# PROST-Net: A Deep Learning Approach to Support Real-Time Fusion in Prostate Biopsy

Luigi Palladino, Bogdan Maris, Alessandro Antonelli, and Paolo Fiorini, *Life Fellow, IEEE*

***Abstract*—Prostate biopsy fusion systems employ manual segmentation of the prostate before the procedure, therefore the image registration is static. To pave the way for dynamic fusion, we introduce PROST-Net, a deep learning (DL) based method to segment the prostate in real-time. The algorithm works in three steps: firstly, it detects the presence of the prostate, secondly defines a region of interest around it, discharging other pixels of the image before the last step which is the segmentation. This approach reduces the amount of data to be processed during segmentation and allows to contour the prostate regardless of the image modality (e.g., magnetic resonance (MRI) or ultrasound (US)) and, in the case of US, regardless of the geometric disposition of the sensor array (e.g., linear or convex). PROST-Net produced a mean Dice similarity coefficient of 86% in US images and 77% in MRI images and outperformed other CNN-based techniques. PROST-Net is integrated in a robotic system–PROST–for trans-perineal fusion biopsy. The robot with PROST-Net gives the potential to track the prostate in real-time, thus reducing human errors during the biopsy procedure.**

***Index Terms*—Deep learning, image segmentation, medical robotics, image fusion.**

## I. Introduction

The prostate cancer is the most common diagnosed cancer among men, with an incidence of 31.1 per 100000 [1]. Diagnosis is by biopsy. The biopsy is usually performed by manually inserting the biopsy needle which is guided through US images following a pre-defined scheme. When pre-operative magnetic resonance images (MRI) are available and the lesion site was identified, a so-called fusion allows for precise sampling. This approach requires physicians to perform manual segmentation of MRI (pre-operative) and sometimes also of US images, just before the biopsy procedure. By minimizing procedure time with automatic segmentation, we could provide a more efficient workflow with improved outcome [2]. The newest and most efficient algorithms for segmentation of medical images apply deep learning techniques and the methods are based on Convolutional Neural Networks (CNNs). CNNs operate usually on two-dimensional (2D) images and most techniques are derived from U-Net [3]. In 2D, TRUS is tracked with a mechanical stepper or other tracking methods in order to obtain a 3D reconstruction. In MRI images, where a common reference system is defined, the 2D contours have already a 3D correspondence. The soft tissue contrast in prostate MRI and a high signal to noise ratio permit to accurately localize the target to be segmented [4]. The US image of the prostate is more challenging due to noise, image artifacts and less availability of data sets, but recent works showed promising results (e.g., [5]). The CNNs based algorithms allows for real-time results, therefore the method can be useful in detecting the deformations of the prostate without the need of a prior deformation model. This will lead to the introduction of new devices for more precise procedures and diagnoses [6].

The objective of this paper is to introduce PROST-Net, a CNN for real-time segmentation of the prostate in US and MRI images. PROST-Net offers support to a robotic prostate biopsy system for automatic fusion and dynamic update of the procedure. Clinical devices for prostate fusion biopsy were developed by Philips Medical Systems, Netherlands (Uronav), Hitachi, Japan (Virual Sonography), Esaote, Italy (Virtual Navigator), Koelis, France (Trinity) and others. None of these systems employ both MRI and US real-time segmentation as implemented with PROST-Net. The real-time US segmentation permits the use of completely automatic fusion techniques as in [8] and the dynamic update of the fusion during the procedure. The algorithm is tested on bi-planar US images (axial and sagittal) that use different configurations of the sensor array (convex and linear), then on MRI-T2 images since both types of data are used when performing fusion in robotic assisted biopsy procedure. A proof of concept of the robotic fusion biopsy was tested on phantoms [9], while the complete validation of the segmentation algorithm is performed on patient data (Figure 2). We describe PROST-Net architecture in Sections II and II-A, while Section III covers the experimental setup and the results.

## II. PROST-Net Architecture

The method we are proposing here derives from Mask R-CNN [7] and U-Net [3] architectures, assembling the best of both for the purpose of prostate segmentation. PROST-Net operates on 2D images. It takes the detection part from Mask R-CNN to find a region of interest (ROI), which is a bounding box that perfectly fits the prostate shape. This will allow to the segmentation sub-network to contour the prostate without the interference of useless information like the pixels outside of the ROI.

