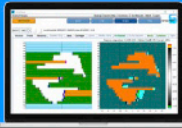


STREAMLINE. OPTIMIZE. TRUST.



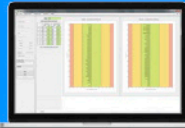
ADAPTIVO



LINACVIEW



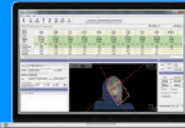
DOSEVIEW



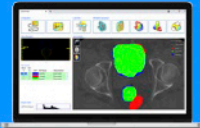
PIPSPRO



QA PILOT



IMSURE



STRUCTSURE
AI QA

COMPLETE INTEGRATED QA

STANDARD **IMAGING**®



[CLICK HERE TO LEARN MORE](#)

Technical note: The effect of image annotation with minimal manual interaction for semiautomatic prostate segmentation in CT images using fully convolutional neural networks

Maysam Shahedi¹ | James D. Dormer¹ | Martin Halicek¹ | Baowei Fei^{1,2,3}

¹ Department of Bioengineering, The University of Texas at Dallas, Richardson, Texas, USA

² Advanced Imaging Research Center, UT Southwestern Medical Center, Dallas, Texas, USA

³ Department of Radiology, University of Texas Southwestern Medical Center, Dallas, Texas, USA

Correspondence

Baowei Fei, Department of Bioengineering, University of Texas at Dallas, 800 W Campbell Road, Richardson, TX 75080, USA.
Email: bfei@utdallas.edu

Funding information

U.S. National Institutes of Health (NIH), Grant/Award Numbers: R01CA156775, R01CA204254, R01HL140325, R21CA231911; the Cancer Prevention and Research Institute of Texas, Grant/Award Number: RP190588

Abstract

Purpose: The goal is to study the performance improvement of a deep learning algorithm in three-dimensional (3D) image segmentation through incorporating minimal user interaction into a fully convolutional neural network (CNN).

Methods: A U-Net CNN was trained and tested for 3D prostate segmentation in computed tomography (CT) images. To improve the segmentation accuracy, the CNN's input images were annotated with a set of border landmarks to supervise the network for segmenting the prostate. The network was trained and tested again with annotated images after 5, 10, 15, 20, or 30 landmark points were used.

Results: Compared to fully automatic segmentation, the Dice similarity coefficient increased up to 9% when 5–30 sparse landmark points were involved, with the segmentation accuracy improving as more border landmarks were used.

Conclusions: When a limited number of sparse border landmarks are used on the input image, the CNN performance approaches the interexpert observer difference observed in manual segmentation.

KEYWORDS

computed tomography, deep learning, prostate, segmentation, user interactions

1 | INTRODUCTION

Computed tomography (CT) imaging is the standard imaging modality used for planning radiotherapy treatments. Segmentation of the target organ and the organs-at-risk is one of the main postimaging steps in radiotherapy treatment planning. However, for soft tissue organs, like the prostate, segmentation of the organ in CT images is challenging, time-consuming, and subject to high interobserver variability.¹

Using computer-assisted segmentation approaches could be an effective solution to decrease the segmentation time and improve the repeatability of the task. Recently, deep learning approaches are widely used for fully automatic medical image segmentation.² They showed fast and accurate segmentation performance in many cases.^{3–7} However, convolutional neural network (CNN) often requires a large amount of training data in

order to achieve good accuracy. One solution to address the accuracy issue is to use complicated algorithms,⁸ which requires intense calibration and fine-tuning; thus, the performance can be sensitive to the data and the selected hyperparameters. Another solution could be incorporating an expert operator interaction for guiding the network. However, deep learning is mostly used for fully automatic segmentation, and applying user inputs is not common in deep learning algorithms. Girum et al. have presented an interactive deep learning-based image segmentation approach and tested it on prostate CT images.⁹ Their approach is a slice-by-slice segmentation technique that needs manually selected points on all the 2D slices in a systematic way. They showed that their algorithm could provide high accuracy and robustness to interclinical image variation.

In this paper, we supervised a fully CNN (U-Net⁷) with minimal manual initialization to study the performance

improvement of the network in image segmentation. We applied the user inputs by annotating the images. We tested our method on three-dimensional (3D) segmentation of the prostate in CT and compared the segmentation results to fully automatic segmentation performed by U-Net on the same dataset. We used a radiologist manual segmentation as the reference standard to evaluate our approach. The U-Net architecture was implemented in 3D to have a fully 3D segmentation, and annotated input images were used as the operator's input. Unlike reference 9, in this study, it is not required to select points on all the slices and the points were not systematically selected. The number of selected points could be flexible (here we tested the algorithm based on 5–30 points per 3D image volume). The main contributions in this study include (1) the idea of using a simple and effective method to incorporate user interaction into CNN training and testing and (2) a less restricted boundary landmark selection (nonsystematic) for initializing the segmentation compared to the state-of-the-art approaches.

