

Henry Ford Health System

Henry Ford Health System Scholarly Commons

Radiation Oncology Articles

Radiation Oncology

2-8-2020

Segmentation of the Prostatic Gland and the Intraprostatic Lesions on Multiparametric Magnetic Resonance Imaging Using Mask Region-Based Convolutional Neural Networks

Zhenzhen Dai

Eric Carver

Chang Liu

James Lee

Wei Wei Zong

See next page for additional authors

Follow this and additional works at: https://scholarlycommons.henryford.com/radiationoncology_articles

Authors

Zhenzhen Dai, Eric Carver, Chang Liu, James Lee, Wei Wei Zong, Milan V. Pantelic, Mohamed A. Elshaikh, and Ning Wen

Scientific Article

Segmentation of the Prostatic Gland and the Intraprostatic Lesions on Multiparametric Magnetic Resonance Imaging Using Mask Region-Based Convolutional Neural Networks

Zhenzhen Dai, MS,^a Eric Carver, PhD,^b Chang Liu, PhD,^a
Joon Lee, MD,^a Aharon Feldman, MD,^a Weiwei Zong, PhD,^a
Milan Pantelic, MD,^b Mohamed Elshaikh, MD,^a and Ning Wen, PhD^{a,*}

^aDepartment of Radiation Oncology, Henry Ford Health System, Detroit, Michigan; and ^bDepartment of Diagnostic Radiology, Henry Ford Health System, Detroit, Michigan

Received 8 July 2019; revised 15 December 2019; accepted 19 January 2020

Abstract

Purpose: Accurate delineation of the prostate gland and intraprostatic lesions (ILs) is essential for prostate cancer dose-escalated radiation therapy. The aim of this study was to develop a sophisticated deep neural network approach to magnetic resonance image analysis that will help IL detection and delineation for clinicians.

Methods and Materials: We trained and evaluated mask region-based convolutional neural networks to perform the prostate gland and IL segmentation. There were 2 cohorts in this study: 78 public patients (cohort 1) and 42 private patients from our institution (cohort 2). Prostate gland segmentation was performed using T2-weighted images (T2WIs), although IL segmentation was performed using T2WIs and coregistered apparent diffusion coefficient maps with prostate patches cropped out. The IL segmentation model was extended to select 5 highly suspicious volumetric lesions within the entire prostate.

Results: The mask region-based convolutional neural networks model was able to segment the prostate with dice similarity coefficient (DSC) of 0.88 ± 0.04 , 0.86 ± 0.04 , and 0.82 ± 0.05 ; sensitivity (Sens.) of 0.93, 0.95, and 0.95; and specificity (Spec.) of 0.98, 0.85, and 0.90. However, ILs were segmented with DSC of 0.62 ± 0.17 , 0.59 ± 0.14 , and 0.38 ± 0.19 ; Sens. of 0.55 ± 0.30 , 0.63 ± 0.28 , and 0.22 ± 0.24 ; and Spec. of 0.974 ± 0.010 , 0.964 ± 0.015 , and 0.972 ± 0.015 in public validation/public testing/private testing patients when trained with patients from cohort 1 only. When trained with patients from both cohorts, the values were as follows: DSC of 0.64 ± 0.11 , 0.56 ± 0.15 , and 0.46 ± 0.15 ; Sens. of 0.57 ± 0.23 , 0.50 ± 0.28 , and 0.33 ± 0.17 ; and Spec. of 0.980 ± 0.009 , 0.969 ± 0.016 , and 0.977 ± 0.013 .

Conclusions: Our research framework is able to perform as an end-to-end system that automatically segmented the prostate gland and identified and delineated highly suspicious ILs within the entire prostate. Therefore, this system demonstrated the potential for assisting the clinicians in tumor delineation.

© 2020 The Author(s). Published by Elsevier Inc. on behalf of American Society for Radiation Oncology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Sources of support: This work was supported by a Research Scholar Grant, RSG-15-137-01-CCE from the American Cancer Society.

Disclosures: The authors declare no competing interests. Dr Wen reports grants from American Cancer Society, during the conduct of the study, personal fees from Varian Medical System, and personal fees from Viewray, outside the submitted work.

Research data are stored in an institutional repository and will be shared upon request to the corresponding author.

* Corresponding author: Ning Wen PhD; E-mail: NWEN1@hfhs.org.

<https://doi.org/10.1016/j.adro.2020.01.005>

2452-1094/© 2020 The Author(s). Published by Elsevier Inc. on behalf of American Society for Radiation Oncology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

Prostate cancer (PCa) is the most common cancer in men in the United States,¹ with an estimated 174,650 new cases and 31,620 deaths expected in 2019.² Radiation therapy (RT) is an effective form of prostate cancer (PCa) treatment and is considered one of the standard treatment options available. Current practice treats the entire prostate with a homogeneous dose distribution.^{3,4} Dose-escalated RT has been shown to improve biochemical progression-free survival at the expense of increased acute and late toxicities.⁵ However, using simultaneous boost technique that limits the volume of dose-escalation to intraprostatic lesions (ILs) may allow for improved dosimetry with a potential to improve the therapeutic ratio as well. A median dose to the entire gland could prevent disease recurrence in the prostate from satellite tumors and significantly reduce the side effects associated with escalated radiation dose to the entire gland. A boosting dose to the ILs can maintain the effectiveness of focal therapy to treat the ILs that are the main determinants of tumor progression and prognosis. For this strategy to be successful, the key requirements included the ability to accurately and reliably identify clinically significant tumors within the prostate gland.

