

Original Research Article

Deep learning-based auto-segmentation of targets and organs-at-risk for magnetic resonance imaging only planning of prostate radiotherapy

Sharif Elguindi^{a,*}, Michael J. Zelefsky^b, Jue Jiang^a, Harini Veeraraghavan^a, Joseph O. Deasy^a, Margie A. Hunt^a, Neelam Tyagi^a^a Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, NY 10065, United States^b Department of Radiation Oncology, Memorial Sloan Kettering Cancer Center, New York, NY 10065, United States

ARTICLE INFO

Keywords:

Deep learning
Autosegmentation
MR-only
U-Net
Prostate cancer

ABSTRACT

Background and purpose: Magnetic resonance (MR) only radiation therapy for prostate treatment provides superior contrast for defining targets and organs-at-risk (OARs). This study aims to develop a deep learning model to leverage this advantage to automate the contouring process.

Materials and methods: Six structures (bladder, rectum, urethra, penile bulb, rectal spacer, prostate and seminal vesicles) were contoured and reviewed by a radiation oncologist on axial T2-weighted MR image sets from 50 patients, which constituted expert delineations. The data was split into a 40/10 training and validation set to train a two-dimensional fully convolutional neural network, DeepLabV3+, using transfer learning. The T2-weighted image sets were pre-processed to 2D false color images to leverage pre-trained (from natural images) convolutional layers' weights. Independent testing was performed on an additional 50 patient's MR scans. Performance comparison was done against a U-Net deep learning method. Algorithms were evaluated using volumetric Dice similarity coefficient (VDSC) and surface Dice similarity coefficient (SDSC).

Results: When comparing VDSC, DeepLabV3+ significantly outperformed U-Net for all structures except urethra ($P < 0.001$). Average VDSC was 0.93 ± 0.04 (bladder), 0.83 ± 0.06 (prostate and seminal vesicles [CTV]), 0.74 ± 0.13 (penile bulb), 0.82 ± 0.05 (rectum), 0.69 ± 0.10 (urethra), and 0.81 ± 0.1 (rectal spacer). Average SDSC was 0.92 ± 0.1 (bladder), 0.85 ± 0.11 (prostate and seminal vesicles [CTV]), 0.80 ± 0.22 (penile bulb), 0.87 ± 0.07 (rectum), 0.85 ± 0.25 (urethra), and 0.83 ± 0.26 (rectal spacer).

Conclusion: A deep learning-based model produced contours that show promise to streamline an MR-only planning workflow in treating prostate cancer.

1. Introduction

Magnetic resonance (MR) only radiation therapy treatment planning for prostate cancer represents a significant paradigm shift from conventional computed tomography (CT) based planning. Superior soft tissue contrast provided by the MRI allows clinicians to define the clinical target volume [CTV] (prostate and seminal vesicles) more accurately and consistently as well as to distinguish organs-at-risk (OARs) near critical portions of the target boundary (bladder, rectal wall, and urethra). Currently, this process is done manually and is not only time-consuming for clinicians but prone to inter-observer variability [1]. It is therefore desirable to develop an automatic segmentation method that will aid in streamlining the current MR-only workflow as well as transition to fast online adaptive planning. Besides the time-saving advantage, an accurate auto-segmentation method could provide the

ability to expand a single clinician's expertise as a learning tool for others new to the treatment modality to improve consistency of contours.

Results from several recent studies have shown that deep learning-based approaches for auto-segmentation represent a significant improvement from older atlas-based methods due to their ability to learn complex sets of features to accurately perform pixel-wise classification of images [2–4]. Once trained, these deep learning models offer a clear computational advantage over atlas-based segmentation because the model performs inference directly on the new data without requiring any additional computations involving the training set. Because these methods are the most accurate machine learning techniques [5], a deep learning approach to automatically segment targets and OARs in prostate radiotherapy was performed. Recent works on deep learning have shown the feasibility to generate highly accurate segmentations

* Corresponding author.

E-mail address: elguinds@mskcc.org (S. Elguindi).<https://doi.org/10.1016/j.phro.2019.11.006>

Received 14 October 2019; Received in revised form 20 November 2019; Accepted 22 November 2019

2405-6316/ © 2019 The Authors. Published by Elsevier B.V. on behalf of European Society of Radiotherapy & Oncology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

for clinical use in radiation therapy, but require large, expert-segmented datasets (on the order of hundreds of patients) [6,7]. The large expert-labeled datasets are required to estimate the high number of parameters of convolutional layers in these deep learning models. MR as the primary imaging modality for treatment planning is relatively new for prostate radiation therapy and consequently, expert target and OAR delineations is difficult to obtain. The limitation of small datasets with expert segmentations, that is frequently encountered in medical imaging, poses a problem for training a deep convolutional network, which typically has a very large number of parameters [8]. A commonly used solution to this problem is transfer learning [9], where the networks trained on a different task or domain, is reused for a new application either by retraining all the filters in the network's layers from scratch [10] or by retraining the old filters and only tuning a defined subset [11]. It has also been shown that using networks trained on other tasks/domains combined with sufficient fine tuning can lead to more accurate performance than when training from scratch [12,13]. Therefore, this work analyzes the use of applying a transfer learning technique for segmenting MR images for prostate radiotherapy and compares it against training from scratch with a U-Net type architecture as well as study its feasibility in generating clinically useable contours.