2 | MATERIALS AND METHODS

2.1 | Data

Our dataset has 92 3D abdominal CT scans from 92 prostate cancer patients. On each image, an expert radiologist manually segmented the prostate. The dataset contained CT images with original size of $512 \times 512 \times 27$ voxels and $0.977 \times 0.977 \times 4.25$ mm³ voxel size. The volume range of the prostate in the dataset based on the manual segmentation labels is from 12.1 to 79.3 cc with average volume of 26.8 cc (standard deviation = 9.6 cc). We randomly selected and used 65% of the patients (60 patients) for training, 10% of the patients (9 patients) for validation, and the remaining 25% (23 patients) were reserved for the final test.

2.2 | Preprocessing

To reduce the intersubject intensity and pattern variability, we extracted one histogram for all the prostate voxels in the training set and removed the 0.5% outliers, 0.25% from each histogram end. The observed Hounsfield unit (HU) range for prostate tissue voxels in the training set was $[-72, 112]$. We truncated the HU range on the CT images within the range of -127 to 128 based on the observed range to have an 8-bit dynamic range (i.e., 256 intensity levels). Then we scaled the dynamic range of the image intensities to the range of zero to one. Since the prostate volume occupied less than 5% of the whole cropped image volume on average, we decided to crop the image volumes to the size of $96 \times 96 \times 15$ voxels around the prostate region to minimize the memory load

for the graphics processing unit (GPU). We automatically cropped the images using the manual segmentation labels as reference to have the whole prostate within the volume.

2.3 | Operator manual interaction

2.3.1 | Experiment I

We included a set of sparse boundary landmarks randomly selected on the prostate surface simulating an operator's input used for guiding the prostate segmentation. We used the manual segmentation label provided by the radiologist for each CT image to select the landmarks. During the random selection of the landmark points, we applied a restriction of a minimum accepted distance between each point pair. To define the minimum accepted distance between the points, we approximated the prostate shape as a sphere with the same volume. Then we evenly distributed the same number of points on the surface of the sphere and found out the distance between the adjacent points. Here we defined the minimum accepted distance between point pairs on the prostate surface as 80% of the measured distance for the points on the sphere. To provide the information of the point locations to the CNN, we made a binary mask of the point with the same size as the image; the voxels of the binary mask had values of one on the location of each selected point and zero everywhere else. We applied a two-dimensional (2D) Gaussian filter to the binary mask to make a normal intensity distribution around the points. This makes a gradient around each point and allows the CNN to extract features from point locations. Annotating a single voxel at each point could make it challenging for the CNN to differentiate between noise and the intended markers. We then subtracted the filtered mask from the image to annotate the image at the location of the selected points. The subtraction helps to have a good contrast between the point and the surrounding tissues regardless of the tissue intensity level. Figure 1 illustrates the annotation process steps. We used the annotated image as the single input channel of a U-Net model. For each training, validation, and test image, we repeated random point selection five times.

2.3.2 | Experiment II

We selected the points at the challenging apex and base regions to study its impact on the algorithm performance. We selected two of the points at the center of the prostate tissue on the inferiormost (apex) and superiormost (base) slices to define the prostate inferior–superior bounds manually for the network. The points were randomly distributed on the inferior third and superior third of the prostate. Any

