm$ 3.756

Bold values indicate better performance of the two results from U-net and our model.

Fig. 5. Segmentation results for test cases shown in transverse CT slices. Red lines: ground truths; green lines: segmentation results; organs from 2nd row to 7th row: bladder, bone marrow, femoral head, rectum, small intestine and spinal cord; oncologist evaluation result from 1st column to 4th column: Acceptable, Minor Revision, Major Revision and Not Acceptable.

Table 2
Evaluation of the prediction results from our proposed model by experienced oncologists.

Anatomy	Acceptable	Minor Revision	Major Revision	Not Acceptable
Bladder	132 (58.15%)	73 (32.16%)	17 (7.49%)	5 (2.20%)
Bone Marrow	285 (52.10%)	213 (38.94%)	49 (8.96%)	0 (0.00%)
Femoral Head Left	83 (68.03%)	34 (27.87%)	5 (4.10%)	0 (0.00%)
Femoral Head Right	80 (64.00%)	44(35.20%)	1 (0.80%)	0 (0.00%)
Rectum	158 (49.84%)	101 (31.86%)	49 (15.46%)	9 (2.84%)
Small Intestine	203 (38.52%)	215 (40.80%)	93 (17.65%)	16(3.04%)
Spinal Cord	173 (58.84%)	106 (36.05%)	14 (4.76%)	1(0.34%)
Mean Percentage	55.64%	34.70%	8.46%	1.20%

block to reweight the importance of different channels, thus preserve the organs with relatively small areas. Note that the proposed method does not make the network deeper than the benchmark model, and this makes the processing time almost the same as the benchmark model. Men K et al. in their study [66] also used the dilation convolution, but they designed a deeper network to improve the performance. Comparatively, our proposed model has fewer parameters, which makes it faster. As Men K reported in their study, their best model could process one patient's CT scans, including 70 slices, in 15 s. Our proposed model can process one slice in 0.06 s or 70 slices in 4.2 s, which is several thousand times faster than manual delineation.

The mean DSC values of our proposed method are 0.924 for the bladder, 0.854 for bone marrow, 0.906 for the femoral head left, 0.900 for the femoral head right, 0.791 for the rectum, 0.833 for the small intestine and 0.827 for the spinal cord. The experimental results in

Fig. 4 and Table 1 demonstrate that our model outperforms the benchmark model. Among all the OARs, bladder achieved the highest DSC value, due to bladder's relatively clear boundary and high contrast between bladder and background. To further compare the accuracy with prior-of-art methods, a comprehensive summary of pelvis organs segmentation results is provided in Table 3. It is not a precise way to compare the evaluation index values directly because many study factors influence the results, including single or multiple data sources, image scanning machines and protocols, cancer types, OAR contouring standards and so on. But the comparison summary still tends to demonstrates that the deep learning based methods outperform the traditional intensity-based, model-based and atlas-based methods. Deep learning based auto-segmentation techniques have shown to provide significant improvements over more traditional approaches suggesting we have entered the fourth generation of auto-segmentation algorithm

Table 3
Summary of state-of-the-art results and results from this work for pelvis organs segmentation.

Method	Image Modality	Cases	Results
2D FCN	CT	60	Bladder DSC = 0.93 [42] Colon DSC = 0.62 Intestine DSC = 0.65 Femoral head left DSC = 0.92 Femoral head right DSC = 0.92 Rectal CTV DSC = 0.88
2D U-Net	CT	85	Prostate DSC = 0.88 [45] Bladder DSC = 0.95 Rectum DSC = 0.92
2D + 3D U-Net	CT	136	Prostate DSC = 0.90 [46] Bladder DSC = 0.95 Rectum DSC = 0.84 Femoral head left DSC = 0.96 Femoral head right DSC = 0.95
Model-based + seeded region-growing	MR	24	Prostate VO = 0.78 [67] Bladder VO = 0.88 Rectum VO = 0.78
Model-based or atlas-based or classic machine learning based or hybrid			Bladder DSC = [0.72, 0.93] [27] Bowel DSC = [0.83, 0.88] Femoral heads DSC = [0.92, 0.94] Rectum DSC = [0.75, 0.77], 0.84 ± 0.08 Spinal cord DSC = [0.69, 0.81]
Atlas-based	CT	26	Prostate DSC = 0.62 [26] Bladder DSC = 0.80 Rectum DSC = 0.66 Hips DSC = 0.69
B-spline registration	CT	10	Bladder DSC > 0.8, HD ≈ 2 [68] Uterus DSC > 0.8, HD ≈ 2 Rectum DSC > 0.8, HD ≈ 2
Our method	CT	105	Bladder DSC = 0.92, HD = 5.18 Bone marrow DSC = 0.85, HD = 1.99 Rectum DSC = 0.79, HD = 5.95 Small intestine DSC = 0.83, HD = 5.28 Spinal cord DSC = 0.83, HD = 3.27 Femoral head left DSC = 0.91, HD = 1.39 Femoral head right DSC = 0.90, HD = 1.44