2. Materials and methods

2.1. Datasets

MR-only simulation and planning for prostate radiotherapy has been routinely used at our institution since June 2016 [14,15]. Of all patients who were treated using a MR-only workflow, one hundred patients with confirmed prostate cancers scanned with axial T2w MRI were retrospectively analyzed and anonymized under an internal review board approved study. The first fifty scans were contoured and curated by two experts so that they can be used to develop the deep learning model using forty for training and ten for validation. They were first reviewed and critiqued by a dosimetrist and/or physicist for consistency of OAR segmentation with institutional standards. These were subsequently re-reviewed and edited as necessary for accuracy by a single board-certified radiation oncologist experienced in MR-only prostate contouring. This helped to ensure consistency in defining critical regions of anatomy unique to intact prostate radiotherapy, particularly the rectosigmoid junction, bladder/prostate interface and the prostate apex. The large bowel was added after review as a constraining structure in the model for the bladder, due to the similar contrast between the bladder wall and bowel loops typically found in regions superior to the prostate. Although not all patients received a rectal spacer gel, they were included in the model for training and evaluation to represent a broader clinical scenario. The second fifty scans were randomly selected from clinically treated patients at our institution, to be used to assess the clinical viability of the model. This dataset was not re-reviewed or edited in any way and represents a range of contouring styles of CTV volumes from different physicians and OARs from physicists/dosimetrists. The clinical parameters of the patient population from both datasets are included in Table A.1 in Appendix A.

2.2. Deep learning architecture – deep LabV3+

Prior works have demonstrated successful results in segmenting medical image sets from small training datasets by repurposing networks constructed and trained on natural images using transfer learning approaches and have shown that these methods were superior to training the networks from only the medical image datasets [16,17]. A fully convolutional network (FCN) method developed by Chen et al. [18] known as DeepLabV3+ was chosen because of its demonstrated high performance segmenting natural images (PASCAL VOC2012 segmentation challenge) [19]. The network architecture was built with an encoding path used to extract features from the input image through

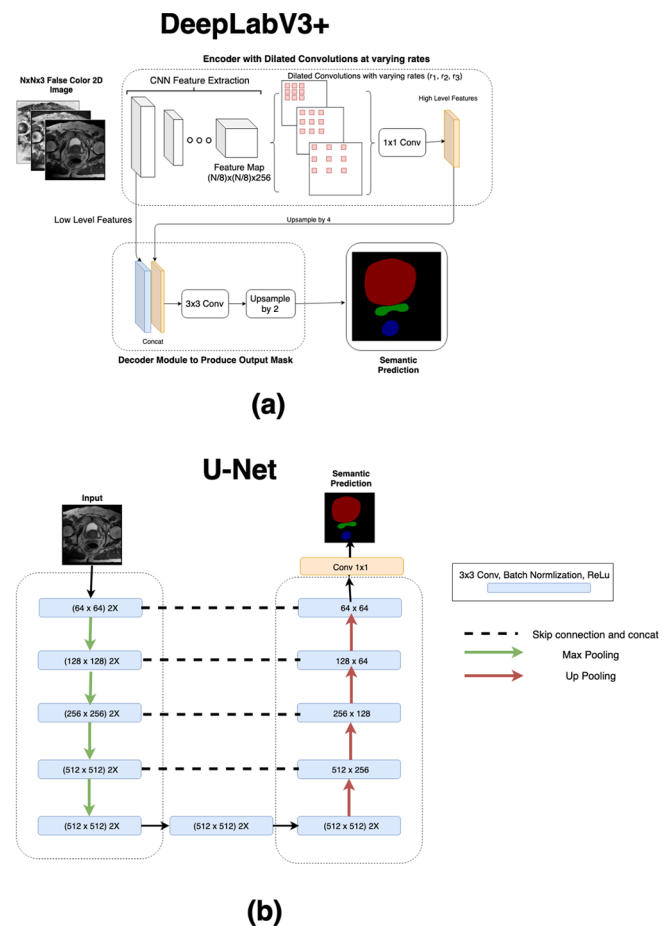


Fig. 1. (a) Schematic of DeepLabV3+ architecture used to produce segmentations. Dilation rates of 12, 24, 36 were used to keep the feature map reduction to one-eighth the input image size (termed output stride = 8) in the encoder module. The entire network except the last 3×3 convolution was initialized with a pre-trained model based on Microsoft Common Objects in Context and PASCAL Visual Object Classes image sets of natural objects. (b) Basic U-Net architecture that utilizes skip connections at each convolution block to keep long range contextual information throughout the encoder/decoder structure.

convolutions, followed by a decoder to recover a probability mask for each pixel. The encoder first extracts feature maps using a modified Xception model proposed by Chollet et al. [20] and then refines it for segmentation using dilated convolutions, which can be applied at varying rates to encode features at different spatial resolutions without reducing the dimensions of the feature space. Such a formulation of encoder enables the network to capture abundant contextual image information while reducing the input image by only a factor of 8 or 16, as opposed to standard convolutional neural networks that routinely reduce feature space up to 32 or 64 times. This in turn allows for a simpler decoder structure to produce the output mask. Details regarding the model configuration used in this work are presented in Fig. 1a. Additional details about the model design and performance in non-medical tasks are presented in a series of publications by Chen, et al. [21–23].

