Deep learning model for automatic prostate segmentation on bicentric T2w images with and without endorectal coil

Deep learning model for automatic prostate segmentation on multi-center T2w images with and without endorectal coil

Davide Barra, Giulia Nicoletti, Arianna Defeudis, Simone Mazzetti, Jovana Panic, Marco Gatti, Riccardo Faletti, Filippo Russo, Daniele Regge, Valentina Giannini

***Abstract*— Automatic segmentation of the prostate on Magnetic Resonance Imaging (MRI) is one of the topics on which research has focused in recent years as it is a fundamental first step in the building process of a Computer aided diagnosis (CAD) system for cancer detection. Unfortunately, MRI acquired in different centers with different scanners lead to images with different characteristics. In this work we propose an automatic algorithm for prostate segmentation, based on a U-Net applying transfer learning method in a multicenter setting. First, T2w images with and without endorectal coil from 80 patients acquired at one center were used in this work as training set and internal validation set. Then, T2w images without endorectal coil from 20 patients acquired at a different institution were used as external validation. The reference standard for this study was manual segmentation of the prostate gland by an expert operator. The results showed a Dice similarity coefficient >85% in both internal and external validation datasets.**

***Clinical Relevance*— This segmentation algorithm could be integrated into a CAD system to optimize computational effort in prostate cancer detection.**

I. Introduction

Prostate cancer (PCa) accounts for about 26% of all cancers diagnosed in men. It is the most common cancer in men and ranks second in mortality after lung cancer [1]. Early diagnosis and detection of PCa is crucial to ensure the best chance of cure [2]. Until publication of the last European and American Urological Guidelines [3][4], the standard diagnostic procedure for PCa diagnosis included prostate specific antigen (PSA) test, digital rectal exam (DRE) and transrectal ultrasound (TRUS) guided biopsy to confirm the presence of the tumor. This workup has shown limits in tumor detection and aggressiveness estimation [5]. In the last decade, multiparametric magnetic resonance imaging (mpMRI) has shown high accuracy in lesion detection and characterization [6] and Computer aided diagnosis (CAD) systems have been proposed to help radiologists to automatically detect PCa, further increasing their detection accuracy, reducing reading time and increasing inter-reader agreement [7].

CAD systems are composed of different modules, including prostate segmentation. The latter is a crucial step for the subsequent recognition of the tumor and is a challenging task due to resolution, artifacts and general appearance of MRI, all of which are strongly influenced by the type of MRI scanner and protocols used for imaging acquisition. The result should be very accurate, to avoid excluding tumor areas from the prostate segmentation.

Recent studies have demonstrated that deep learning methods to segment medical image has led to significant improvements with respect to other approaches, especially with the increasingly robust development of Convolutional Neural Networks (CNNs) and U-Net [8]–[12]. Most of previous studies validated their algorithms on internal datasets through cross validation or through images from the same center/centers used to train the nets. These approaches have the limit of not considering the differences that may exist in the acquisition of images in different centers and with different MRI scanners. To the best of our knowledge, only Zavala-Romero *et al.* [9] and Liu *et al.* [10] proposed a multi-MRI vendor and a multi-site network. They showed an average Dice similarity coefficient between 89% and 91.7%.

In this work, we propose an automatic algorithm based on a U-Net to segment the whole prostate gland on T2w MRI acquired both with and without endorectal coil. The algorithm was trained with images from one center and validated using both images from the same center and from a different hospital.

II. Material and Methods

A. Dataset and reference standard

T2w images at the Candiolo Cancer Institute (Center A) were acquired using a 1.5T scanner (Optima MR450w, GE Healthcare, 106 Milwaukee, WI, USA) and with either a single 32-channel phased-array coil or with a 32-channel phased-array coil combined with an endorectal coil (Medrad, Indianola, Pa). All T2w axial sequences have a field of view of 160mm x 160mm with voxel size of 0.31mm x 0.31mm x 3mm. T2w images acquired at A.O.U Città della Salute e della Scienza (Center B) were acquired using a 1.5T scanner (Achieva, Philips Medical System, Eindhoven, The Netherlands) with field of view of 180mm x 180mm and voxel size of 0.47mm x 0.47mm x 3.3mm. An expert operator with five years of experience manually contoured the prostate gland in all acquired images. The local ethics committee approved this retrospective study. The training set used for developing the network consists of both T2w images with and without endorectal coil from the Center A. Patients from Center B were not used in the learning process and they were used as an external validation dataset.