Malignant ILs can be identified on magnetic resonance imaging and are known to correlate with tumor aggressiveness. Multiparametric magnetic resonance imaging (mp-MRI) combines anatomic T1- and T2-weighted imaging (T2WI) with diffusion-weighted (DWI) and perfusion-weighted sequences and plays an essential role in the diagnosis, risk stratification, staging, and treatment guidance in PCa.⁶ Over the decades, significant advancements have been made in image acquisition technologies (including mp-magnetic resonance imaging [MRI]). These allowed for the visualization of not only structural anatomy but also vascular and functional properties of the prostate gland. However, the large amount of data has hindered the reproducibility and efficacy of image interpretation.⁷ In addition, inter-observer variability and clinician fatigue⁸ can restrict accurate interpretations before therapeutic interventions. Therefore, computer-aided diagnostic systems have been established to help improve clinical practice with MRI-based automated prostate and IL segmentation.

Recent progress in image segmentation has involved convolutional neural network (CNN) based models. Many segmentation models fall into 2 classes. The first class does not rely on the region proposal algorithm. U-Net,⁹ for example, is a classic model widely used in biomedical image segmentation tasks. The underlying fundamentals of the second class rely on region proposals such as the mask region-based CNN (Mask R-CNN) model. It is widely used in semantic segmentation, object localization, and instance segmentation of natural images.¹⁰

Methods and Materials

Patient cohorts

A total of 120 patients were divided into 2 cohorts in this study. Cohort 1 (public patients) included 78 randomly selected patients from the International Society for Optics and Photonics-American Association of Physicists in Medicine-National Cancer Institute Prostate MR Gleason Grade Group Challenge (PROSTATEx-2 Challenge), which is a data set of prostate MRI studies conducted by American Association of Physicists in Medicine along with the International Society for Optics and Photonics and the National Cancer Institute.^{11–13} Each patient was read under the supervision of an expert radiologist with more than 20 years of experience. Areas of suspicion were indicated by the radiologist using a point marker. Then MR-guided biopsy was followed in the suspicious area. Confirmation scans were performed with biopsy needle in situ to conform the accurate localization. Images were acquired on 2 types of Siemens 3 Tesla MR scanners, the MAGNETOM Trio and Skyra. Axial T2WIs were acquired using a turbo spin echo sequence with a resolution of around 0.5 mm in plane and slice thickness of 3.6 mm. Axial DWI sequences were acquired using a single-shot echo planar imaging sequence with a resolution of 2 mm in plane and 3.6 mm slice thickness and with diffusion-encoding gradients in 3 directions. Three b-values were acquired (50, 400, and 800 s/mm²), and the apparent diffusion coefficient (ADC) map was calculated using the scanner software. All images were acquired without an endorectal coil. Each patient from cohort 1 was proved to have only one lesion.

Cohort 2 (private patients) included 42 patients who underwent mp-MRI scans at our institution. A transrectal ultrasound-guided fine-needle biopsy was performed to confirm the presence of PCa. All images were acquired using a 3 Tesla MR scanner (Ingenia; Philips Medical System, Best, the Netherlands). Axial T2WIs were obtained using fast spin echo (TE/TR: 4389/110 ms, flip angle: 90 degrees) with a resolution of 0.42 mm in plane and slice thickness of 2.4 mm. Axial DWIs were obtained (TE/TR: 4000/85 ms, flip angle: 90 degrees) with a resolution of 1.79 mm in plane and slice thickness of 0.56 mm. The voxel-wise ADC map was constructed using 2 DWIs with 2 b-values (0, 1000 s/mm²). The radiologists annotated the ILs in the MR images during the diagnosis, and the targeted biopsy was performed with the MRI-defined lesions superimposed using the T2WI on the transrectal ultrasound images. The ILs were delineated by 3 different clinicians from our institute with the reference to both radiology and pathology reports for consistency.

Three data sets are traditionally used to build deep neural network-based models. The training set is used to fit the model; the validation set provides an unbiased

evaluation of model fit in selecting model and tuning model hyper-parameters, and the testing set provides an unbiased evaluation in the final model fit. For prostate segmentation, 54 public patients (1085 slices in total and an average of 20.1 ± 1.4 slices per patient) were randomly selected as the training set. The model was validated using 12 public patients and tested using 12 public patients and 16 private patients, respectively. Compared with the prostate, ILs have more variance in shape, size, and location. Thus, more patient samples were included in the testing set to evaluate the robustness of the model performance on the IL delineation. The model was trained using 45 public patients (614 slices in total and an average of 13.6 ± 2.4 slices per patient) and was validated with 10 public patients and tested with 23 public patients and 42 private patients. We also trained the model with mix cohorts of the same 45 public patients and 21 additional private patients (314 slices in total and an average of 15.0 ± 5.3 slices per patient) to compare the model performances.

The long dimension of T2 images was resized to 384, and the short dimension was resized to keep the same width-to-length ratio as the original image and padded by pixels with zero value to extend to the same size of 384. The ADC map was resampled using bilinear interpolation and rigidly registered to T2WI using a software developed in house. All images were normalized slice by slice before having the histogram equalized. For prostate segmentation, we used T2WI only in training, validation, and testing, as the prostate gland can be well-defined by the morphologic imaging. For IL segmentation, only the region of the prostate was analyzed that prostate patches were cropped out based on the results of the prostate segmentation. Coregistered ADC map and T2WI were combined as the model inputs (the first channel was T2WI and the second one was the ADC map).

Extension of the model into 3-dimensional space

Figure 1 shows the 2D Mask R-CNN architecture used in our paper. A detailed description of network architecture and implementation of the Mask R-CNN model is provided in Appendix 1 (available online at <https://doi.org/10.1016/j.adro.2020.01.005>). in the supplement. Many models based on 3-dimensional (3D) CNNs^{14–16} are available for volumetric segmentation. However, we found it difficult to directly extend the Mask R-CNN model to perform 3D prostate and IL segmentation as training was highly resource intensive. Therefore, we developed a method to extend the model to perform a more efficient volumetric segmentation.