In our implementation, the ROI extracted by the region proposal network (Figure 1) is given as input to the second part of the network which exploits the U-Net architecture, but with application of residual blocks and dropout layers (Figure 1). This approach permitted to take advantage of the strong part of a regional network, but also the typical aspects of using a "U-Net like" architecture, which works better on medical images, where typically the amount of data is limited. The introduction of residual blocks allowed us to have deeper contraction and expansion steps, while the introduction of dropout layers is used for regularization, to prevent overfitting. We tested the benefits of the modifications proposed by performing two types of ablation studies. The first one used the original U-Net architecture as segmentation head. This was meant to assess the benefit of adding an

Manuscript received July 20, 2021; revised October 3, 2021; accepted January 7, 2022. Date of publication January 25, 2022; date of current version May 18, 2022. This article was recommended for publication by Associate Editor G. Mylonas and Editor P. Dario upon evaluation of the reviewers' comments. This work was supported by the European Research Council (ERC-2019-POC) through the European Union's Horizon 2020 Research and Innovation Programme (PROST) under Grant 875523. *(Corresponding author: Luigi Palladino.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional and/or National Research Committee (under Application No. 3167CESC, and performed in line with the Helsinki Declaration).

Luigi Palladino, Bogdan Maris, and Paolo Fiorini are with the Department of Computer Science, University of Verona, 37134 Verona, Italy (e-mail: luigi.palladino@univr.it).

Alessandro Antonelli is with the Department of Urology, Verona University Hospital, 37126 Verona, Italy.

Digital Object Identifier 10.1109/TMRB.2022.3145667

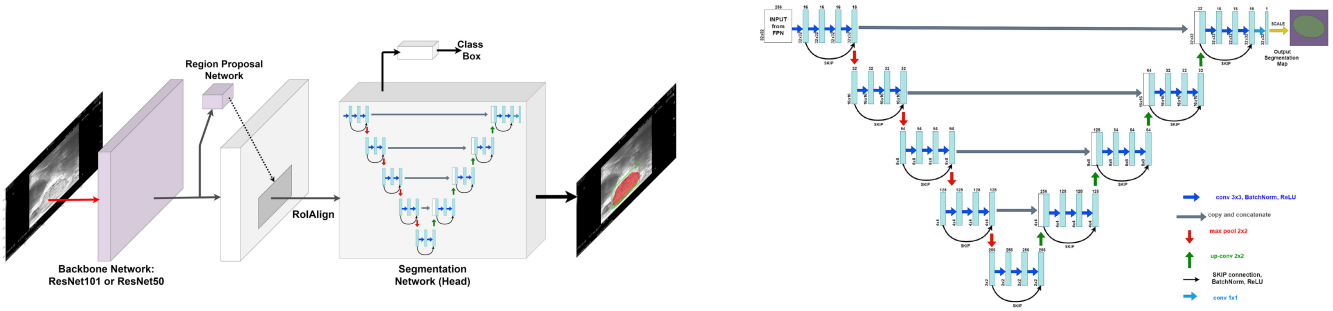


Fig. 1. **Left:** PROST-Net architecture as derived from the seminal work [7]. Each sub-network is optimized to fit PROST-Net. **Right:** PROST-net segmentation sub-network inspired by U-Net. Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the boxes. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations. Black arrows are skip connections.