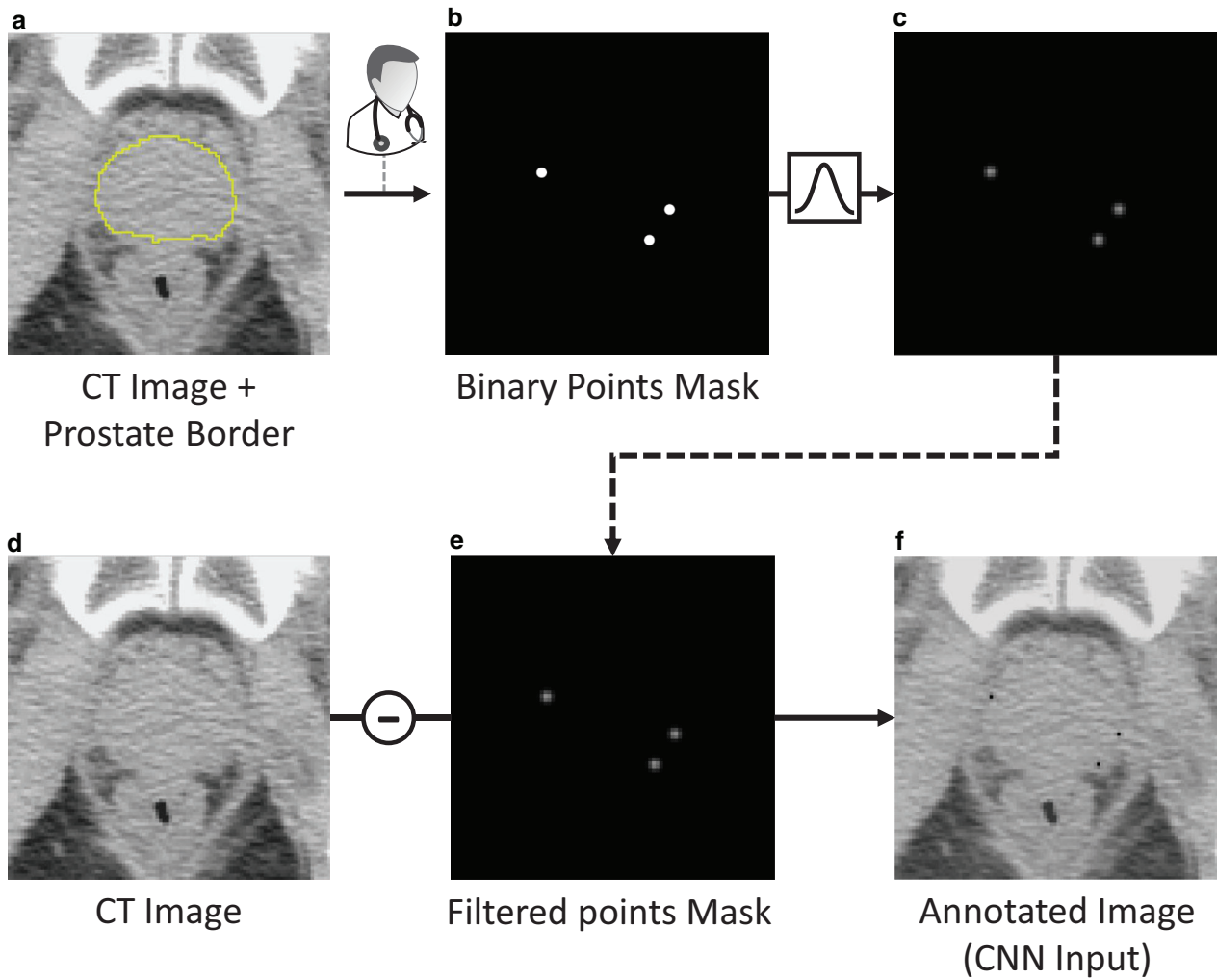


FIGURE 1 Image annotation with the selected points. (a and d) The prostate CT image; the prostate boundary overlaid on (a) in yellow as a reference, (b) the binary mask of the selected surface points, (c and e) the normal intensity distribution around the points after Gaussian filtering, and (f) the image annotated at the location of the selected points

voxel beyond the defined prostate bounds misclassified by the network was labeled as background during postprocessing.

2.3.3 | Experiment III

To quantify the sensitivity of the segmentation results to the interobserver variability in point selection, we conducted a simulation study by in-plane (intraslice) perturbation of the landmark points positions. To calculate the position of each perturbed point, we randomly sampled from 2D Gaussian distributions with mean defined on the originally selected landmark point on the reference prostate contour. The standard deviation of the Gaussian distribution was selected as 2 pixels. We used the perturbed landmark points for training and testing the network with the same setting used for Experiment I.