DSC, dice similarity coefficient; HD, Hausdorff distance (in mm); VO, volume overlap ($|X \cap Y|/|X \cup Y|$).

development in radiotherapy [16]. Our proposed method has comparable performance as other deep learning methods in bladder, femoral head left and femoral head right, superior performance in intestine, but inferior performance in rectum. After investigation, we found 2 special data in our test dataset, gauzes were plugged in the cervix, thus led to large rectum segmentation error. The mean DSC value increased to 0.85 if we removed these two data from the test dataset. The DSC values from [46] seems higher than other CNN based models due to the usage of 3D U-Net. Theoretically, a 3D U-Net based model has better ability to extract 3D structured features. It's predictable that we would further improve our model performance if we extend our model to 3D in future studies. On the other hand, we used 2D DSC value to evaluate the algorithm slice by slice, which is lower than 3D DSC value that considers all slices together as a whole volume, as oncologists also evaluate the contours in 2D.

An oncologist evaluated the model and reached the conclusion that the predicted results from our proposed model were highly acceptable. As mentioned in Table 2, 55.64% of the segmentations in the test set are acceptable without any revision and 34.70% need minor revisions. Moreover, the model is deterministic, which means that it can provide consistent segmentation results and avoid the inter-observer and intra-observer variabilities of manual OARs delineation.

Although the oncologist's evaluation result demonstrates that more than 90% segmentation results are highly acceptable to be used in radiation therapy planning. There are still 9.66% segmentation results need major revisions or are unacceptable. As shown in Table 2, most of the poorly performed organs are small intestine and rectum. We analysed the ground truth masks of these two organs and found they had massive diversity of sizes, shapes and locations for different patients and even for different CT scans from a same patient. As radiotherapy instructions, the rectum and colon must be completely empty before CT scanning. Also the CT scanning time should be precise enough to acquire the images with contrast enhanced properly. However, it is hard to control the rectum and small intestine conditions in real world. Furthermore, some part of the rectum and most part of the small intestine, have very unclear boundaries that even an experienced radiotherapy oncologist have to delineate the boundary by experience instead of by the images. While different oncologists may have different preference when delineating a same small intestine's contour. Comparatively, organs with bone structure were much more consistent such as femoral heads and bone marrow, thus our proposed model can achieve better performance. Consequently, less segmentation results of small intestine or rectum were scored as acceptable.

OARs have various sizes, shapes, and locations and some of them may have several isolated regions and unclear boundaries, which makes it a challenging segmentation task. Even though we achieved promising results for the test set, it is still necessary to collect more data from more different data sources. The data that we used in this study were from only one hospital, which means that the model may not apply to data that are scanned in different situations. A larger dataset and more data sources can make the model more robust and have a better generalization capability.

5. Conclusion

OARs segmentation is a time consuming but key step for radiotherapy planning. We proposed a deep learning-based model that can process one slice in 0.06 s when running on a GTX 1080 GPU. The model is deterministic and can provide consistent results. To overcome the challenges of OARs segmentation on patients with cervical cancer, the model used context aggregation blocks to gain more semantic contextual features and reduce the number of false positives and used reweighted channel importance with Squeeze-Extract blocks. The experiments and oncologist's evaluations demonstrate the effectiveness and efficiency of the proposed method.