Pre-processing was done on the T2w scans to multiple axial false color images, consisting of separate red, green, and blue channels. The MR scans were first down-sampled from 16-bit to 8-bit grayscale images, resulting in a tolerable mean average reduction in signal-to-noise ratio of 0.96 ± 0.2 . The red channel is the unprocessed 8-bit copy of the original axial image slice. The green channel applies an imaging processing technique called Contrast Limited Adaptive Histogram Equalization on the axial image using a tile size of 8×8 pixels and a clip limit of 10% [24]. The blue channel is the red channel

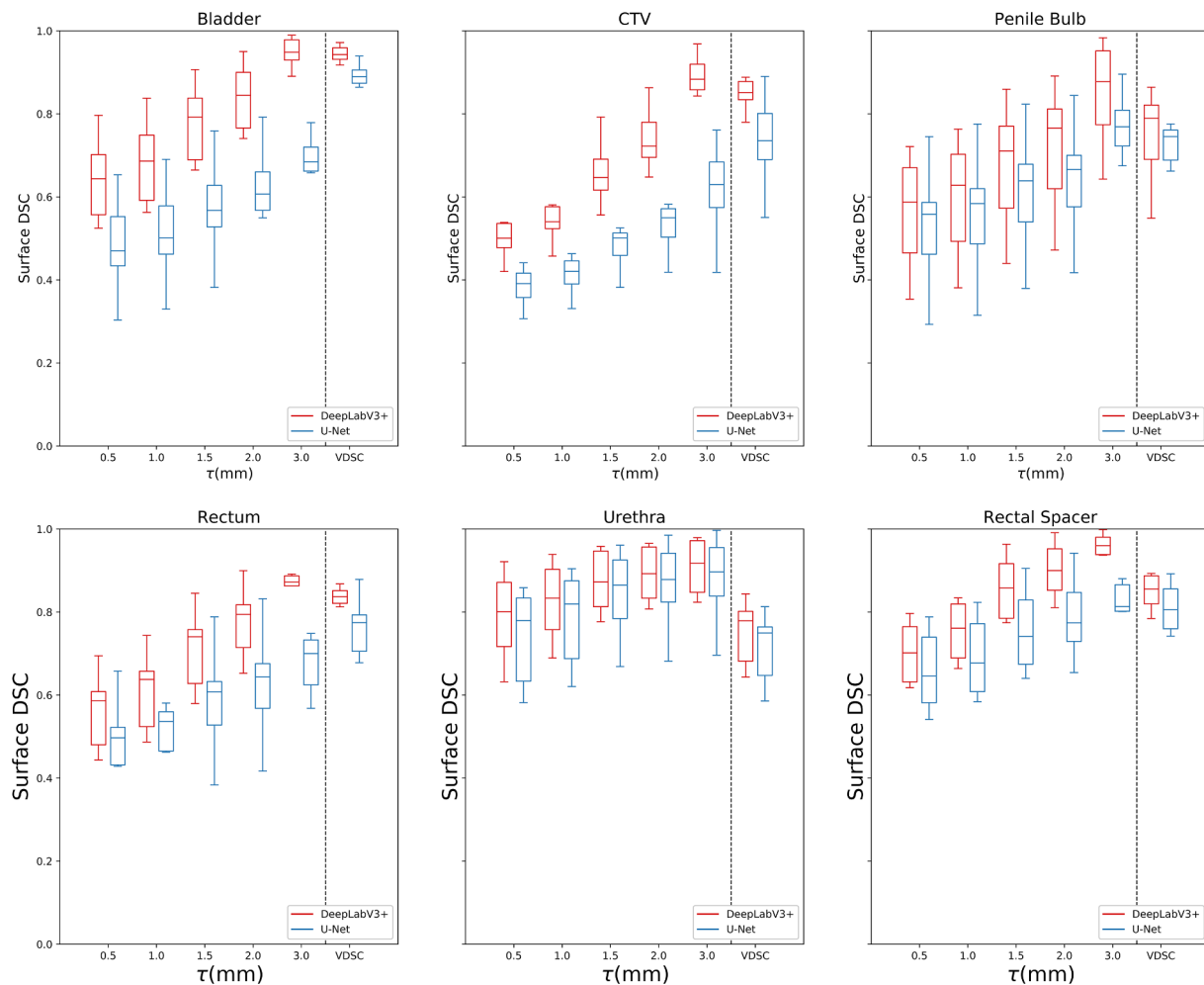


Fig. 2. Boxplots (Red – DeepLabV3+ ; Blue – U Net) for each structure of surface Dice similarity coefficient (SDSC) as a function of τ for the 10 patients reviewed by a single expert radiation oncologist. The data is plotted for the range of τ from sub pixel (0.5 mm) to the approximate voxel size (3 mm). An additional box plot of the volumetric Dice similarity coefficient (VDSC) is included for comparison. Box plots are shown with an inter-quartile range from 25 to 75 percent with the horizontal line representing median value. Upper and lower whisker represent the 5 and 95 percentiles. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

inverted, by subtracting every pixel by 255 and then renormalizing to 0–255.

Each associated structure set was converted into a corresponding masked array. The six structures are defined fully by anatomical features and are not considered regions with ill-defined boundaries except the rectosigmoid junction and the inferior portion of the penile bulb. The urethra was defined to start at the bladder and stop at the most superior aspect of the penile bulb.

A pretrained model that is publicly available was used as an initialization point for both the encoder and decoder portions of the DeepLabV3+ architecture, except the last 3×3 layer, which was adjusted to the output number of structures, including the constraining structure, large bowel (7). The model weights for all other layers were initialized using the same network pretrained on the Microsoft Common Objects in Context (COCO) and PASCAL Visual Object Classes (VOC) challenge datasets. This dataset consists of millions of annotated and segmented photos of objects in their natural context [19,25]. Retraining was performed on all the layers of the network on a high-performance GPU cluster with four NVIDIA GTX 1080i GPUs with 11 GB of memory each. To use a sufficiently large batch size (16 images per batch) so that training could be performed with batch normalization, images were cropped to 256×256 patches. Data augmentation was applied to each 2D image through scaling, cropping, and rotations using 5 random permutations per axial image. The full network was

then re-trained with learning rate set initially at 0.007. A polynomial learning rate decay was used with decay factor of 0.1 and power 0.9. Training was carried out for 20 epochs. An epoch represents one iteration where the entire training dataset is passed through the neural network. A class-weighted cross-entropy loss was implemented to reduce the influence of data imbalance. Early stopping was implemented according to the accuracy of volumetric Dice similarity coefficient (VDSC) on a selected five validation patients. Training time was approximately 10 h.