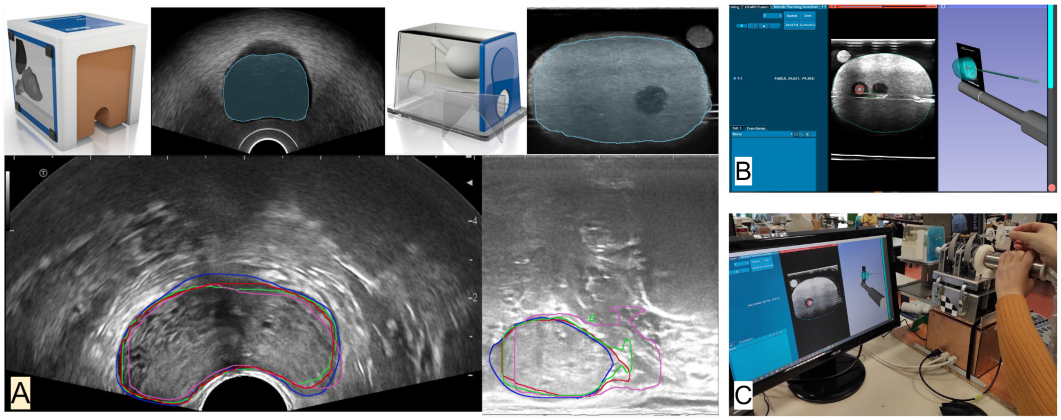


Fig. 2. **A:** On top row CIRS 070 (left) and CIRS 053 (right) prostate phantoms. Axial US taken from CIRS070 (left) and sagittal US from CIRS053 (right) with corresponding segmentation masks. At the bottom axial (left) and sagittal (right) US images of the prostate ans segmentation contours. The blue line shows the ground truth. The other lines show the result of the automatic segmentation performed with PROST-Net (Red), Mask R-CNN with ResNet50 (Green) and Mask R-CNN with ResNet101 (Pink). **B:** GUI, implemented in 3D Slicer(https://www.slicer.org/), of the surgical vision support system presenting the segmentation of the prostate shape. The segmentation using PROST-Net was integrated through a python module. **C:** Proposed surgical vision support system tested on a robotic system for prostate biopsy [9].

extra convolution layer. The second added the extra convolution layer without including the skip connections. This was done to confirm the importance of the added residual blocks.

PROST-Net network was initialized with the original COCO dataset training weights [10] to improve performance and significantly reduce the training time applying fine-tuning, as demonstrated in [11]. The network architecture is divided in three parts (see Figure 1): the backbone part, which uses ResNet101 or ResNet50 [10], the Region Proposal Network (RPN) [12] for the bounding-box (or region of interest ROI) selection and the head of the network, which outputs the segmentation mask. The loss function was adapted from [7] to our specific case with only one class: the prostate. Training of the network develops in two phases: the first one consists of fine-tuning of head layers, freezing the entire ResNet backbone; the second one unfreezes the layers of the backbone up to the 4th stage. In this phase, the learning rate (LR) is reduced by a factor of 10 from the previous training so that the model will converge easier. The optimizer used is Adam. We start with a LR of 0.001. Further implementation details at: https://gitlab.com/altairLab.

### A. *PROST-Net's Head Network for Segmentation*

While PROST-Net backbone and RPN are essentially the same as the one used in Mask R-CNN, the head part for segmentation derives from U-Net architecture (Figure 1). The contraction path consists of four levels of convolution and max pooling layers. The main difference of PROST-Net's segmentation head from U-Net lays on the presence of three convolutional layers instead of the typical two. This deepening in convolution block is compensated by adding a skip connection as the one used in ResNet [10] between the first and third convolution with an Add operation. This operation transforms the convolution block in a residual block capable of better fitting training data, without suffering from vanishing of the gradient. The deepening of the network permitted to lower the number of channels in the network at each layer, with a peak of 256 channels as opposed to original 1024.

## III. Experimental Setup and Results

A first comparative experiment was performed on realistic prostate phantoms. The images were acquired using an Ultrasonix US machine (BK Medical, Peabody, Massachusetts, USA) from two standard synthetic phantoms (CIRS 070 and CIRS 053, Computerized Imaging Reference Systems, Inc. (CIRS), Norfolk, Virginia, USA) containing the prostate with urethra and seminal vesicles, the bladder, and the rectum (Figure 2). The dataset is composed by axial and sagittal 2D images. From this dataset we kept 20% of the images for testing and the remaining images were split with a ratio of 0.8 in training set and 0.2 in validation set. Ground truth was created using a semi-automatic procedure. An average DC precision score of 80% was obtained on the using U-Net architecture. With Mask R-CNN configured to have ResNet101 as backbone, we pushed the DC precision score up to

TABLE I
SEGMENTATION PERFORMANCE ON 2D IMAGES

| Architecture | | Dice Coefficient | | | Imaging |
|---|---|---|---|---|---|
| *Method* | *Backbone* | *Train* | *Validate* | *Test* | *Type* |
| U-Net | NA | 0.8264 | 0.8112 | **0.8002** | Phantom US |
| Mask R-CNN | ResNet101 | 0.8315 | 0.8219 | **0.8575** | Phantom US |
| PROST-Net | ResNet50 | 0.8753 | 0.8787 | **0.8912** | Phantom US |
| Number of Images | | 2201 | 550 | 687 | Phantom US |
| Mask R-CNN | ResNet101 | 0.9022 | 0.8980 | **0.8436** | US |
| Mask R-CNN | ResNet50 | 0.9068 | 0.9060 | **0.8587** | US |
| PROST-Net | ResNet50 | 0.9207 | 0.9192 | **0.8693** | US |
| Ablation Test 1 | ResNet50 | | | **0.8306** | US |
| Ablation Test 2 | ResNet50 | | | **0.8539** | US |
| Number of Images | | 668 | 172 | 664 | US |
| PROST-Net | ResNet50 | 0.7876 | 0.7834 | **0.7773** | MRI |
| Ablation Test 1 | ResNet50 | | | **0.7100** | MRI |
| Ablation Test 2 | ResNet50 | | | **0.7524** | MRI |
| Number of Images | | 18642 | 2128 | 5148 | MRI |
| PROST-Net | ResNet50 | | | **0.7709** | MRI (Local Hospital) |
| Number of Images | | | | 735 | MRI (Local Hospital) |