Prostate segmentation

We found that 2-dimensional (2D) Mask R-CNN introduced false positive in superior/inferior slices beyond the

prostate gland within the pelvic region and false negative in the apex and base regions of the prostate gland. To train the model for prostate segmentation, a smallest rectangle area covering the entire prostate gland to the maximum was estimated based on the training set. Images with the prostate gland were masked by the clinician's contour and were labeled as "prostate"; the remaining images were masked by the rectangle area and were labeled as "nonprostate" during training. For each slice, delineated contours by the model were scored by the probability of being "prostate" or "nonprostate," and the highest scoring contour and its label was regarded as the final result. This significantly decreased the false-positive and false-negative rate.

Intraprostatic lesion segmentation

Because prostate cancer is multifocal disease and it is unknown how many ILs each patient has, we designed the Mask R-CNN model to perform multifocal PCa segmentation and selected 5 top suspicious ILs for each patient, which is consistent with the procedure for MRI-guided prostate biopsy. Each IL was located with a bounding box and for probability of being a lesion. The lesion then was delineated on each slice within the bounding box. Because multiple ILs could be detected on the same slice, we set a threshold to determine whether these delineated lesions were represented separate entities or the same lesion. When the dice similarity coefficient (DSC) of 2 contours was greater than 0.5, these 2 contours and their bounding boxes were unified into new ones, and their probability scores were averaged. After this step, contours on all slices were ranked by their scores and the highest scoring contour was selected as a seed contour. Next we only selected the contours on the adjacent slices having the highest DSC with the seed contour, and both the DSC and its score were expected to be larger than the cutoff thresholds. The selected adjacent contour was regarded as the new seed contour and was used to find its next superior/inferior adjacent lesion. This process was iterated until all slices were traversed or no adjacent contour could be identified. Then these selected seed contours were combined together to define the IL volume. Algorithm 1 (available online at <https://doi.org/10.1016/j.adro.2020.01.005>). in supplement describes the definition of the most suspicious volumetric IL after contour union in detail. The same procedure was repeated to define the volume of the next IL until 5 suspicious lesions were delineated in total. The cutoff threshold of score and DSC were 0.7 and 0.41, respectively, which were fine-tuned on the validation set.

Evaluation metrics

The metrics used to evaluate the model included the DSC, 95th percentile Hausdorff distance (HD), sensitivity

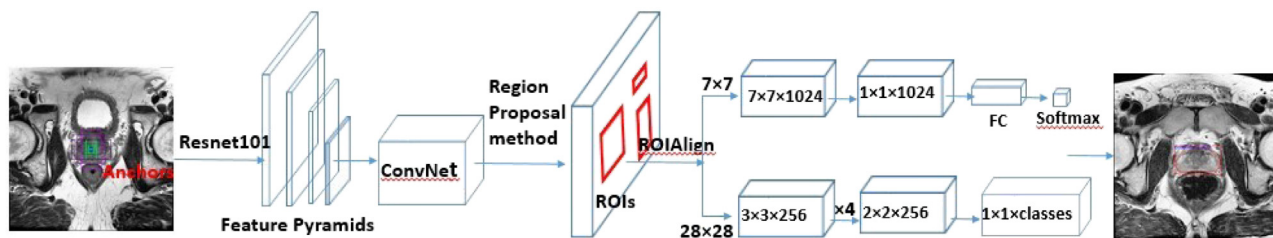


Figure 1 General Mask-RCNN network architecture used in our paper. Predefined anchors with different scales at one location are shown as purple bounding boxes on the input image. Cubes are represented by kernel size \times kernel size \times number of filters, above branch is used for classification, bottom branch is used for segmentation.

(Sens.), and specificity (Spec.). DSC (Equation 1) evaluates how well 2 binary sets match with a DSC of 1 representing a perfect match of prediction and contoured mask. HD (Equation 2) measures how far apart 2 subsets within a metric space are from each other. Sens. (Equation 4), also known as the true positive rate, measures the proportion of actual positives that are correctly identified; Spec. (Equation 3), also known as the true negative rate, measures the proportion of actual negatives that are correctly identified. We also defined the agreement rate (Equation 5) as a degree to which model's segmentation results concur with clinicians' segmentation results. The formulas for these metrics are shown below:

$$DSC = \frac{2 \cdot \langle y_{true}, \bar{y}_{pred} \rangle}{\langle y_{true}, y_{true} \rangle + \langle \bar{y}_{pred}, \bar{y}_{pred} \rangle} \quad (1)$$

where y_{true} is the ground truth contour, \bar{y}_{pred} is the model's contour.

$$HD(A, B) = \max_{a \in A} \{ \min_{b \in B} \{ d(a, b) \} \} \quad (2)$$

where a and b are points of sets A and B , respectively; and $d(a, b)$ is the Euclidean distance between these points; 95th percentile HD says that 95% of $\min_{b \in B} \{ d(a, b) \}$ is below this amount. A is the ground truth contour, B is the model's contour.

$$Spec. = \frac{DTN_{object}}{TN_{object}} \quad (3)$$

$$Sens. = \frac{DTP_{object}}{TP_{object}} \quad (4)$$

For prostate segmentation, $DTP_{prostate}$ and $DTN_{prostate}$ denote the number of slices detecting and not detecting the prostate, respectively; and $TP_{prostate}$ and $TN_{prostate}$ denote the total number of slices truly containing the prostate and not containing the prostate, respectively. In addition, 161 positive and 73 negative slices were used in the validation for cohort 1; 156 positive and 84 negative slices in the testing for cohort 1; and 216 positive and 246 negative slices in the testing for cohort 2. For IL segmentation, DTP_{IL} and DTN_{IL} denote the number of pixels in the detected lesion and background, respectively,

whereas TP_{IL} and TN_{IL} denote the total number of pixels within the lesion and background, respectively. Pixels are counted within the entire prostate. DSC was calculated based on 3D volume for both prostate and lesion segmentation.

$$Agreement = \frac{dIL}{cIL} \quad (5)$$

where dIL denotes the number of detected lesions having DSC greater than 0.1 with lesions contoured by clinicians, cIL denotes the number of lesions contoured by clinicians.