2.3. Deep learning architecture – U-Net

Comparison of the transfer learning approach outlined in Section 2.2 was done using the default U-Net architecture [26], that is composed of 4 max-pooling for subsampling and 4 up-pooling for resampling to obtain the segmentation at the image resolution. Skip connections were used to connect the features from the encoding path with the decoding path to prevent loss of resolution. The network consisted of convolutional blocks, with each block consisting of convolution, batch normalization, and ReLU activation. The network used 17.2 M parameters with 64 channels used in the last layer, which was converted into a single channel through a 1×1 convolution. Details regarding this network are presented in Fig. 1b. This network was trained from scratch using the 40 training image sets. Additional data

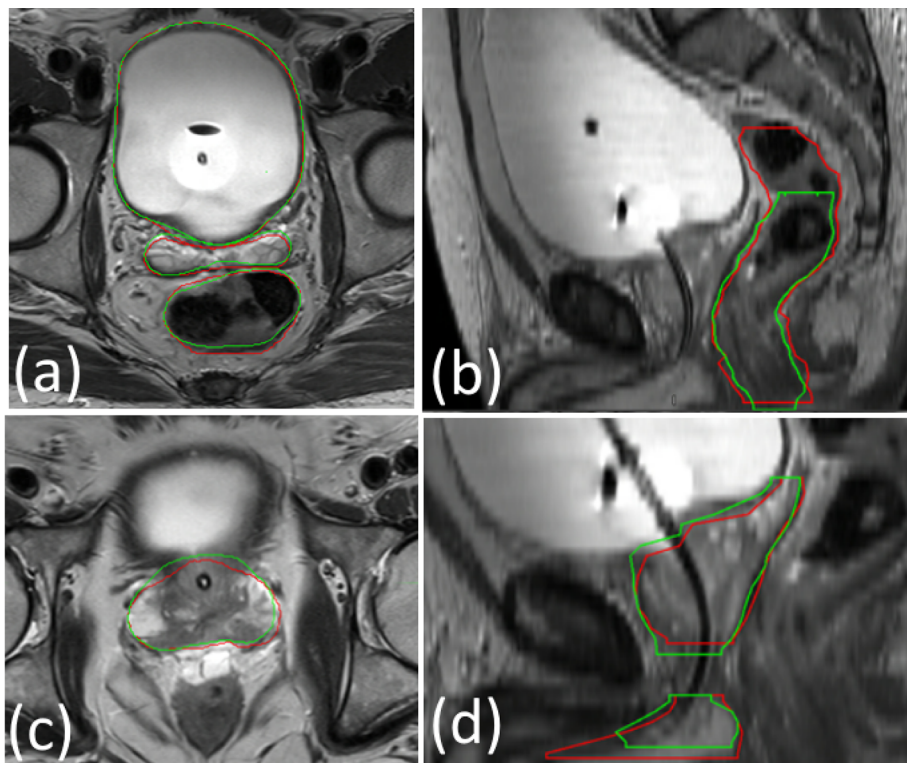


Fig. 3. Example model output, green (DeepLabV3+ model), red (expert) for 4 different patient anatomies. (a) Axial slice of bladder, rectum, and seminal vesicles. (b) Sagittal example of model determining rectum/sigmoid junction shorter than expert contour. (c) Small discrepancy near bladder/prostate interface (d) Example of expert penile bulb extending anteriorly and posteriorly from model output. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

augmentation consisting of rotation, mirroring, translation, and scaling was performed to increase the training set size using per-epoch augmentation such that new datasets were created using each epoch. The algorithm was trained using 2D patches of size 256×256 obtained through cropping the images. The algorithm was optimized using ADAM optimizer and the training was done for 200 epochs with an initial learning rate of 0.0001. The model with the highest VDSC on the validation set was chosen. The VDSC scores from validation and testing sets between the U-Net and DeepLabV3+ results were chosen as the statistical comparison end points using a paired, two-sided Wilcoxon test.

2.4. Surface Dice similarity coefficient (SDSC)

SDSC is a metric proposed by Nikolov, et al. to capture the inherent inter-observer variability found in radiation therapy manual segmentation [27]. Specifically, SDSC compares the Dice similarity coefficient between two surfaces adding an additional parameter, τ , with units of distance, which reflects this inter-observer uncertainty. Any difference of the surface boundary below τ would be considered clinically acceptable and therefore would correlate with a segmentation needing no adjustment from experts. Selection of this parameter is subjective to the amount and type of user base it is drawn from and can drastically change the Dice similarity coefficient value reported. Rather than attempting to select a value for comparison, SDSC is analyzed as a function of τ , noting that faster convergence to 1.0 correlates with segmentations likely needing little to no adjustment.

2.5. External dataset – PROMISE12

Analysis was also performed on the PROMISE12 challenge dataset [28] to establish the generalizability of our model on an external dataset. This public dataset consists of fifty axial T2-weighted MR images on which the prostate gland alone was contoured. We stratify the scans into two parts, ones with MR endorectal coil and ones with out. Comparison was performed with and without re-training DeepLabV3+. For

retraining, the dataset was randomized into forty training cases and ten validation cases

3. Results

3.1. Comparison against U-Net

Fig. 2 shows SDSC scores for DeepLabV3+ and U-net as a function of τ for the six structures of interest in this study for the validation set, compared directly next to VDSC. Sample output of the DeepLabV3+ model to the expert contours are shown in Fig. 3. The same analysis is shown on the test set in Fig. 4. DeepLabV3+ outperformed U-Net for bladder, rectum, CTV, penile bulb and rectal spacer ($P < 0.001$), urethra was statistically similar ($P = 0.307$).