2.4 | Fully convolutional neural network architecture training

In this work, we customized a U-Net for 3D images, with the architecture shown in Figure 2. We trained the U-Net in two different conditions: (1) using the original CT images and (2) using the annotated CT images for training and testing the models. For both conditions, we used the Adadelta optimizer with a loss function based on “soft Dice” similarity coefficient⁵ defined as follows:

$$\mathcal{L} = 1 - \frac{2 \sum_{ijk} (p(x_{ijk}) \cdot y_{ijk})}{\sum_{ijk} (p(x_{ijk})) + \sum_{ijk} (y_{ijk})} \quad (1)$$

where $\sum_{ijk}()$ is summation of matrix elements across all the three axes, $p(x_{ijk})$ is the probability value at coordinate (i, j, k) on the output probability map, and y_{ijk} is the value of the reference binary segmentation mask at

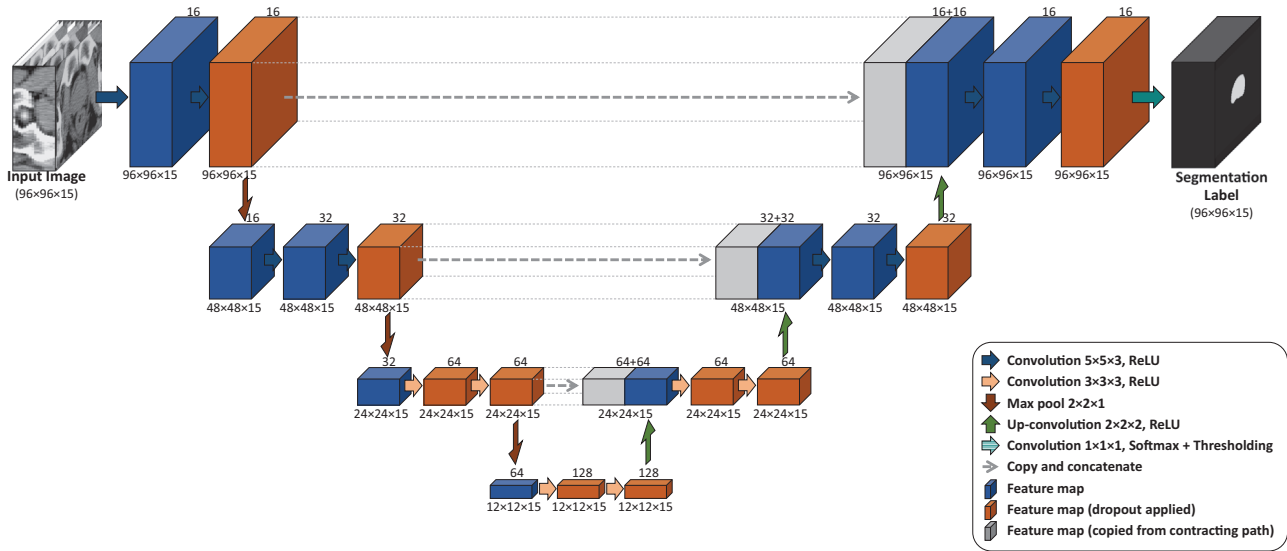


FIGURE 2 A four-level, 3D U-Net architecture used in this study. The number of feature maps and the size of each feature map at each layer have been mentioned above and below the layer, respectively. To avoid overfitting and to make the network more stable, at the layers shown in orange, 40% of the activations are randomly dropped out

the corresponding coordinate. y_{ijk} is either one (for the prostate voxels) or zero (for the background voxels).

We used the TensorFlow¹⁰ machine learning API in Python to implement the U-Net model on a desktop computer with 512 GB of memory and NVIDIA TITAN Xp GPU.

2.5 | Evaluation metrics

We used Equations (2)–(4) to measure F_1 -score or Dice similarity coefficient¹¹ (DSC), sensitivity (or recall) rate (SR), and precision rate (PR), respectively, for comparing the algorithm segmentation results against manual segmentation.

$$\text{DSC} = \frac{2(S \cap R)}{S + R} \times 100 \quad (2)$$

$$\text{SR} = \frac{S \cap R}{R} \times 100 \quad (3)$$

$$\text{PR} = \frac{S \cap R}{S} \times 100 \quad (4)$$

where S is the algorithm segmentation volume and R is the reference segmentation volume. $(S \cap R)$ is the volume of the overlap between S and R . All the metrics are reported in percent in this paper.

3 | RESULTS

Table 1 shows the results of Experiment I. It shows the mean \pm standard deviation of DSC, SR, and PR values

measured on the test data using the automatic segmentation as well as using 5, 10, 15, 20, and 30 randomly selected points. We used one-way ANOVA test followed by ad hoc t -tests (one-tailed) to detect statistically significant improvement through the use of image annotation. Figure 3 illustrates the qualitative results for three sample patients. Table 2 shows the evaluation metrics values for different regions of interest (ROIs) including whole prostate gland as well as prostate base, mid-gland, and apex subregions using 20 selected points. Here the superiormost and the inferiormost thirds of the prostate on reference segmentation labels were called base and apex, respectively, and the middle third called mid-gland. Table 3 shows the results of Experiment II when the points were selected at the apex and base regions. Each metric value in Table 3 was compared to the corresponding value in Table 1, with statistically significant improvements shown in bold.