Results

Results of prostate segmentation are shown in Table 1. We selected the model that achieved the highest DSC on validation patients and tested the model using both public and private patients. DSC, 95 HD, Sens., and Spec. were 0.88 ± 0.04 , 6.05 ± 2.39 (mm), 0.93, and 0.98, respectively on validation patients; 0.86 ± 0.04 , 6.19 ± 2.38 (mm), 0.95, and 0.85, respectively on public testing patients; and 0.82 ± 0.05 , 8.94 ± 4.09 (mm), 0.95, and 0.90, respectively on private testing patients. Figure 2 shows a sample of prostate segmentation results on T2WIs by clinician and Mask R-CNN.

Results of IL segmentation are shown in Table 2. When training with public patients only, agreement, DSC of detection, Sens., and Spec. were 80%, 0.62 ± 0.17 , 0.55 ± 0.30 , and 0.974 ± 0.010 respectively on validation

Table 1 Results of prostate segmentation

Evaluation	DSC	95 HD (mm)	Sens.	Spec.
12 public validation patients	0.88 ± 0.04	6.05 ± 2.39	0.93	0.98
12 public testing patients	0.86 ± 0.04	6.19 ± 2.38	0.95	0.85
16 private testing patients	0.82 ± 0.05	8.94 ± 4.09	0.95	0.90

Abbreviations: DSC = dice similarity coefficient; Sens. = sensitivity; Spec = specificity.

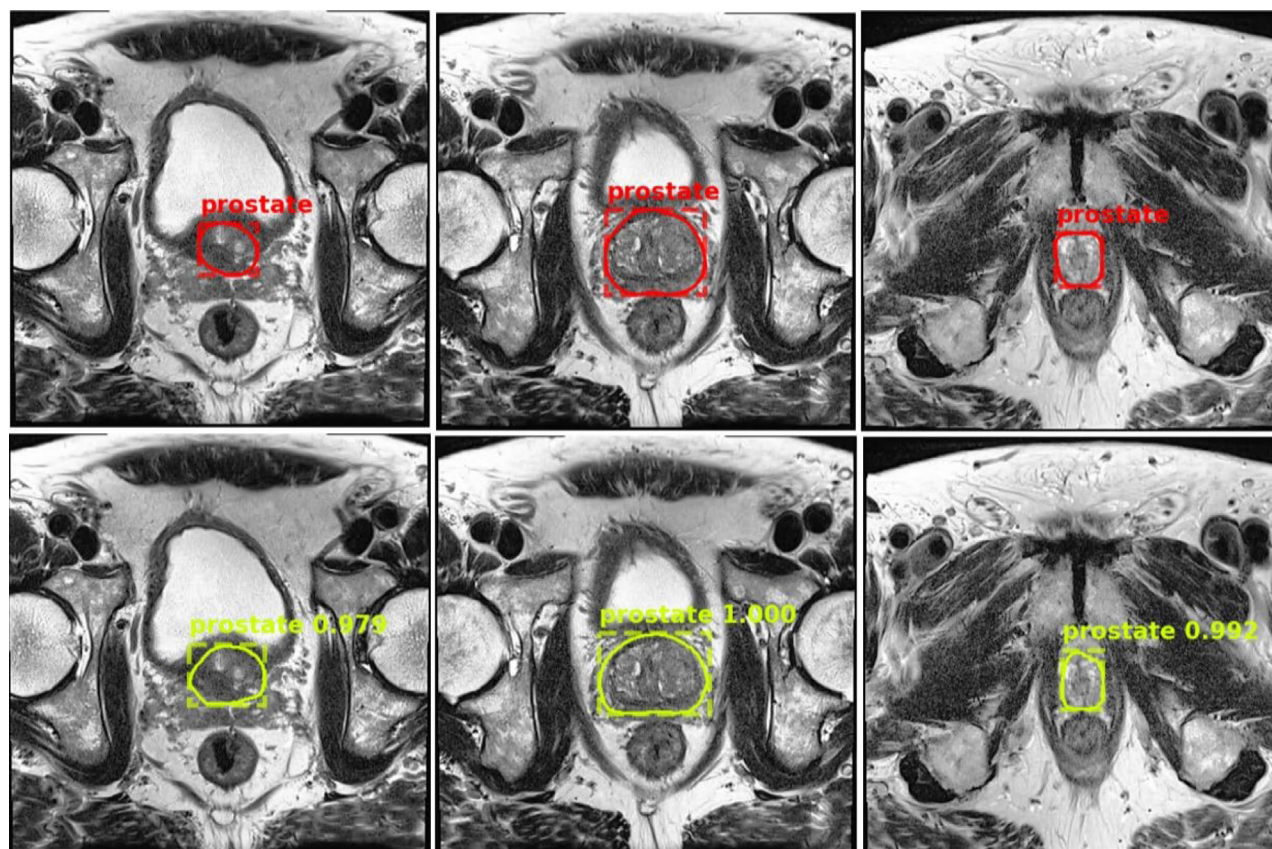


Figure 2 Prostate segmentation results on 3 slices of T2-weighted images from one patient. Ground truth by the clinician (top rows) is shown with prostate contour and bounding box; mask region-based convolutional neural network prediction (bottom rows) is shown with bounding box, prostate contour, and prediction class and score.

patients; 87%, 0.59 ± 0.14 , 0.63 ± 0.28 , and 0.964 ± 0.015 , respectively on public testing patients; and 47%, 0.38 ± 0.19 , 0.22 ± 0.24 , 0.972 ± 0.015 , respectively on private testing patients. When training with mix cohorts of patients, the values were 90%, 0.64 ± 0.11 , 0.57 ± 0.23 , and 0.980 ± 0.009 , respectively, on validation patients; 83%, 0.56 ± 0.15 , 0.50 ± 0.28 , and 0.969 ± 0.016 , respectively on public testing patients; and 63%, 0.46 ± 0.15 , 0.33 ± 0.17 , and 0.977 ± 0.013 , respectively, on private testing patients. [Figure 3](#) shows a sample of IL segmentation results on T2WIs by clinician and Mask R-CNN.