3.2. Performance on PROMISE12

Average VDSC using DeepLabV3+ for the prostate gland on the PROMISE12 dataset without training was 0.59 with the endorectal coil (24 scans) and 0.78 without (26 scans). Average VDSC after retraining for the validation set is 0.95 ± 0.01 . Fig. 5 shows an example output of the model for each scenario.

4. Discussion

MR-Only radiation treatment planning for prostate is becoming increasingly common among radiation oncology centers. This is also accelerated by the adoption of new advancements in linear accelerator technology, whereby daily MR imaging in the treatment position is possible. In the work presented here, we aim to demonstrate the ability of a deep learning approach to utilize the improved soft tissue contrast provided by MR imaging to create a deep-learning based model that can help streamline the MR-Only workflow.

Many works have cited using U-Net [26], or modified U-Net type architectures, in lieu of training larger state-of-the-art FCNs. This is because U-Net architectures (a) are more shallow and have fewer

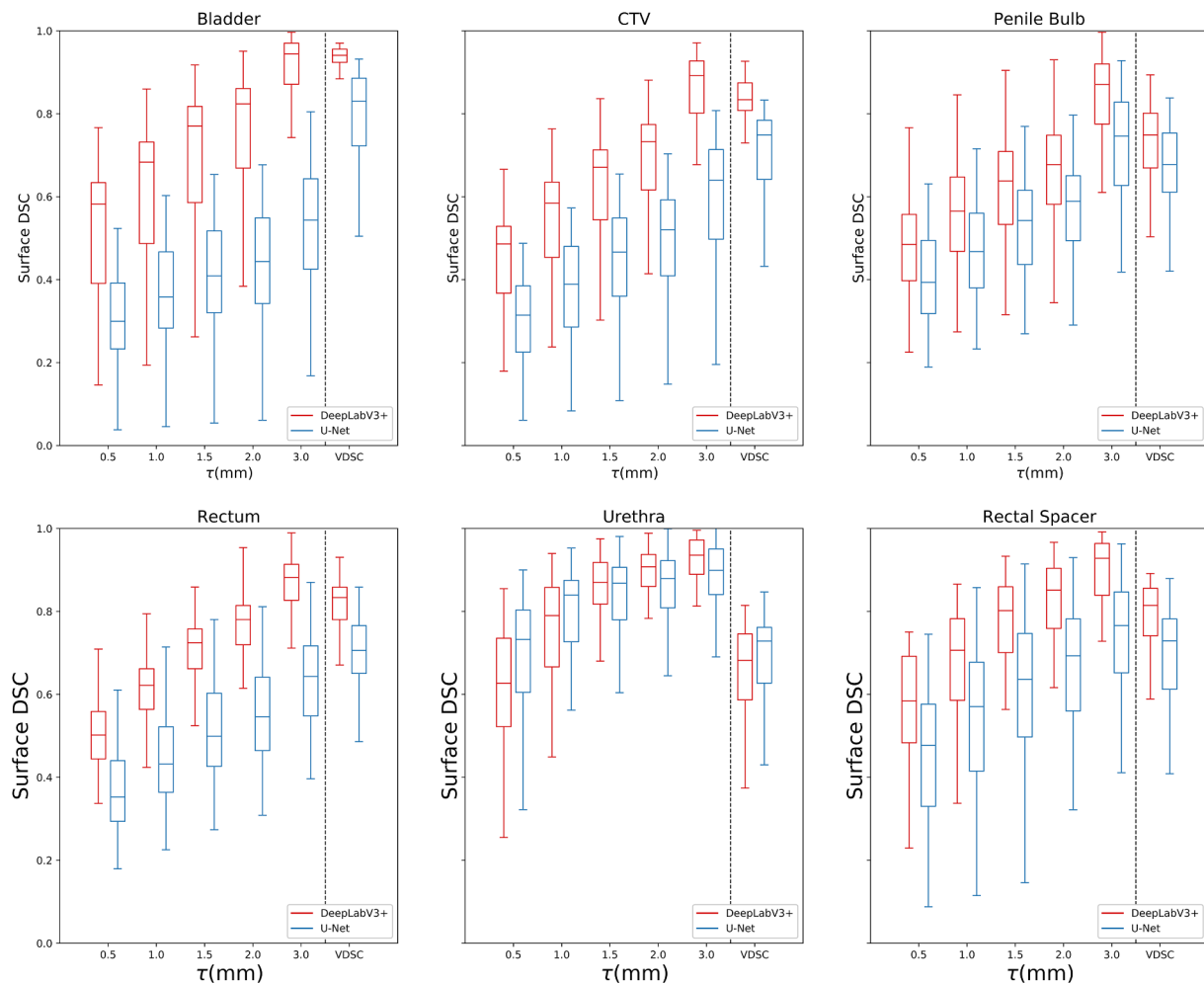


Fig. 4. Boxplots (Red – DeepLabV3+; Blue – U Net) for each structure of surface Dice similarity coefficient (SDSC) as a function of τ with the clinically used test set. The data is plotted for the range of τ from sub pixel (0.5 mm) to the approximate voxel size (3 mm). An additional box plot of the volumetric Dice similarity coefficient (VDSC) is included for comparison. Box plots are shown with an inter-quartile range from 25 to 75 percent with the horizontal line representing median value. Upper and lower whisker represent the 5 and 95 percentiles. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