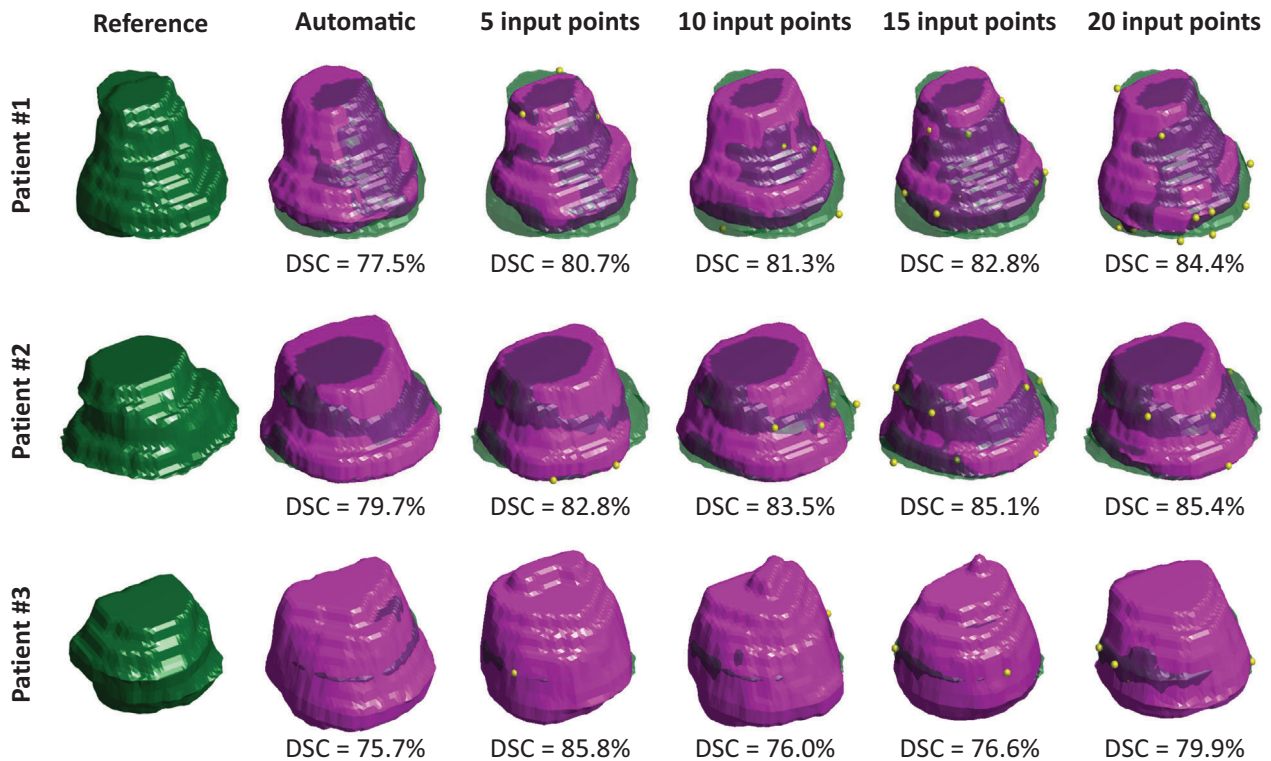
To evaluate the precision of the segmentation algorithm when different sets of points are selected, we applied a pairwise comparison between the results of the five repetitions for each prostate using DSC metric. For each prostate, we had five segmentations based on five different point selections. We measured DSC between each of the two segmentations and calculated the mean and the standard deviations of all the DSC values. In an ideal situation, the five repetitions should be identical and the DSC values should be 100%. Table 4 shows the average of the mean and standard deviation values across the test images.

Table 5 shows the results of Experiment III. The values in the table were compared to the corresponding values in Table 1 using two-tailed t -test. The values in bold shows the statistically significant difference detected between each pair of values.