Acknowledgement

The work that was described in this paper was supported by a grant from the Ministry of Science and Technology of the People's Republic of China (Grant Number 2016YFC0105207).

References

- [1] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;68:394–424. <https://doi.org/10.3322/caac.21492>.
- [2] Delaney G, Jacob S, Featherstone C, Barton M. The role of radiotherapy in cancer treatment: estimating optimal utilization from a review of evidence-based clinical guidelines. *Cancer* 2005;104(6):1129–37.
- [3] Chen KQ. Systematic evaluation of atlas-based auto segmentation (ABAS) software for adaptive radiation therapy in cervical cancer. *China J Radio Med print* 2015;35(2):111–3.
- [4] Weiss E, Richter S, Krauss T, Metzelthn SI, Hille A, Pradier O, et al. Conformal radiotherapy planning of cervix carcinoma: differences in the delineation of the clinical target volume: a comparison between gynaecologic and radiation oncologists. *Radiother Oncol* 2003;67:87–95.
- [5] Geets X, Daisne JF, Arcangeli S, Coche E, Poel de M, Duprez T, 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.
- [6] Saarnak AE, Boersma M, Van Bunnigen BFM, Wolterink R, Steggerda MJ. Inter-observer variation in delineation of bladder and rectum contours for brachytherapy of cervical cancer. *Radiother Oncol* 2000;56:37–42. [https://doi.org/10.1016/S0167-8140\(00\)00185-7](https://doi.org/10.1016/S0167-8140(00)00185-7).
- [7] Nelms BE, Tomé WA, Robinson G, Wheeler J. Variations in the contouring of organs at risk: test case from a patient with oropharyngeal cancer. *Int J Radiat Oncol Biol Phys* 2012 Jan 1;82(1):368–78.
- [8] Vinod SK, Jameson MG, Min M, Holloway LC. Uncertainties in volume delineation in radiation oncology: a systematic review and recommendations for future studies. *Radiother Oncol* 2016;121(2):169–79.
- [9] Vinod SK, Min M, Jameson MG, Holloway LC. 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.
- [10] Mukesh M, Benson R, Jena R, Hoole A, Roques T, Scrase C, 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):e530–6.
- [11] M van H. Errors and margins in radiotherapy. *Seminars in radiation oncology* 2004;14(1):52–64.
- [12] Hong T, Tome W, Chappell R, Harari P. Variations in target delineation for head and neck IMRT: a international multi-institutional study. *Int J Radiat Oncol Biol Phys* 2004;60:S157–8.
- [13] Daisne J, Blumhofer A. Atlas-based automatic segmentation of head and neck organs at risk and nodal target volumes: a clinical validation. *Radiat Oncol* 2013;8:154.
- [14] Costa Ma J, Delingette H, Novellas Sb, Ayache N. Automatic segmentation of bladder and prostate using coupled 3D deformable models. *Med Image Comput Comput Assist Interv* 2007;10:252–60.
- [15] Acosta O, Dowling J, Drean G, Simon A, de Crevoisier R, Haigron P. Multi-atlas-based segmentation of pelvic structures from CT scans for planning in prostate cancer radiotherapy. In: El-Baz AS, Saba L, Suri J, editors. *Abdomen Thorac Imaging An Eng Clin Perspect*, Boston, MA: Springer US; 2014, p. 623–56. doi:10.1007/978-1-4614-8498-1_24.
- [16] Cardenas CE, Yang J, Anderson BM, Court LE, Brock KB. Advances in auto-segmentation. *Semin Radiat Oncol* 2019;29:185–97. <https://doi.org/10.1016/j.semradi.2019.02.001>.
- [17] Wong WKH, Leung LHT, Kwong DLW. Evaluation and optimization of the parameters used in multiple-atlas-based segmentation of prostate cancers in radiation therapy. *Br J Radiol* 2016;89:20140732. <https://doi.org/10.1259/bjr.20140732>.
- [18] Plassard AJ, McHugo M, Heckers S, Landman BA. Multi-scale hippocampal parcellation improves atlas-based segmentation accuracy. *Proc SPIE–the Int Soc Opt Eng* 2017;10133. <https://doi.org/10.1117/12.2254425>.
- [19] Tourbier S, Hagmann P, Cagneaux M, Guibaud L, Gorthi S, Schaer M, et al. Automatic brain extraction in fetal MRI using multi-atlas-based segmentation. *Proc SPIE* 2015;9413.
- [20] Court L, McCarroll R, Kisling K, Zhang L, Yang J, Jhingran A, et al. Full automation of radiation therapy treatment planning. *Phys Med* 2016;O20(32):147. <https://doi.org/10.1016/j.ejmp.2016.07.028>.
- [21] Fortunati V, Verhaart RF, Van Der Lijn F, Niessen WJ, Veenland JF, Paulides MM, et al. Tissue segmentation of head and neck CT images for treatment planning: a multiatlas approach combined with intensity modeling. *Med Phys* 2013;40. doi:10.1118/1.4810971., 19.
- [22] Tao C, Yi J, Chen N, Ren W, Cheng J, Tung S, et al. Multi-subject atlas-based auto-segmentation reduces interobserver variation and improves dosimetric parameter consistency for organs at risk in nasopharyngeal carcinoma: a multi-institution clinical study. *Radiother Oncol* 2015;115:407–11. <https://doi.org/10.1016/j.radonc.2015.05.012>.