parameters thus making them easier to train with only a small representative dataset and (b) have been shown to have the capacity to learn the underlying relationship between pixel values and segmentations, thereby providing potentially clinically usable results [29,30]. The benefit of larger networks, like DeepLabV3+, is the ability for further abstraction from the original image which may allow for more complex relationships to be understood in the underlying data, the trade-off being larger datasets required to train these deeper convolutions effectively. In addition, hyper parameter choices, such as training policies, image process/data augmentation techniques, the choice of loss function, etc., can have large effects on the results of these networks because encoding and decoding layers are difficult to develop regardless of network capacity. By adjusting the input dataset (MR images) with simple pre-processing rules to align it more closely with the pre-trained images (ImageNet, MS COCO, PASCAL VOC), this work demonstrates that a large, high-capacity network can effectively relearn with a small (as compared to natural image segmentation problems) training set. Of note, the pre-processing does influence the model performance. Preliminary evaluation using the single raw image copied across the three channels produced sub-optimal results as compared with the method demonstrated here. An alternative to this pre-processing is to convert the pre-training natural images to grayscale, which would eliminate the need for the pre-processing step.

In this instance, a basic U-Net was shown to underperform on the

training/validation dataset when compared with DeepLabV3+. In many patient image sets, the U-Net model suffered from false positive results outside the intended organ of interest which attributed to its underperformance. We suspect that the benefit of pre-training allowed the DeepLabV3+ encoder network to be more discriminatory against similar contrast tissue in the surrounding anatomy. This may also explain why the Foley-defined urethra (foreign object) was equally contoured between both networks.

Fig. 6 shows a few examples of the discrepancies between the better performing DeepLabV3+ model and clinical contours. In general, we find that the contour quality is more consistent in the model data than what is used clinically as shown in images (a), (b) and (c) in Fig. 6. We suspect that the larger inter-quartile ranges shown in SDSC in the test set can be mostly attributed to such variability. The SDSC analysis also demonstrates that the bulk of the contour surface (80–90%) is within one voxel distance ($\tau = 3$ mm) away from the clinically used contours showing that most of the differences are subtle, allowing physicians the ability to adjust contours upon review, particularly the CTV, to match clinical concerns such as the risk for extra-prostatic extension based on biopsy and/or radiographic findings. The ability of the physicians to act as a co-pilot to the auto segmentations, reviewing/revising as needed, will help to streamline our MR-only planning workflow. After an internal QA process, we have implemented this auto-contour method into our MR-only planning workflow. We intend to report on the results of

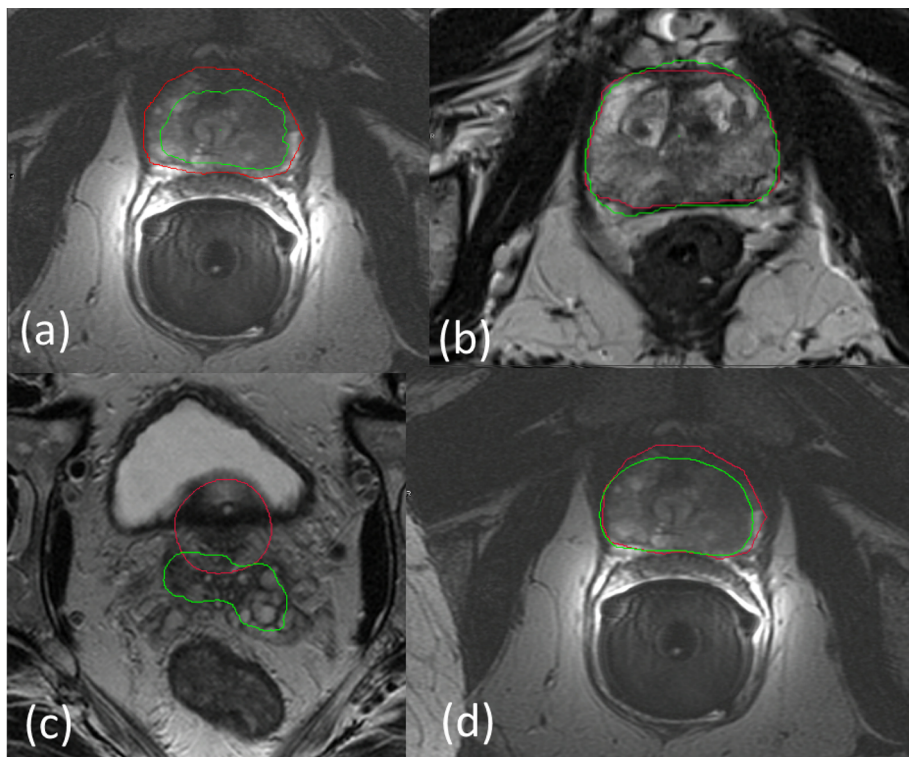


Fig. 5. Example model output, green (DeepLabV3+ model), red (clinical contour) on the PROMISE12 training set. (a) Model fails in the presence of a MR rectal coil which is not used at our institution. (b) Model performance is improved on cases without a coil. Our model was trained to detect seminal vesicles as shown in (c), which reduces VDSC. (d) When the model is trained using 40 patients from the provided dataset, DeepLabV3+ performance is significantly improved with or without the coil on the other 10 validation patient sets. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