TABLE 1 Quantitative results of 3D image segmentation performance on the test data when 0, 5, 10, 15, 20, and 30 points were selected

Condition	# Images (3D)	DSC (%)	SR (%)	PR (%)
Automatic	23	81.5 ± 4.5	91.5 ± 5.7	74.4 ± 9.1
5 input points	23 × 5	83.2 ± 3.8 ^{ab}	87.7 ± 8.9 ^{hi}	80.6 ± 8.7 ^{mn}
10 input points	23 × 5	83.4 ± 3.8 ^{cd}	88.2 ± 7.1 ⁱ	80.2 ± 8.6 ^{op}
15 input points	23 × 5	83.9 ± 3.5 ^{ef}	89.8 ± 6.4 ^k	79.8 ± 8.9 ^{qr}
20 input points	23 × 5	85.7 ± 3.4 ^{aceg}	89.7 ± 6.7 ^{hl}	83.0 ± 7.4 ^{moqs}
30 input points	23 × 5	90.8 ± 1.9 ^{bdfg}	95.3 ± 3.4 ^{ijkl}	87.0 ± 3.7 ^{nprs}

For each test image, random point selection was repeated five times. The values in bold indicate statistically significant improvements compared to the automatic method ($p < 0.05$). Corresponding letters show statistically significant differences between semiautomatic approaches using different number of points ($p < 0.05$).

**FIGURE 3** Qualitative results for three sample cases. The green semitransparent shapes are the reference segmentations provided by an expert radiologist, and the solid purple shapes show the algorithm results. The used points for annotation have been shown in yellow dots. The DSC value for each condition was provided underneath**TABLE 2** Quantitative results of 3D image segmentation performance on different ROIs when 20 points were selected

Condition	# Images (3D)	DSC (%)	SR (%)	PR (%)
Whole gland	23 × 5	85.7 ± 3.4	89.7 ± 6.7	83.0 ± 7.4
Base	23 × 5	82.9 ± 7.2	88.8 ± 12.8	80.4 ± 11.0
Mid-gland	23 × 5	89.9 ± 2.8	92.6 ± 4.6	87.9 ± 6.1
Apex	23 × 5	82.5 ± 6.6	87.0 ± 9.8	80.5 ± 11.4

For each test image, random point selection was repeated five times.

4 | DISCUSSION

Comparing the results shown in Table 1 for different train and test conditions, it is shown that using manually selected points fused to the input image could help the CNN segment the prostate with increased accuracy. Using only five input points could improve the DSC by about 2%. The results also showed that increasing the number of input points could improve the segmentation performance. In Table 1, although the SR value for automatic segmentation is about 2%–4% higher than that of the point-based segmentations (when 5–20 input points were selected), the PR value is about 5%–13% lower

TABLE 3 Quantitative results of 3D image segmentation performance when points were selected only at apex and base regions

Condition	# Images (3D)	DSC (%)	SR (%)	PR (%)
5 input points	23 × 5	85.1 ± 2.7 ^{abcd}	89.6 ± 7.9 ^{kl}	82.1 ± 6.8 ^{rs}
10 input points	23 × 5	85.8 ± 2.9 ^{aefg}	89.3 ± 7.5 ^{mn}	83.4 ± 6.3 ^t
15 input points	23 × 5	86.9 ± 2.4 ^{behj}	90.6 ± 6.6 ^{op}	84.3 ± 6.3 ^{ru}
20 input points	23 × 5	88.1 ± 2.0 ^{cfhj}	93.1 ± 4.0 ^{kmoq}	84.0 ± 4.8 ^v
30 input points	23 × 5	90.1 ± 2.1 ^{dgiij}	95.0 ± 2.6 ^{lnpq}	85.8 ± 4.2 ^{stuv}

For each test image, random point selection was repeated five times. Corresponding letters show statistically significant differences between approaches correspond to different numbers of points ($p < 0.05$). The values in bold are significantly improved compared to the corresponding value in Table 1 ($p < 0.05$).

which means a higher false positive rate. Moreover, it is important to note that the SR and PR values need to be interpreted together to have a better understanding of the algorithm performance. For example, a high SR and a low PR means the algorithm is oversegmenting. Therefore, the results in Table 1 show that using the points help the network to mitigate the oversegmentation observed in automatic segmentation in this case.