D.B., G.N, A.D., S.M., V.G., M.G., R.F. and D.R. are with University of Turin, Department of Surgical Sciences, via Genova 3, 10126 Torino, Italy (corresponding author email: davide.barra@unito.it).

F.R., D.R. are with Candiolo Cancer Institute, FPO -IRCCS, Strada Provinciale 142, km 3.95, Candiolo (TO), Italy.

J.P. is with Polytechnic of Turin, Department of Electronics and Telecommunications, Torino, Italy

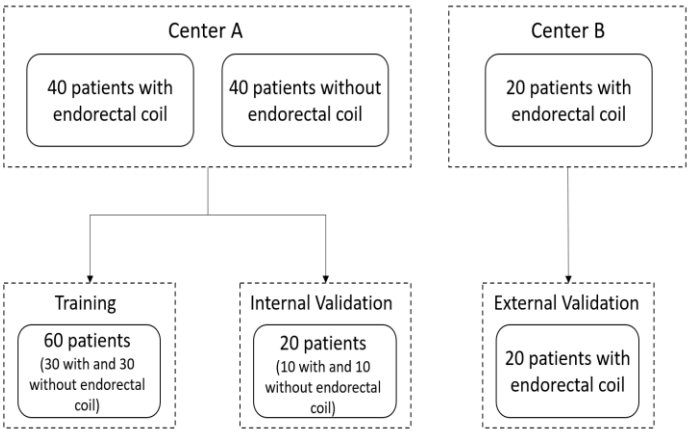

Figure 1. Composition of the dataset

B. Automatic Segmentation Algorithm

The automatic segmentation algorithm was based on a U-Net [13] that is a network characterized by two different paths: the descent one (encoder) and the ascent one (decoder). Differently from the standard convolutional neural network (CNN), where the original image is converted into a vector and the features maps are extracted, the features maps learned in the contraction phase (encoder) are used in the expansion phase (decoder) to get a new output image of the same size as the input image. Each pixel of the output image is characterized by a value that represents the probability of belonging to a class. For this particular purpose, we applied the transfer learning method building a U-Net where the encoder path is a known convolutional neural network called Residual neural network (ResNet) [14].

ResNet-18 is an 18-layers deep convolutional neural network made up of several residual blocks. The peculiarity that makes ResNet unique compared to other CNN is the presence of the skip connections that link the original input to the output of each convolutional block. This strategy helps to make deeper networking and reaching outstanding results in segmentation tasks [14].

Since our goal is to obtain a network that is able to segment images acquired in different modalities and from different centers, it is important to avoid overfitting. To this end, we exploited the concept of transfer learning. More specifically, we used the ResNet-18 with pretrained weights on the ImageNet database [15] which allows faster and better convergence. Each layer from which the encoding path is composed consists of two connections from the input. The first one goes through two 3x3 convolutional blocks followed by a batch normalization layer used to make artificial neural networks faster and more stable and followed by an activation function (ReLU, Rectified Linear Unit). The second one skips all these functions and is added to the first connection at the end of the layer.

Each decoding path is characterized by a transposed convolutional function to up-sample feature maps followed by a batch normalization and a ReLU, concluding with several concatenations of feature maps from the encoding path performed at different level of the network.

The training set used for network learning consists of both T2w images with and without endorectal coil coming from the Center A. Patients from Center B were kept out the learning process and used as an external validation dataset. All MRI slices were fed to the network, even those without the prostate, to allow the model to learn where the prostate bounds were located. The original image was resized to 256 x 256 pixels to ease the network training and reduce computational cost. Furthermore, all slices of each patient were standardized with the z-score formula to obtain zero mean and unit standard deviation:

$$Img_{normalized} = \frac{Img - \mu}{\sigma} \quad (1)$$

Where μ and σ are respectively the mean value and the standard deviation of each image slice. This normalization process was also applied to the T2w images of center B.

Given the small size of the prostate gland compared to the whole image size, the dataset was unbalanced. To overcome this issue, a class-weighting process was performed. It consists of giving a different weight to both the most represented class and the least represented class. The aim is to give in the training phase a higher weight to the classification errors of the smallest class and at the same time reduce the weight of the elements of the largest class. This modification of the weights will affect the training process because the network learning is focused more on one class than the other.