85% (Table I). U-Net training time was about 10 hours, while training time of Mask R-CNN and PROST-Net was less than 2 hours on a commercial laptop with an Nvidia RTX 2060m GPU. The difference in training time underlines the importance of the transfer learning in the initialization of the CNNs. This becomes of critical importance in a clinical scenario, where the data is added constantly and a server cluster of GPUs is not always available. These results helped us on establishing the training strategy described in the previous section, which came very useful when we dealt with patient data.

A second phase of experiments involved the use of a US dataset of patients that included both axial and sagittal images recorded with Hitachi Aloka 70 (Hitachi Healthcare Americas, Twinsburg, USA) and a biplanar probe. The data and the ground truth were obtained with the support of physicians after the approval of the local ethics committee and include data from 33 different patients, of which 12 are used for training and 11 for testing. Ground truth was obtained using a semi-automatic interface written in MEVISLAB. Each segmentation was validated by a physician. We compared the results of three different algorithms: the first two employed standard Mask-R-CNN but with different backbones (ResNet101 and ResNet50 respectively), while the third one used PROST-Net with ResNet50 as backbone.

Finally, a third phase of experiments explored the application of PROST-Net on 2D slices extracted from 3D MRI images. The training was performed on data taken from Cancer Imaging Archive [13]. PROST-Net produced promising results reaching 78% DC precision (Figure 3). The model was then tested on MRI data corresponding to 6 new patients acquired from local hospitals in accordance with ethical approval. We achieved a DC score of 77%. The precision obtained with these data was in line with the previous results showing that our method is ready to be implemented in a clinical scenario.

### A. Discussion

The accuracy was calculated using the Dice Coefficient (DC). Overall, for each optimization of the algorithm (U-Net, Mask R-CNN with different backbones, and PROST-Net) we obtained an improvement in accuracy in both phantom tests and patient tests (Table I) and for both US and MRI.

A further analysis of ROI performance was made for PROST-Net (Table II). We used mean Average Precision (mAP) adapted to the binary case to measure *detection* performance as described in [10]. To decide between false and positive cases we performed Intersection

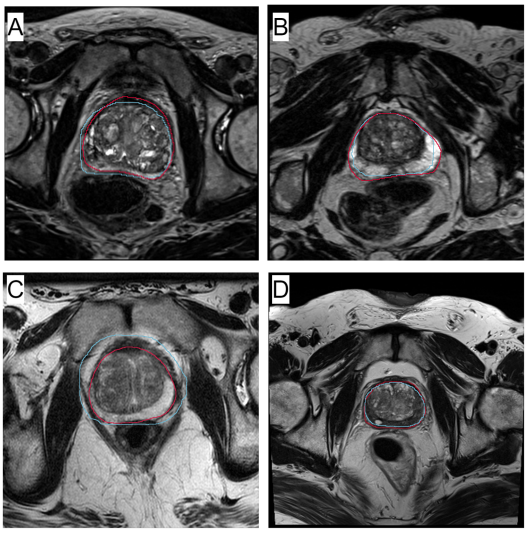


Fig. 3. Comparison between ground truth prostate shape (Red) and the automatic segmentation performed with PROST-Net (Blue). **A**, **B** and **C** show the different types of MRI images present in [13] at respective resolutions of 256x256, 512x512 and 672x672. **D** show results performed on an MRI image from local hospital of resolution 448x448.

TABLE II
ROI PERFORMANCE ON 2D IMAGES

| | IoU = 0.5 | IoU = 0.7 | IoU = 0.5:0.95 | Type | Test |
|---|---|---|---|---|---|
| mAP | 0.9065 | 0.7596 | 0.6101 | US | 664 |
| mAP | 0.9191 | 0.8062 | 0.6555 | MRI | 5148 |

over Union (IoU) between ground truth and predicted bounding box with different threshold values. We evaluated also the memory consumption and the real-time feasibility of PROST-Net. The computed peak memory at prediction time is 251 MB. This was confirmed by measuring the memory consumption during execution. A base memory footprint of 1.1 GB was observed with an average peak of memory to 1.34 GB during prediction. The prediction time per frame measured was 370ms on average. This is compatible with low frame rate real time scenario such as US image streams. Measurement was performed on a commercial laptop so it can be considered feasible for real time in most machines.