To facilitate a more comprehensive and unbiased evaluation of the Mask R-CNN's performance, we trained a 2D U-Net and a 3D U-Net and calculated the DSCs of the prostate contours using the same 12 testing patients used for the prostate segmentation. DSCs were 0.85 ± 0.03 and 0.83 ± 0.07 using 2D U-Net and 3D U-Net, respectively.

Discussion

Conventional research applied classic machine learning and statistical graph models to prostate and IL

Table 2 Lesion detection and segmentation results

Training	Evaluation	DSC of detection	Agreement	Sens.	Spec.
45 public patients	10 public validation patients	0.62 ± 0.17	80%	0.55 ± 0.30	0.974 ± 0.010
	23 public testing patients	0.59 ± 0.14	87%	0.63 ± 0.28	0.964 ± 0.015
	42 private testing patients	0.38 ± 0.19	47%	0.22 ± 0.24	0.972 ± 0.015
45 public patients + 21 private patients	10 public validation patients	0.64 ± 0.11	90%	0.57 ± 0.23	0.980 ± 0.009
	23 public testing patients	0.56 ± 0.15	83%	0.50 ± 0.28	0.969 ± 0.016
	21 private testing patients	0.46 ± 0.15	63%	0.33 ± 0.17	0.977 ± 0.013

Abbreviations: DSC = dice similarity coefficient; Sens. = sensitivity; Spec = specificity.

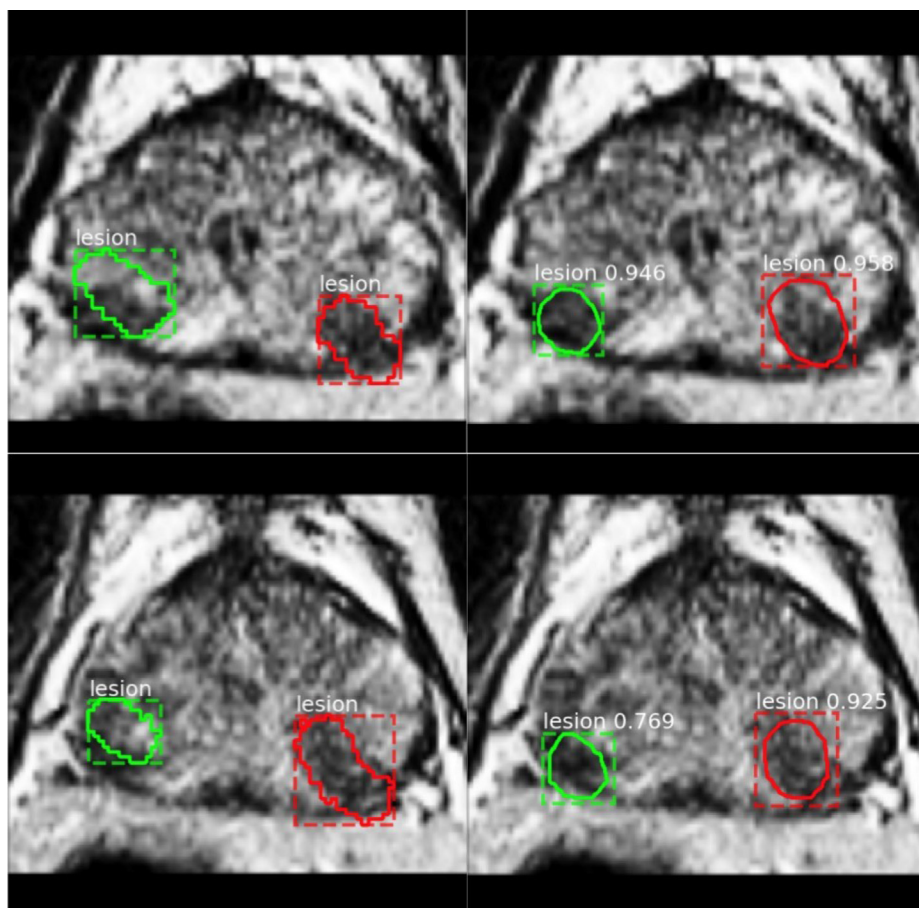


Figure 3 Lesion segmentation on 2 continuous slices from one patient from our institute, with 2 lesion identified and contoured (green and red) on T2-weighted images. Ground truth by the clinician (left column) is shown with lesion contour and bounding box; prediction of 2 agreed candidate lesions by the mask region-based convolutional neural network (right column) is shown with bounding box, lesion contour, and prediction class and score.

segmentation. Neural network–based models for prostate and IL segmentation have only been developed during the past 3 years and are still in need of significant improvements. We performed a literature review (Table 3)¹⁷⁻³⁵ of previous works as our baseline and compared their experiment set ups and results to show the potential of deep neural networks on prostate and IL segmentation.