For the construction of the model Keras and Tensorflow libraries [16] were used. The network was trained for 40 epochs with a batch size of 4 and using Adam optimizer with β1=0.9 and β2=0.999 and a learning rate of 0.001. Data augmentation in the training set was not performed to avoid image deformations, consequently losing the anatomical information about the prostate.

C. Loss Function

The loss function used to train this network is a combination of the Dice similarity coefficient-based loss function and the Focal loss function.

$$Loss\ Function = Dice\ Loss + Binary\ Focal\ Loss \quad (2)$$

Where Dice Loss and Focal Loss are the following:

$$DiceLoss = 1 - Dice\ similarity\ coefficient \quad (3)$$

$$Focal\ Loss\ (p_t) = -\alpha_t(1 - p_t)^{\gamma} \log(p_t) \quad (4)$$

In particular, the Dice Loss is usually used for segmentation problems [17], since it evaluates overlap between two segmentations, regardless imbalance between the two classes. Focal loss function [18] is an improvement of the basic cross-entropy function -log(p_t) where p_t is the probability score of the model for the class t. The Focal loss adds a factor to help the model handle misclassified errors and in general hard examples. It focuses training on hard negatives (samples misclassified as positive) and hard positives (samples misclassified as negative) reduces the priority on easy examples (samples correctly classified). In particular the higher γ, the lower is the loss function for correctly classified examples. Instead α adds different weights to the classes to change the importance between positive and negative examples.

D. Evaluation Metrics

Dice Similarity Coefficient (DSC), Precision (Pr) and Recall (Re) were used to evaluate the performance of the algorithm on prostate segmentation. The Dice similarity coefficient relates the elements in common between two groups with respect to the total number of elements. Precision is the portion of elements indicated as positive by the model that are truly positive. Recall is the portion of the truly positive elements that have been correctly identified by the model.

$$DSC = \frac{2TP}{2TP+FP+FN} \quad (5)$$

$$Pr = \frac{TP}{TP+FP} \quad (6)$$

$$Re = \frac{TP}{TP+FN} \quad (7)$$

Where TP, FP and FN are respectively the total number of true positives, false positives, and false negatives.

These metrics were applied both on the entire prostate gland and on different prostate regions (base, middle and apex). These regions were automatically located as following: the first two slices segmented by the algorithm were considered as the base of the prostate, the last two as the apex and the remaining ones as the middle of the gland.

III. Results

The composition of the training and the validation sets is shown in Fig.1.

	Training set		Internal Validation set (Center A)		External Validation set (Center B)
	No endorectal coil	Endorectal coil	No endorectal coil	Endorectal coil	No endorectal coil
DSC	0.91	0.95	0.90	0.89	0.87
Precision	0.93	0.96	0.95	0.92	0.89
Recall	0.91	0.94	0.89	0.87	0.86

Table 1. Mean values of Dice Similarity Coefficient (DSC), Precision and Recall for the prostate segmentation in the training set and internal and external validation sets.

The algorithm achieves DSC of 0.89 on average considering both internal and external validation sets. Differences between internal and external validation sets (Table 1) are not significantly different. In all cases Precision is higher than Recall, meaning that the net tends to include a lower number of false positive than false negative voxels. However, differences are not significant and precision is higher than 0.89 in all datasets.

When we considered the metrics obtained on the different prostate regions, we noticed a drop in performance in the apex. The mean DSC for the T2w images without endorectal coil from internal validation set is 0.71 with a very high standard deviation of 0.18. Similarly, for T2w with endorectal coil the mean DSC in the apex is 0.79 with standard deviation of 0.09. This reduction in performance near the apex also affects the external validation of Center B which shows a mean DSC of 0.73 with a standard deviation of 0.14.

	Base (Dsc ± std)	Middle (Dsc ± std)	Apex (Dsc ± std)
Internal validation set No endorectal coil (center A)	0.78 ± 0.13	0.94 ± 0.14	0.71 ± 0.18
Internal validation set endorectal coil (center A)	0.86 ± 0.08	0.92 ± 0.06	0.79 ± 0.09
External validation set No endorectal coil (center B)	0.82 ± 0.08	0.91 ± 0.05	0.73 ± 0.14

Table 2. Mean values of Dice Similarity Coefficient (DSC) and Standard deviation (std) for the prostate segmentation in the training and validation set.

Fig. 3 shows three examples of automatic prostate segmentation on three different patients, one for each validation set. A slice is shown near the base, one in the middle and one in correspondence with the apex.