The DSC value observed in this study started from about 82% for automatic segmentation and increased to about 91% when up to 30 landmark points were involved. The results in Table 2 show that the algorithm performed more accurately at the midland subregion compared to apex and base subregions, while the results of Table 3 indicate that DSC could be improved in general using points distributed on challenging apex and base regions only. However, comparing the last rows of Tables 1 and 3 shows that when 30 input points were used, the performance of the algorithm was significantly better when the points were well distributed on the prostate surface. Zhang et al.⁸ reported a lower average DSC (88%) with substantially higher standard deviation (11%) compared to our best segmentation results (DSC of $\sim 91\% \pm 2\%$) using a complicated state-of-the-art algorithm based on general adversarial network. Their method is a 2D slice-by-slice segmentation. Comparing our results to the results of a nondeep learning method based on incorporation of patient-specific knowledge (DSC = $85\% \pm 3\%$)¹² also demonstrates that our approach outperformed their method even when only 10 points have been used at apex and base regions.

Comparing the DSC trend (from 81.5% to 90.8% using 0 to 30 input points, respectively) to the average interexpert observer difference in manual prostate segmentation in CT on the same test dataset ($\sim 92\%$)¹³ shows that using minimal user interaction could improve the segmentation results to approach the expert observer performance. The DSC score between the two manual segmentations were measured as the interobserver difference, and the average of DSC scores across the test set was reported as the average interobserver difference.

The high values in Table 4 show that the algorithm is precise and the segmentation performance is robust to

the changes in the selected landmark points. The average variability in the results increased up to about 8% when the selected landmarks were changed. This variability is zero for automatic segmentation, as no landmark point is used (highest precision). Interpreting the results in Tables 1 and 4 shows that adding the landmark points and increasing the number of the selected points increased the accuracy of the algorithm while its precision dropped slightly. In automatic segmentation, the CNN uses features extracted from the image only. However, when the images are annotated with landmarks, the CNN uses a combination of the features extracted from the input image and the input points. The greater the number of points used, the fewer the features extracted from the image as the CNN is more focused on the landmarks. Therefore, the variability in the location of the randomly selected landmarks impacted the precision of the algorithm especially when the number of selected landmarks was increased. By applying random perturbation to the location of the points to mimic interobserver variation in selecting the point, we did not detect a statistically significant drop in the DSC values when 5–15 input points were used (see Table 5). However, using more input points decreased the DSC by up to about 3%. By adding more input points, the CNN relies on the point information more and the perturbation could have a higher impact on its performance.

In this study, we used a uniform random or semirandom distribution of the points on the prostate surface to initialize the algorithm. This approach could be considered as a general form of systematic approaches for point selection (similar to what used in references 14 and 9) and might be generalizable to those conditions. However, the findings of this research study need to be tested and confirmed using a more realistic point selection such as systematic or free form point selection approaches.

As another observation, we have seen some cases where unexpectedly the network results did not pass through the input points (e.g., see 20 input points case for patient #1 in Figure 3). This may be due to the stronger image-derived features extracted by the network compared to the features extracted based on the point annotation. Modifying the annotation approach

TABLE 4 The precision of the segmentation algorithm

	5 input points	10 input points	15 input points	20 input points	30 input points
Precision (%)	97.0 ± 0.3	96.7 ± 0.4	96.5 ± 0.6	95.7 ± 0.5	91.9 ± 1.2

The precision was measured based on five repetitions in random landmark point selections for each test image. The precision value for each case is measured using the average DSC value calculated between different algorithm result pairs.

TABLE 5 Quantitative results of 3D image segmentation performance when the selected points on the reference contour were perturbed randomly around the border points

Condition	# Images (3D)	DSC (%)	SR (%)	PR (%)
5 input points	23 × 5	81.7 ± 3.8	84.2 ± 10.7	81.3 ± 8.9
10 input points	23 × 5	82.5 ± 4.1	86.5 ± 8.1	80.2 ± 8.8
15 input points	23 × 5	83.1 ± 3.6	86.9 ± 8.8	81.1 ± 8.9
20 input points	23 × 5	84.3 ± 2.9	89.4 ± 6.4	80.7 ± 7.7
30 input points	23 × 5	87.7 ± 2.5	92.9 ± 4.6	83.4 ± 5.4

For each test image, random point selection was repeated five times. The values in bold are significantly different from the corresponding value in Table 1 ($p < 0.05$).

could be helpful for better detection of the points by CNN and might improve the segmentation performance.

Using this method could decrease the prostate segmentation time. Our previous studies showed that selecting border points on CT images for initializing the segmentation took less than 20 s per 3D image¹⁴, which is several times shorter than fully manual segmentation time.

The idea of supervising a CNN-based segmentation by manual interaction could be tested on other datasets for different organs.