In this study, we provided an unbiased evaluation of Mask R-CNN's performance using an independent patient cohort from our institution. For prostate segmentation, the validation and testing results showed promise on the PROSTATEx-2 Challenge data set. When using private patients as an independent testing cohort, we observed slightly decreased performance in the DSC, whereas sensitivity and specificity remained equivalent (Table 1). We hypothesized that this is due to variations in prostate delineation among clinicians. Other possible explanations include variations in image quality and small size of data set. Although U-Net is a more elegant fully convolutional network (FCN), it analyzes the entire image as for

segmentation. Mask R-CNN differs from this kind of segmentation model in that it is based on region proposal network, which selects regions of interest and then performs pixel-to-pixel segmentation using FCN on the selected regions of interest. We provided an unbiased comparison between U-Net and Mask R-CNN using the same 12 public testing patients. We compared the DSC calculated by 2D and 3D U-Net with one calculated by the Mask R-CNN and concluded that Mask R-CNN slightly outperformed both U-Nets in prostate segmentation. Mask R-CNN may work better in lesion segmentation than FCN owing to its ability to detect possible lesion patches first, instead of a direct pixel prediction.

The segmentation of ILs is challenging owing to the small volume, prostate tissue heterogeneity, and the often a subtle appearance of tumor, and this leads to interobserver variability in defining the ground truth. We first explored differences between contours by 2 clinicians on the same lesion in 19 patients. The DSC was sensitive to the relative size of the target and was generally low in the evaluation of IL segmentation. The DSC was 0.67 ± 0.21 ,

Table 3 Literature review of publications for the prostate and IL segmentation

Publication	Task	Method	Result	Evaluation
Tian et al ¹⁷	Prostate segmentation	Graph cut	DSC = $87.0\% \pm 3.2\%$	MICCAI 2012 Promise12 challenge
Mahapatra and Buhmann ¹⁸	Prostate segmentation	Super pixel + random forests + graph cut	DSC = 0.81	MICCAI 2012 Promise12 challenge
Guo et al ¹⁹	Prostate segmentation	Stacked sparse auto-encoder + deformable segmentation	DSC = 0.871 ± 0.042	66 T2WIs
Milletari et al ²⁰	Prostate segmentation	V-Net + dice-based loss	DSC = 0.869 ± 0.033	Trained with 50 MRI scans Test with 30 MRI scans
Zhu et al ²¹	Prostate segmentation	Deeply supervised CNN	DSC = 0.885	Trained with 77 patients Tested with 4 patients
Yu et al ¹⁴	Prostate segmentation	Volumetric convolutional neural network	DSC = 89.43%	MICCAI 2012 Promise12 challenge
Toth and Madabhushi ²²	Prostate segmentation	Landmark-free AAM	DSC = $88\% \pm 5\%$	Tested with 108 studies
Liao et al ²³	Prostate segmentation	Stacked independent subspace analysis + sparse label	DSC = $86.7\% \pm 2.2\%$	30 T2WIs
Vincent et al ²⁴	Prostate segmentation	AAM	DSC = 0.88 ± 0.03	MICCAI 2012 Promise12 challenge
Klein et al ²⁵	Prostate segmentation	Atlas matching	Median DSC varied between 0.85 and 0.88	Leave-one-out test with 50 clinical scans
Li et al ²⁶	Prostate segmentation	RW	DSC = $80.7\% \pm 5.1\%$	30 MR volumes
Kohl et al ²⁷	IL segmentation	Adversarial networks	DSC = 0.41 ± 0.28 Sens. = 0.55 ± 0.36 Spec. = 0.98 ± 0.14	Four-fold cross-validation on 55 patients with aggressive tumor lesions
Cameron et al ²⁸	IL detection	Morphology, asymmetry, physiology and size model	Accuracy (Acc.) = $87\% \pm 1\%$ Sens. = $86\% \pm 3\%$ Spec. = $88\% \pm 1\%$	13 patients
Chung et al ²⁹	IL segmentation	Radiomics-driven CRF	Sens. = 71.47% Spec. = 91.93% Acc. = 91.17% DSC = 39.13%	20 patients
Artan et al ³⁰	IL segmentation	Cost-sensitive support vector machine + CRF	Sens. = 0.84 ± 0.19 Spec. = 0.48 ± 0.22 DSC = 0.35 ± 0.18	21 patients
Artan et al ³¹	IL localization	RW	Sens. = 0.51 Jakkard = 0.44	10 patients
Artan et al ³²	IL segmentation	RW	Sens. = 0.62 ± 0.23 Spec. = 0.89 ± 0.10 DSC = 0.57 ± 0.21	16 patients with lesions in peripheral zone only
Ozer et al ³³	IL segmentation	Relevance vector machine	Spec. = 0.78 Sens. = 0.74 DSC = 0.48	20 patients
Artan et al ³⁴	IL segmentation	Cost-sensitive CRF	Sens. = 0.73 ± 0.25 Spec. = 0.75 ± 0.13 Acc. = 0.71 ± 0.18 DSC = 0.45 ± 0.28	10 patients with lesions in peripheral zone only
Liu et al ³⁵	IL segmentation	Fuzzy Markov random fields	Spec. = 89.58% Sens. = 87.50% Acc. = 89.38% DSC = 0.6222	11 patients

Abbreviations: AAM = active appearance models; CNN = convolutional neural network; CRF = conditional random field; DSC = dice similarity coefficient; IL = intraprostatic lesions; MRI = magnetic resonance imaging; RW = random walker; T2WIs = T2-weighted images; Sens. = sensitivity; Spec = specificity.