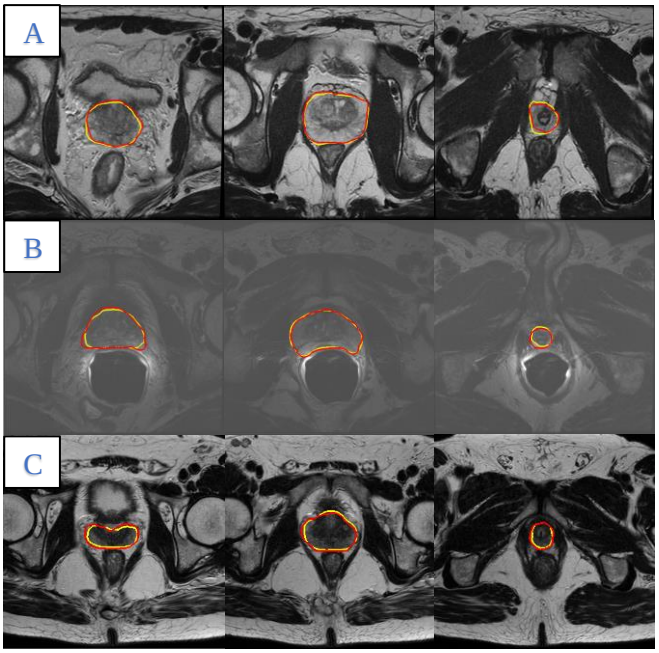

Figure 2. Qualitative comparison between automatic segmentation (yellow line) and manual segmentation (red line) of the prostate. Three sample slices corresponding to the base, middle and apex of the prostate are shown for each data set evaluated. Row A shows T2w images of Centre A without endorectal coil, row B shows T2w images of Centre A with endorectal coil, row C shows T2w images without endorectal coil from Center B.

The entire automatic algorithm takes approximately 3.4s per patient to segment the whole prostate, starting from the original image, not resized or normalized yet. The average processing time of a single slice is 0.17s.

IV. Discussions and Conclusion

In this study, an automatic algorithm able to work with multi-center T2w images with and without endorectal coil was developed. This algorithm shows performances comparable to previous works but has the advantage to segment the prostate on T2w images acquired both with and without endorectal coil.

These results shows an improvement compared to Tian et al. [8] which reached a mean DSC of 0.885 using a deep CNN model on 140 patients from two open-source prostate segmentation datasets, the PROMISE12 challenge [19] and the ISBI2013 [20]. Conversely, Zhu et al. [9] built a cascaded U-net to segment the prostate and the peripheral zone using diffusion-weighted images (DWIs) and T2w, while Nie and Shen [10] solved the blurry boundaries issues building a semantic guided strategy to learn discriminative features, obtained slightly better performances on the internal validation set reaching respectively a mean DSC of 0.92 and 0.932. However, our method has the advantage of being multi-centric while these studies worked on a singular dataset, performing training and validation of the network with images from the same center.

The mean DSC value calculated on our external validation set (Center B) is comparable to those obtained by Zavala-Romero et al. [11], who developed a 3D U-net using axial, coronal and sagittal T2-w images from different MRI centers and to Liu et al. [12] who implemented a Multi-site network that focuses on the heterogeneity of prostate MRIs from three different centers. They achieved mean DSC values of 0.89 and 0.92, respectively showing higher results than ours, but without performing external validation.

Our study has also some limitations. First, the algorithm showed lower segmentation results in both the base and the apex. Indeed, the model did not clearly identify the prostate boundaries, especially in the first and last slices of the volume. This might be due to the difficulty of manually segmenting the apex, even among experienced operators. This issue was also present at the prostate base, but with less consistency given the larger size of the gland.

Future developments of this U-net segmentation algorithm will focus on the improvement of the segmentation in both base and apex of the prostate. Additional training will be performed on a larger number of patients.

In conclusion, the presented algorithm can provide an important support in the identification of the prostate area to move on to the subsequent detection of PCa, also drastically reducing the segmentation times.

ACKNOWLEDGMENT

The research leading to these results has received funding from the Fondazione AIRC under IG2017 - ID.20398 project – P.I. Regge Daniele and from the European Union's Horizon 2020 research and innovation programme under grant agreement no 952159.

REFERENCES

[1] R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal, "Cancer Statistics, 2021," *CA. Cancer J. Clin.*, vol. 71, no. 1, pp. 7–33, 2021, doi: https://doi.org/10.3322/caac.21654.