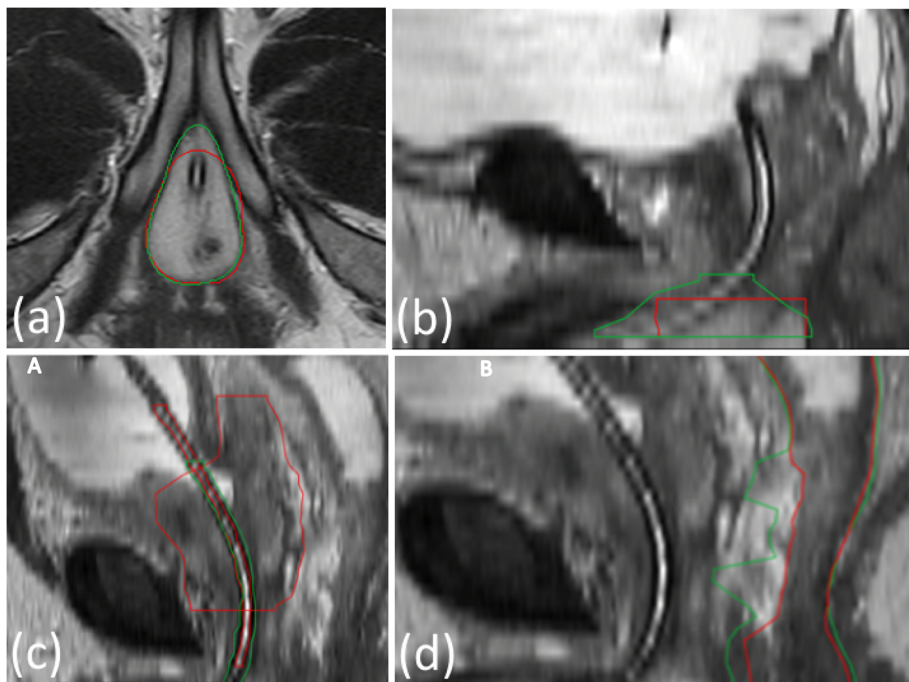


Fig. 6. Example model output, green (DeepLabV3+ model), red (clinical contour) for 2 different patient anatomies. (a)(b) Model shows more consistent detection/definition than clinical contoured penile bulb (c) Urethra used clinically does not about the end of the urethral lumen, model was trained to include the whole space. Clinically used urethra extends further into bladder, which is not needed in context of radiotherapy planning. (d) Same patient as (c), rectal spacer/rectum boundary is unique, model incorrectly classifies many slices of spacer. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

this pilot period when enough data and feedback have been collected.

Finally, we analyzed our model with an external dataset (PROMISE12). The trained model using our internal data performed average on MRs without the endorectal coil but failed in the instance of the coil in place. This is primarily due to the training set not including such cases. Also, VDSC suffers somewhat in the no-coil setting as our clinical model includes the seminal vesicles. To demonstrate that this method can work well in other contexts, we retrained DeepLabV3+ exclusively on a subset of the PROMISE12 training data and achieve VDSC on the order of 0.95 on the validation set of ten patients (mixed with and without endorectal coil). This demonstrates that, given

enough training examples, the model can be generalizable to T2-weighted scans from different scanners as well as acquisition using different radiofrequency coils.

In conclusion, we investigated the use of a deep learning network, DeepLabV3+, combined with a transfer learning technique to accurately learn and segment relevant targets and OARs in the context of MR-only prostate radiation therapy. This work shows that using pre-trained weights from a model trained on a very large unrelated dataset can adapt learned features to the problem at hand.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Neelam Tyagi reports having a research agreement from Philips Healthcare, outside the submitted work. Michael Zelefsky reports being a consultant for Augmenix, outside the submitted work. Margie Hunt

Appendix A

Table A.1

Table A.1
Summary of patient clinical parameters.

# of patients	100
Age	68.9 (53–87)
Clinical T Stage	60
	T1c
	T2a
	T2b
	T2c
	T3a,b
Initial PSA	18
Gleason Score	9.29 (0.58–64.33)
	6
	7 (3 + 4)
	7 (4 + 3)
	8 (4 + 4)
	9
Prostate volume (cc)	14
	41
	21
	12
	12
	44.4 (12.8–96)