demonstrating high interobserver variability in defining IL boundaries by trained clinicians. The DSC of our model was 0.59 ± 0.14 when training and testing with the public cohort, which was in the same order of the comparison between clinicians' performances. Most works²⁸⁻³⁵ on IL segmentation were performed with less than 25 clinical scans, and others^{32,34,35} focused on lesions in the peripheral zone only. In our study, we validated and tested the Mask R-CNN model to identify the ILs within the entire prostatic gland using 75 clinical scans from both public and private patient cohorts. To the best of our knowledge, the highest performance was achieved by Liu et al with DSC = 0.6222, Sens. = 87.50%, and Spec. = 89.58%.³⁵ However, they tested a small cohort of 11 patients, and the study was limited to the lesions in the peripheral zone. We identified only one study applying the neural network-based model for IL segmentation²⁷ with DSC = 0.41 ± 0.28 , Sens. = 0.55 ± 0.36 , and Spec. = 0.98 ± 0.14 . Similar to this study, we achieved relatively low Sens. and high Spec. compared with the classic models. There are 2 explanations of the performance of our model: (1) Spec. was calculated based on the whole prostate volume instead of a single MRI slice, leading to the high Spec.; (2) classic models predict more false-positives, whereas Mask R-CNN predicts lesions with size more close to a real lesion. This can potentially facilitate PCa diagnosis and treatment (eg, dose-escalated radiation therapy to the ILs). The performance of Mask R-CNN dropped when training with patients from cohort 1 and testing with cohort 2. We hypothesize this was due to the different acquisition parameters of the DWIs between the 2 patient cohorts, which led to different data distributions between the 2 cohorts. The ADC maps were created from 3 b values (50, 400, and 800 s/mm²) in cohort 1, which were different from the 2 b values (0, 1000 s/mm²) in cohort 2. The variability of the signal intensity values in ADC maps was reduced using the 3-point b values.³⁶ The model generalization is challenging considering MR acquisition and reconstruction techniques are significantly different along with different types of scanners and imaging protocols. Recent research has been working on developing generative models.^{37,38} We investigated how to improve the model generalization. We added a small number of the patient data from cohort 2 and fine-tuned the model using the mixture of cohort 1 and cohort 2 patients. The results showed an increase of the average value and a decrease of the variance of the DSC, Sens., and Spec. when testing with cohort 2. The overall performance of the model was improved, indicating that it was a viable solution to generalize the model by fine-tuning it with a small sample size of an independent cohort.

Definitive radiation therapy for prostate cancer involves treating the entire prostate gland with a homogeneous dose. Newer studies suggest a local control benefit with an additional simultaneous integrated boost to the ILs using a

number of different radiation therapy techniques (eg, IMRT, SBRT, brachytherapy). It is in designing the appropriate IL boost volume that auto segmentation may add value by decreasing physician labor. Assuming physicians have adequate experience in mp-MRI interpretation, they must bring up multiple sequences of the mp-MRI data set into the appropriate image viewer including, at minimum, the T2-weighted and the diffusion-weighted axial images, and scroll through the individual slices to determine the site of the IL. The performance of IL segmentation in the present study may allow radiation oncologists to quickly determine whether or not they are in agreement with the model prediction. The model prediction can also be served as the secondary check for quality control. With a relatively high Dice coefficient, the radiation oncologist may feel comfortable accepting the auto segmentation. In cases where there is little to no agreement between the physician and the auto segmentation (including failure to delineate a target), it would also raise the alarm for the clinician to carefully evaluate and justify the manually delineated DIL contour.

As health care costs continue to rise at a rapid rate, there is mounting pressure to cut costs and to improve efficiency. Contouring is usually the single most time-consuming activity for a radiation oncologist, and advances in auto segmentation may make the process more efficient. Overall, the current performance of IL segmentation is not perfect, but it does offer the possibility of augment a radiation oncologist on the IL delineation. As more data are made available, we may reasonably expect that the performance of IL segmentation will improve over time.

Conclusions

Our research framework is able to work as an end-to-end system. It automatically segmented the prostate gland and identified highly suspicious volumetric lesions within the entire prostate directly from the clinical MRI scans without human intervention, thereby demonstrating the potential to assist the clinician in tumor delineation. Future research of validating imaging findings with histopathologic images to map spatial extend of tumor foci is warranted for accurate delineation of ILs for radiation treatment.

Supplementary data

Supplementary material for this article can be found at <https://doi.org/10.1016/j.adro.2020.01.005>.

References

1. Smith RA, Andrews KS, Brooks D, et al. Cancer screening in the United States, 2018: A review of current American Cancer Society guidelines and current issues in cancer screening. *CA Cancer J Clin.* 2018;68:297-316.