- [23] Kosmin M, Ledsam J, Romera-Paredes B, Mendes R, Moinuddin S, de Souza D, et al. Rapid advances in auto-segmentation of organs at risk and target volumes in head and neck cancer. *Radiother Oncol* 2019;135:130–40. <https://doi.org/10.1016/j.radonc.2019.03.004>.
- [24] Nikolov S, Blackwell S, Mendes R, De Fauw J, Meyer C, Hughes C, et al. Deep learning to achieve clinically applicable segmentation of head and neck anatomy for radiotherapy 2018:1–31. arXiv:1809.04430v1.
- [25] Sims R, Isambert A, Grégoire V, Bidault F, Fresco L, Sage J, et al. A pre-clinical assessment of an atlas-based automatic segmentation tool for the head and neck. *Radiother Oncol* 2009;93:474–8. <https://doi.org/10.1016/j.radonc.2009.08.013>.
- [26] Ayadi M, Delpon G, Lafay F, Meyer P. Evaluation of ABAS TM: multi-center study in the case of prostate cancer. *Phys Med* 2011;27. <https://doi.org/10.1016/j.ejmp.2011.06.032>.
- [27] Sharp G, Fritscher KD, Pekar V, Peroni M, Shusharina N, Veeraraghavan H, et al. Vision 20/20: perspectives on automated image segmentation for radiotherapy. *Med Phys* 2014;41:50902. <https://doi.org/10.1118/1.4871620>.
- [28] Aljabar P, Heckemann RA, Hammers A, Hajnal JV, Rueckert D. Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. *Neuroimage* 2009;46(3):726–38.
- [29] Long J, Shelhamer E. Fully convolutional networks for semantic segmentation. *IEEE Trans Pattern Anal Mach Intell* 2017;39(4):640–51. <https://doi.org/10.1109/TPAMI.2016.2572683>.
- [30] Badrinarayanan V, Kendall A, Cipolla R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans Pattern Anal Mach Intell* 2017;39:2481–95. <https://doi.org/10.1109/TPAMI.2016.2644615>.
- [31] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF, editors. *Med Image Comput Comput Interv – MICCAI 2015*, Cham: Springer International Publishing; 2015, p. 234–41.
- [32] Jegou S, Drozdal M, Vazquez D, Romero A, Bengio Y. The one hundred layers tiramisu: fully convolutional densenets for semantic segmentation. *IEEE Comput Soc Conf Comput Vis Pattern Recognit Work* 2017 2017:1175–83. <https://doi.org/10.1109/CVPRW.2017.156>.
- [33] Zhao H, Shi J, Qi X, Wang X, Jia J. [PSPNet]pyramid scene parsing network — supplementary material 2012:7–8. doi:10.1109/CVPR.2017.660.
- [34] Chen LC, Zhu Y, Papandreou G, Schroff F, Adam H. Encoder-decoder with atrous separable convolution for semantic image segmentation. *LNCS Lect Notes Comput Sci (Including Subser Lect Notes Artif Intell Lect Notes Bioinformatics)* 2018;11211:833–51. https://doi.org/10.1007/978-3-030-01234-2_49.
- [35] Ibragimov B, Xing L. Segmentation of organs-at-risks in head and neck CT images using convolutional neural networks. *Med Phys* 2017;44:547–57. <https://doi.org/10.1002/mp.12045>.
- [36] Chan JW, Kearney V, Haaf S, Wu S, Bogdanov M, Reddick M, et al. A convolutional neural network algorithm for automatic segmentation of head and neck organs at risk using deep lifelong learning. *Med Phys* 2019;46:2204–13. <https://doi.org/10.1002/mp.13495>.