Mask R-CNN with our training pattern outperformed U-Net on the same dataset, ascertaining the superiority of the region-based approach over the classical one usually employed for semantic segmentation. It seem so be confirmed the thesis that considering only the pixel in the ROI containing the prostate as possible pixel for the final mask produce better segmentation results than considering the whole pixels in the image.

The difference is small, but on our data, of the two implementations, the one with ResNet50 performed better than the one with ResNet101. For these reason we chose ResNet50 also for the PROST-Net testing. Overall, PROST-Net outperformed Mask R-CNN and U-Net in the first case study and produced slightly better results also with patient data against Mask R-CNN. The improvements of the results are due to the inclusion of a modified U-Net. This was confirmed by the ablation study on PROST-Net. The ablation results are included in Table I. PROST-Net performed in a similar way also in the MRI dataset. The MRI segmentation is performed offline, while the 3D US segmentation and reconstruction from 2D frames is performed only once, before the procedure, rotating the tracked probe. The segmentation time of 370ms, that produces about 3 frames per second,

is enough to give real-time feedback to account for prostate deformations during the biopsy. These characteristics allowed us to include the method in the PROST robotic setup and to test the segmentation-based fusion on prostate phantoms [9]. On patient data, we tested the output of PROST-Net for the fusion of the data from the public dataset [13] using a 3D Slicer prostate fusion module [8].

## IV. Conclusion

This study investigated the implementation of a segmentation network able to contour the prostate in US and MRI images. PROST-Net optimized Mask R-CNN to improve the segmentation accuracy and to is able to work with multiomodal prostate images. The MRI-US segmentation will be used for the fusion of pre-operative images with intra-operative data. The continous segmentation of US images will make it possible to update the initial alignment and to compute local deformation. These will lead to update the position of the target in real-time during the procedure.

The realization of such a complex system is not an easy task and involves the synergy of several domains: mechatronic systems, software design and clinical support. Compared with other methods that employs CNN for real-time segmentation of the prostate, PROST-Net is, as far as we know, the first approach to handle different image modalities and to be integrated in a robotic system. Real-time segmentation and robotic tracking will allow dynamic update of the MRI-US fusion.

## References

[1] C. H. Pernar, E. M. Ebot, K. M. Wilson, and L. A. Mucci, "The epidemiology of prostate cancer," *Cold Spring Harbor Perspect. Med.*, vol. 8, no. 12, 2018, Art. no. a030361.

[2] N. Orlando, D. J. Gillies, I. Gyacskov, C. Romagnoli, D. D'Souza, and A. Fenster, "Automatic prostate segmentation using deep learning on clinically diverse 3D transrectal ultrasound images," *Med. Phys.*, vol. 47, no. 6, pp. 2413–2426, 2020.

[3] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent.*, 2015, pp. 234–241.

[4] E. M. A. Anas, P. Mousavi, and P. Abolmaesumi, "A deep learning approach for real time prostate segmentation in freehand ultrasound guided biopsy," *Med. Image Anal.*, vol. 48, pp. 107–116, Aug. 2018.

[5] N. Ghavami *et al.*, "Integration of spatial information in convolutional neural networks for automatic segmentation of intraoperative transrectal ultrasound images," *J. Med. Imag.*, vol. 6, no. 1, 2018, Art. no. 11003.

[6] F. J. Siepel, B. Maris, M. K. Welleweerd, V. Groenhuis, P. Fiorini, and S. Stramigioli, "Needle and biopsy robots: A review," *Current Robot. Rep.*, vol. 2, pp. 73–84, Jan. 2021.

[7] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2961–2969.

[8] "SlicerRt/SegmentRegistration." May 2017. [Online]. Available: https://github.com/SlicerRt/SegmentRegistration

[9] B. Maris *et al.*, "Toward autonomous robotic prostate biopsy: A pilot study," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 16, no. 8, pp. 1393–1401, 2021.

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2016, pp. 770–778.

[11] M. A. Morid, A. Borjali, and G. Del Fiol, "A scoping review of transfer learning research on medical image analysis using ImageNet," *Comput. Biol. Med.*, vol. 128, Jan. 2021, Art. no. 104115.

[12] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," 2015, *arXiv:1506.01497*.

[13] S. Natarajan, A. Priester, D. Margolis, J. Huang, and L. Marks. "Prostate MRI and Ultrasound With Pathology and Coordinates of Tracked Biopsy (Prostate-MRI-US-Biopsy)." 2020. [Online]. Available: https://wiki.cancerimagingarchive.net/x/BQAWB