2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin.* 2019;69:7-34.
3. Dinh CV, Steenbergen P, Ghobadi G, et al. Magnetic resonance imaging for prostate cancer radiotherapy. *Phys Med.* 2016;32:446-451.
4. Moghanaki D, Turkbey B, Vapiwala N, et al. Advances in prostate cancer magnetic resonance imaging and positron emission tomography-computed tomography for staging and radiotherapy treatment planning. *Semin Radiat Oncol.* 2017;27:21-33.
5. Dearnaley DP, Jovic G, Syndikus I, et al. Escalated-dose versus control-dose conformal radiotherapy for prostate cancer: Long-term results from the MRC RT01 randomised controlled trial. *Lancet Oncol.* 2014;15:464-473.
6. Ghai S, Haider MA. Multiparametric-MRI in diagnosis of prostate cancer. *Indian J Urol.* 2015;31:194.
7. Lee H, Grosse R, Ranganath R, Ng A Y. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In: *Proceedings of the 26th Annual International Conference on Machine Learning, Montreal, Quebec, Canada, June 14-18.* 2009:609-616.
8. Dosovitskiy A, Tobias Springenberg J, Brox T. Learning to generate chairs with convolutional neural networks. In: *Proceedings of the Institute of Electrical and Electronics Engineers (IEEE) Conference on Computer Vision and Pattern Recognition, Boston, Massachusetts, June 7-12.* 2015:1538-1546.
9. 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.* Cham: Springer. 2015:234-241.
10. He K, Gkioxari G, Dollár P, Girshick R. Mask r-cnn. In: *Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, October 22-29.* 2017:2961-2969.
11. Litjens G, Debats O, Barentsz J, Karssemeijer N, Huisman H. Computer-aided detection of prostate cancer in MRI. *IEEE Trans Med Imaging.* 2014;33:1083-1092.
12. Litjens G, Debats O, Barentsz J, Karssemeijer N, Huisman H. ProstateX Challenge data. *The Cancer Imaging Archive.* 2017. <https://doi.org/10.7937/K9TFCIA.2017.MURS5CL>.
13. Clark K, Vendt B, Smith K, et al. The Cancer Imaging Archive (TCIA): Maintaining and operating a public information repository. *IEEE Trans Med Imaging.* 2013;26:1045-1057.
14. Yu L, Yang X, Chen H, Qin J, Heng P A. Volumetric ConvNets with mixed residual connections for automated prostate segmentation from 3D MR images. In: *31st Association for the Advancement of Artificial Intelligence (AAAI) Conference on Artificial Intelligence, San Francisco, California, February 4-9.* 2017.
15. Kamnitsas K, Ledig C, Newcombe VFJ, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med Image Anal.* 2017;36:61-78.
16. Dou Q, Chen H, Jin Y, Yu L, Qin J, Heng PA. 3D deeply supervised network for automatic liver segmentation from CT volumes. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention, Athens, Greece, October 17-21* Springer, Cham. 2016:149-157.
17. Tian Z, Liu L, Zhang Z, Fei B. Superpixel-based segmentation for 3D prostate MR images. *IEEE Trans Med Imaging.* 2016;35:791-801.
18. Mahapatra D, Buhmann JM. Prostate MRI segmentation using learned semantic knowledge and graph cuts. *IEEE Trans Biomed Eng.* 2014;61:756-764.
19. Guo Y, Gao Y, Shen D. Deformable MR prostate segmentation via deep feature learning and sparse patch matching. *IEEE Trans Biomed Eng.* 2016;35:1077-1089.
20. Milletari F, Navab N, Ahmadi SA. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: *2016 Fourth International Conference on 3D Vision (3DV), Stanford, California, October 25-28.* 2016:565-571. IEEE.
21. Zhu Q, Du B, Turkbey B, Choyke PL, Yan P. Deeply-supervised CNN for prostate segmentation. In: *2017 International Joint Conference on Neural Networks, Anchorage, Alaska, May 14-19.* 2017: 178-184. IEEE.
22. Toth R, Madabhushi M. Multifeature landmark-free active appearance models: Application to prostate MRI segmentation. *IEEE Trans Biomed Eng.* 2012;31:1638-1650.
23. Liao S, et al. Representation learning: A unified deep learning framework for automatic prostate MR segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention.* Springer; 2013.
24. Vincent G, Guillard G, Bowes M. Fully automatic segmentation of the prostate using active appearance models. In: *MICCAI Grand Challenge: Prostate MR Image Segmentation.* 2012:2.
25. Klein S, van der Heide UA, Lips IM, van Vulpen M, Staring M, Pluim JP. Automatic segmentation of the prostate in 3D MR images by atlas matching using localized mutual information. *Med Phys.* 2008;35:1407-1417.
26. Li A, Li C, Wang X, Eberl S, Feng DD, Fulham M. Automated segmentation of prostate MR images using prior knowledge enhanced random walker. In: *2013 International Conference on Digital Image Computing: Techniques and Applications, Hobart, Tasmania, Australia, November 26-28.* 2013:1-7. IEEE.
27. Kohl S, et al. Adversarial networks for the detection of aggressive prostate cancer. 2017.
28. Cameron A, Khalvati F, Haider MA, Wong A. MAPS: A quantitative radiomics approach for prostate cancer detection. *IEEE Trans Biomed Eng.* 2016;63:1145-1156.
29. Chung AG, Khalvati F, Shafiee MJ, et al. Prostate cancer detection via a quantitative radiomics-driven conditional random field framework. *IEEE Access.* 2015;3:2531-2541.
30. Artan Y, Haider MA, Langer DL, et al. Prostate cancer localization with multispectral MRI using cost-sensitive support vector machines and conditional random fields. *IEEE Trans Image Process.* 2010;19:2444-2455.
31. Artan Y, Haider MA, Yetik IS. Prostate cancer segmentation using multispectral random walks. In: *International Workshop on Prostate Cancer Imaging.* Springer; 2010.
32. Artan Y, Haider MA, Langer DL, Yetik IS. Semi-supervised prostate cancer segmentation with multispectral MRI. In: *2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Rotterdam, Netherlands, April 14-17.* 2010:648-651. IEEE.
33. Ozer S, Haider MA, Langer DL, et al. Prostate cancer localization with multispectral MRI based on relevance vector machines. In: *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Boston, Massachusetts, June 28-July 1.* 2009:73-76. IEEE.
34. Artan Y, Langer DL, Haider MA, et al. Prostate cancer segmentation with multispectral MRI using cost-sensitive conditional random fields. In: *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Boston, Massachusetts, June 28-July 1.* 2009:278-281. IEEE.
35. Liu X, Langer DL, Haider MA, Yang Y, Wernick MN, Yetik IS. Prostate cancer segmentation with simultaneous estimation of Markov random field parameters and class. *IEEE Trans Med Imaging.* 2009;28:906-915.
36. Park SY, Kim CK, Park BK, Kwon GY. Comparison of apparent diffusion coefficient calculation between two-point and multipoint B value analyses in prostate cancer and benign prostate tissue at 3 T: Preliminary experience. *AJR Am J Roentgenol.* 2014;203: W287-W294.
37. Doersch C. 2016. Tutorial on variational autoencoders. arXiv preprint arXiv:1606.05908.
38. Goodfellow IJ, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. In: *Advances in Neural Information Processing Systems.* 2014:2672-2680.