- [37] 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;45:4558–67. <https://doi.org/10.1002/mp.13147>.
- [38] Feng X, Qing K, Tustison NJ, Meyer CH, Chen Q. Deep convolutional neural network for segmentation of thoracic organs-at-risk using cropped 3D images. *Med Phys* 2019;46(5):2169–80. <https://doi.org/10.1002/mp.13466>.
- [39] Dong X, Lei Y, Wang T, Thomas M, Tang L, Curran WJ, et al. Automatic multiorgan segmentation in thorax CT images using U-net-GAN. *Med Phys* 2019;46(5):2157–68.
- [40] Yang J, Veeraraghavan H, Armato 3rd SG, Farahani K, Kirby JS, Kalpathy-Kramer J, et al. Autosegmentation for thoracic radiation treatment planning: a grand challenge at AAPM 2017. *Med Phys* 2018;45(10):4568–81.
- [41] Fu YB, Mazur TR, Wu X, Liu S, Chang X, Lu YG, et al. A novel MRI segmentation method using CNN-based correction network for MRI-guided adaptive radiotherapy. *Med Phys* 2018;45(11):5129–37.
- [42] Men K, Dai JR, Li YX. Automatic segmentation of the clinical target volume and organs at risk in the planning CT for rectal cancer using deep dilated convolutional neural networks. *Med Phys* 2017;44(12):6377–89.
- [43] Dolz J, Xu X, Rony J, Yuan J, Liu Y, Granger E, et al. Multi-region segmentation of bladder cancer structures in MRI with progressive dilated convolutional networks. 2018. arXiv preprint arXiv:1805.10720.
- [44] van der Heyden B, Wohlfahrt P, Eekers DBP, Richter C, Terhaag K, Troost EGC, et al. Dual-energy CT for automatic organs-at-risk segmentation in brain-tumor patients using a multi-atlas and deep-learning approach. *Sci Rep* 2019;9:1–9. <https://doi.org/10.1038/s41598-019-40584-9>.
- [45] Kazemifar S, Balagopal A, Nguyen D, McGuire S, Hannan R, Jiang S, et al. Segmentation of the prostate and organs at risk in male pelvic CT images using deep learning. *Biomed Phys Eng Express* 2018;4. <https://doi.org/10.1088/2057-1976/aad100>.
- [46] Balagopal A, Kazemifar S, Nguyen D, Lin MH, Hannan R, Owringi A, et al. Fully automated organ segmentation in male pelvic CT images. *Phys Med Biol* 2018;63. <https://doi.org/10.1088/1361-6560/aaf11c>.
- [47] Wang C, Connolly B, de Oliveira Lopes PF, Frangi AF, Smedby Ö. Pelvis Segmentation using multi-pass U-net and iterative shape estimation BT – computational methods and clinical applications in musculoskeletal imaging. In: Vrtovec T, Yao J, Zheng G, Pozo JM, editors., Cham: Springer International Publishing; 2019, p. 49–57.
- [48] Hu J, Shen L, Albanie S, Sun G, Wu E. Squeeze-and-excitation networks. *IEEE Trans Pattern Anal Mach Intell* 2019;1. doi:10.1109/TPAMI.2019.2913372.
- [49] De Fauw J, Ledsam JR, Romera-Paredes B, Nikolov S, Tomasev N, Blackwell S, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat Med* 2018;24:1342–50. <https://doi.org/10.1038/s41591-018-0107-6>.
- [50] Jiang H, Rong R, Wu J, Li X, Dong X, Chen EZ. Skin lesion segmentation with improved C-UNet networks. *BioRxiv* 2018;382549. <https://doi.org/10.1101/382549>.