For this study, we used a set of well-distributed, randomly selected surface landmarks to mimic an operator's interaction. However, in a real situation, the points manually selected by an operator are not necessarily well distributed on the prostate surface and thus our assumption is not always true. The number of selected points from case to case could also be different and the CNN model might need to be trained and tested using variable numbers of selected points. For this work, we used the hyperparameters from our previous study for prostate segmentation in CT images.¹³ We also avoided max pooling along z direction according to our previous observations.¹³ Optimizing the hyperparameters of the network to this specific task could improve the results. As another limitation of this study, the image volumes were automatically cropped around the prostate region due to the limitation of the GPU memory capacity in accommodating the image and the feature maps. This added a step of user interaction (i.e., defining a generous bounding box around the prostate) to the segmentation process. In future, we need to study the impact of prostate localization accuracy and the cropping size on the algorithm performance. We compared the results of this study to the interobserver difference observed in manual segmentation to take the variability into account. In future, for a more comprehensive evaluation of the

algorithm, it is required to compare the results to multiple expert observers.

5 | CONCLUSIONS

We proposed a segmentation approach to guide the CNN with a set of selected border landmarks for a better CT segmentation performance. We used U-Net as a full CNN because it is widely used in medical image segmentation. We annotated CNN's input images using the selected points to incorporate the manual interaction in training and testing the network. The results showed an improved segmentation performance by using a limited number of sparse border landmarks annotated on the image. By selecting 30 landmark points, the CNN performance was close (DSC < 2%) to the average interexpert observer difference observed in manual segmentation.

ACKNOWLEDGMENTS

This research was supported in part by the U.S. National Institutes of Health (NIH) grants (R01CA156775, R01CA204254, R01HL140325, and R21CA231911) and by the Cancer Prevention and Research Institute of Texas (CPRIT) grant RP190588.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- Smith WL, Lewis C, Bauman G, et al. Prostate volume contouring: a 3D analysis of segmentation using 3DTRUS, CT, and MR. *Int J Radiat Oncol Biol Phys*. 2007;67:1238-1247.
- Du G, Cao X, Liang J, Chen X, Zhan Y. Medical image segmentation based on u-net: a review. *J Imaging Sci Technol*. 2020;64(2): pp. 20508-1-20508-12(12).
- Lee J-G, Jun S, Cho Y-W, et al. Deep learning in medical imaging: general overview. *Korean J Radiol*. 2017;18:570-584.

4. Lai M, Deep learning for medical image segmentation. arXiv:1505.02000, 2015.
5. Milletari F, Navab N, Ahmadi S-A, V-net: fully convolutional neural networks for volumetric medical image segmentation. In: *2016 Fourth International Conference on 3D Vision (3DV)*, 2016, 565-571.
6. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal.* 2017;42:60-88.
7. 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*, 2015, 234-241.
8. Zhang Z, Zhao T, Gay H, Zhang W, Sun B. ARPM-net: a novel CNN-based adversarial method with Markov random field enhancement for prostate and organs at risk segmentation in pelvic CT images. *Med. Phys.* 2021;48:227-237.
9. Girum KB, Créhange G, Hussain R, Lalande A. Fast interactive medical image segmentation with weakly supervised deep learning method. *Int J Comput Assist Radiol Surg.* 2020;15:1437-1444.
10. Abadi M, Barham P, Chen J, et al. Tensorflow: a system for large-scale machine learning. *OSDI.* 2016:265-283.
11. Dice LR. Measures of the amount of ecologic association between species. *Ecology.* 1945;26:297-302.
12. Ma L, Guo R, Tian Z, et al. Combining population and patient-specific characteristics for prostate segmentation on 3D CT images. *Med Imag 2016: Image Process.* 2016;9784:978427.
13. Shahedi M, Halicek M, Dormer JD, Schuster DM, Fei B. Deep learning-based three-dimensional segmentation of the prostate on computed tomography images. *J. Med Imaging.* 2019;6:025003.
14. Shahedi M, Halicek M, Guo R, Zhang G, Schuster DM, Fei B. A semiautomatic segmentation method for prostate in CT images using local texture classification and statistical shape modeling. *Med. Phys.* 2018;45:2527-2541.

How to cite this article: Shahedi D, Dormer JD, Halicek M, Fei B. Technical note: The effect of image annotation with minimal manual interaction for semiautomatic prostate segmentation in CT images using fully convolutional neural networks. *Med. Phys.* 2022;49:1153–1160.
<https://doi.org/10.1002/mp.15404>