[2] Y. K. Fong and B. Djavan, "Early detection of prostate cancer.," *Rev. Urol.*, vol. 7, no. 1, pp. 63–64, 2005.

[3] M. A. Bjurlin *et al.*, "Update of the Standard Operating Procedure on the Use of Multiparametric Magnetic Resonance Imaging for the Diagnosis, Staging and Management of Prostate Cancer.," *J. Urol.*, vol. 203, no. 4, pp. 706–712, Apr. 2020, doi: 10.1097/JU.0000000000000617.

[4] N. Mottet *et al.*, "EAU-ESTRO-SIOG Guidelines on Prostate Cancer. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent.," *Eur. Urol.*, vol. 71, no. 4, pp. 618–629, Apr. 2017, doi: 10.1016/j.eururo.2016.08.003.

[5] E. Lecornet *et al.*, "The accuracy of different biopsy strategies for the detection of clinically important prostate cancer: a computer simulation.," *J. Urol.*, vol. 188, no. 3, pp. 974–980, Sep. 2012, doi: 10.1016/j.juro.2012.04.104.

[6] L. C. Brown *et al.*, "Multiparametric MRI to improve detection of prostate cancer compared with transrectal ultrasound-guided prostate biopsy alone: the PROMIS study.," *Health Technol. Assess.*, vol. 22, no. 39, pp. 1–176, Jul. 2018, doi: 10.3310/hta22390.

[7] V. Giannini *et al.*, "Multiparametric magnetic resonance imaging of the prostate with computer-aided detection: experienced observer performance study.," *Eur. Radiol.*, vol. 27, no. 10, pp. 4200–4208, Oct. 2017, doi: 10.1007/s00330-017-4805-0.

[8] Z. Tian, L. Liu, Z. Zhang, and B. Fei, "PSNet: prostate segmentation on MRI based on a convolutional neural network.," *J. Med. imaging (Bellingham, Wash.)*, vol. 5, no. 2, p. 21208, Apr. 2018, doi: 10.1117/1.JMI.5.2.021208.

[9] Y. Zhu *et al.*, "Fully automatic segmentation on prostate MR images based on cascaded fully convolution network.," *J. Magn. Reson. Imaging*, vol. 49, no. 4, pp. 1149–1156, Apr. 2019, doi: 10.1002/jmri.26337.

[10] D. Nie and D. Shen, "Semantic-guided Encoder Feature Learning for Blurry Boundary Delineation," *CoRR*, vol. abs/1906.0, 2019, [Online]. Available: http://arxiv.org/abs/1906.04306.

[11] O. Zavala-Romero *et al.*, "Segmentation of prostate and prostate zones using deep learning : A multi-MRI vendor analysis.," *Strahlenther. Onkol.*, vol. 196, no. 10, pp. 932–942, Oct. 2020, doi: 10.1007/s00066-020-01607-x.

[12] Q. Liu, Q. Dou, L. Yu, and P. A. Heng, "MS-Net: Multi-Site Network for Improving Prostate Segmentation With Heterogeneous MRI Data.," *IEEE Trans. Med. Imaging*, vol. 39, no. 9, pp. 2713–2724, Sep. 2020, doi: 10.1109/TMI.2020.2974574.

[13] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *CoRR*, vol. abs/1505.0, 2015, [Online]. Available: http://arxiv.org/abs/1505.04597.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *CoRR*, vol. abs/1512.0, 2015, [Online]. Available: http://arxiv.org/abs/1512.03385.

[15] J. Deng, W. Dong, R. Socher, L. Li, Kai Li, and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255, doi: 10.1109/CVPR.2009.5206848.

[16] F. Chollet, "Keras," *GitHub*. https://github.com/keras-team/keras.

[17] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso, "Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations," *CoRR*, vol. abs/1707.0, 2017, [Online]. Available: http://arxiv.org/abs/1707.03237.

[18] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal Loss for Dense Object Detection.," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, Feb. 2020, doi: 10.1109/TPAMI.2018.2858826.

[19] G. Litjens *et al.*, "Evaluation of prostate segmentation algorithms for MRI: the PROMISE12 challenge.," *Med. Image Anal.*, vol. 18, no. 2, pp. 359–373, Feb. 2014, doi: 10.1016/j.media.2013.12.002.

[20] N. Bloch, "Nc-isbi 2013 challenge: automated segmentation of prostate structures," *Cancer Imaging Arch.*, 2015.