- [51] Li X, Chen H, Qi X, Dou Q, Fu CW, Heng PA. H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes. *IEEE Trans Med Imaging* 2018;37(12):2663–74.
- [52] Lachinov D, Vasiliev E, Turlapov V. Glioma segmentation with cascaded Unet. 2018. ArXiv Preprint ArXiv:1810.04008.
- [53] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:1511.07122 (2015).
- [54] Peng C, Zhang X, Yu G, Luo G, Sun J. Large Kernel matters — improve semantic segmentation by global convolutional network, 2017, p. 1743–51. doi:10.1109/CVPR.2017.189.
- [55] Oktay O, Schlemper J, Folgoc L Le, Lee M, Heinrich M, Misawa K, et al. Attention U-Net : learning where to look for the pancreas. *MIDL* 2018. arXiv preprint arXiv:1804.03999.
- [56] Kamnitsas K, Bai W, Ferrante E, McDonagh S, Sinclair M, Pawlowski N, et al. Ensembles of multiple models and architectures for robust brain tumour segmentation. In: Crimi A, Bakas S, Kuijf H, Menze B, Reyes M, editors. *Brainlesion Glioma, Mult. Sclerosis, Stroke Trauma*. Brain Inj., Cham: Springer International Publishing; 2018, p. 450–62.
- [57] Wang ZY, Ji SW. Smoothed dilated convolutions for improved dense prediction. *ACM* 2018. arXiv:1808.08931.
- [58] Paszke A, Chanan G, Lin Z, Gross S, Yang E, Antiga L, et al. Automatic differentiation in PyTorch 2017:1–4.
- [59] Kingma DP, Jimmy B. Adam: a method for stochastic optimization. 2014. arXiv preprint arXiv:1412.6980.
- [60] Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. *AISTATS* 2010.
- [61] Ioffe S, Christian S. Batch normalization: accelerating deep network training by reducing internal covariate shift. 2015. arXiv preprint arXiv:1502.03167.
- [62] Nair V, Hinton GE. Rectified linear units improve restricted Boltzmann machines. *Proc. 27th Int. Conf. Int. Conf. Mach. Learn., USA: Omnipress*; 2010, p. 807–14.
- [63] Dice LR. Measures of the amount of ecologic association between species. *Ecology* 1945;26:297–302. <https://doi.org/10.2307/1932409>.
- [64] Huttenlocher DP, Klanderman GA, Rucklidge WJ. Comparing images using the Hausdorff distance. *IEEE Trans Pattern Anal Mach Intell* 1993;15:850–63. <https://doi.org/10.1109/34.232073>.
- [65] Raghu M, Zhang CY, Kleinberg J, Bengio S. Transfusion: understanding transfer learning for medical imaging. 2019. arXiv:1902.07208.
- [66] Men K, Zhang T, Chen X, Chen B, Tang Y, Wang S, 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;50:13–9. <https://doi.org/10.1016/j.ejmp.2018.05.006>.
- [67] Pasquier D, Lacomberie T, Vermandel M, Rousseau J, Lartigau E, Betrouni N. Automatic segmentation of pelvic structures from magnetic resonance images for prostate cancer radiotherapy. *Int J Radiat Oncol Biol Phys* 2007;68:592–600. <https://doi.org/10.1016/j.ijrobp.2007.02.005>.
- [68] Langerak T, Heijkoop S, Quint S, Mens JW, Heijmen B, Hoogeman M. Towards automatic plan selection for radiotherapy of cervical cancer by fast automatic segmentation of cone beam CT scans. *Proc Med Image Comput Comput Assisted Intervention* 2014:528–35.