References

- [1] Fiorino C, Reni M, Bolognesi A, Cattaneo GM, Calandrino R. Intra- and inter-observer variability in contouring prostate and seminal vesicles: implications for conformal treatment planning. *Radiother Oncol* 1998. [https://doi.org/10.1016/S0167-8140\(98\)00021-8](https://doi.org/10.1016/S0167-8140(98)00021-8).
- [2] Guo Y, Gao Y, Shen D. Deformable MR prostate segmentation via deep feature learning and sparse patch matching. first ed. Elsevier Inc.; 2017. <https://doi.org/10.1016/B978-0-12-810408-8.00012-2>.
- [3] Granger E, Desrosiers C. Multi-region segmentation of bladder cancer structures in MRI with progressive dilated convolutional networks. (arXiv:1805.10720v2 [cs.CV] UPDATED), (n.d.). <https://doi.org/10.1002/mp.13240>.
- [4] Liang S, Tang F, Huang X, Yang K, Zhong T, Hu R, et al. Deep-learning-based detection and segmentation of organs at risk in nasopharyngeal carcinoma computed tomographic images for radiotherapy planning. *Eur Radiol* 2018. <https://doi.org/10.1007/s00330-018-5748-9>.
- [5] Lecun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521:436–44. <https://doi.org/10.1038/nature14539>.
- [6] Men K, Chen X, Zhang Y, Zhang T, Dai J, Yi J, et al. Deep Deconvolutional neural network for Target segmentation of nasopharyngeal cancer in Planning computed Tomography images. *Front Oncol* 2017;7:1–9. <https://doi.org/10.3389/fonc.2017.00315>.
- [7] Fu Y, Mazur TR, Wu X. A novel MRI segmentation method using CNN-based correction network for MRI-guided adaptive radiotherapy. (n.d.). <https://doi.org/10.1002/mp.13221>.
- [8] Erhan D, Manzagol P-A, Bengio Y, Bengio S, Vincent P. The difficulty of training deep architectures and the effect of unsupervised. *Int Conf Artif Intell Stat* 2009;5:153–60. http://machinelearning.wustl.edu/mlpapers/paper_files/AISTATS09_ErhanMBBV.pdf.
- [9] Pan SJ, Yang Q. A survey on transfer learning. *IEEE Trans Knowl Data Eng* 2010;22:1345–59. <https://doi.org/10.1109/TKDE.2009.191>.
- [10] Gao M, Bagci U, Lu L, Wu A, Buty M, Shin HC, et al. Holistic classification of CT attenuation patterns for interstitial lung diseases via deep convolutional neural networks. *Comput Methods Biomech Biomed Eng Imaging Vis* 2018. <https://doi.org/10.1080/21681163.2015.1124249>.
- [11] Chen H, Ni D, Qin J, Li S, Yang X, Wang T, et al. Standard plane localization in fetal ultrasound via domain transferred deep neural networks. *IEEE J Biomed Heal Informatics* 2015. <https://doi.org/10.1109/JBHI.2015.2425041>.
- [12] Tajbakhsh N, Shin JY, Gurudu SR, Hurst RT, Kendall CB, Gotway MB, et al. Convolutional neural networks for medical image analysis: full training or fine tuning? *IEEE Trans Med Imag* 2016. <https://doi.org/10.1109/TMI.2016.2535302>.
- [13] Ribeiro E, Uhl A, Wimmer G, Häfner M. Exploring deep learning and transfer learning for colonic polyp classification. *Comput Math Methods Med* 2016. <https://doi.org/10.1155/2016/6584725>.
- [14] Tyagi N, Fontenla S, Zelefsky M, Chong-Ton M, Ostergren K, Shah N, et al. Clinical workflow for MR-only simulation and planning in prostate. *Radiat Oncol* 2017. <https://doi.org/10.1186/s13014-017-0854-4>.
- [15] Tyagi N, Fontenla S, Zhang J, Cloutier M, Kadbi M, Mechalakos J, et al. Dosimetric and workflow evaluation of first commercial synthetic CT software for clinical use in pelvis. *Phys Med Biol* 2017. <https://doi.org/10.1088/1361-6560/aa5452>.
- [16] Ghafoorian M, Mehrtash A, Kapur T, Karsssemeijer N, Marchiori E, Pesteie M, et al., Transfer learning for domain adaptation in MRI: Application in brain lesion segmentation, in: *Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, 2017. https://doi.org/10.1007/978-3-319-66179-7_59.
- [17] Shin H-C, Roth HR, Gao M, Lu L, Xu Z, Nogues I, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans Med Imaging* 2016. <https://doi.org/10.1109/TMI.2016.2528162>.
- [18] Chen L-C, Zhu Y, Papandreou G, Schroff F, Adam H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation; 2018. <https://doi.org/10.1159/000018039>.
- [19] Everingham M, Van Gool L, Williams CKI, Winn J, Zisserman A. The pascal visual object classes (VOC) challenge. *Int J Comput Vis* 2010;88:303–38. <https://doi.org/10.1007/s11263-009-0275-4>.
- [20] Chollet F. Xception: deep learning with depthwise separable convolutions. 2017 IEEE Conf. Comput. Vis. Pattern Recognit. IEEE; 2017. p. 1800–7. <https://doi.org/10.1109/CVPR.2017.195>.
- [21] Chen L-C, Collins MD, Zhu Y, Papandreou G, Zoph B, Schroff F, Adam H, Shlens J. Searching for efficient multi-scale architectures for dense image prediction; 2018. <https://doi.org/10.1063/1.3631751>.
- [22] Chen L-C, Papandreou G, Kokkinos I, Murphy K, Yuille AL. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Trans Pattern Anal Mach Intell* 2016. <https://doi.org/10.1109/TPAMI.2017.2699184>.
- [23] Chen L-C, Papandreou G, Schroff F, Adam H. Rethinking Atrous Convolution for Semantic Image Segmentation; 2017. <https://doi.org/10.1002/cncr.24278>.
- [24] Zhu Y, Huang C. An adaptive histogram equalization algorithm on the image gray level mapping. *Phys Procedia* 2012;25:601–8. <https://doi.org/10.1016/j.phpro.2012.03.132>.
- [25] Andriluka M, Pishchulin L, Gehler P, Schiele B. 2D human pose estimation: New benchmark and state of the art analysis, *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit*; 2014, 3686–3693. <https://doi.org/10.1109/CVPR.2014.471>.
- [26] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: *Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*; 2015. https://doi.org/10.1007/978-3-319-24574-4_28.
- [27] 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. p. 1–31. <http://arxiv.org/abs/1809.04430>.
- [28] Litjens G, Toth R, Van De Ven W, Hoeks C, Kerkstra S, Malmberg F, et al. Evaluation of prostate segmentation algorithms for MRI: the PROMISE12. *Med Image Anal* 2015;18:359–73. <https://doi.org/10.1016/j.media.2013.12.002.Evaluation>.
- [29] Alom MZ, Hasan M, Yakopcic C, Taha TM, Asari VK. Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation. *Desalination* 2018. <https://doi.org/10.1016/j.desal.2007.07.021>.
- [30] Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-net: Learning dense volumetric segmentation from sparse annotation. In: *Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, 2016. https://doi.org/10.1007/978-3-319-46723-8_49.