CNN-based Prostate Zonal Segmentation on Magnetic Resonance Images

Changhee HAN[1,a)]   Leonardo RUNDO[2,b)]   Jin ZHANG[1,c)]   Ryuichiro HATAYA[1,d)]
Yudai NAGANO[1,e)]   Giancarlo MAURI[2,f)]   Hideki NAKAYAMA[1,g)]

Abstract

Although prostate cancer is the most common cancer in American males, prostate Magnetic Resonance Imaging (MRI) remains challenging. Besides whole prostate gland segmentation, differentiating between the unclear boundary of the Central Gland and Peripheral Zone can lead to differential diagnosis, since the frequency and severity of tumors differ in these regions. To tackle this issue, we compare three Convolutional Neural Network (CNN)-based architectures: SegNet, U-Net, and pix2pix. Moreover, this study evaluates CNN's generalization ability on two multi-centric MRI datasets using a mixed scheme: (i) training on either each dataset or both datasets and (ii) testing on both datasets. The results show that training on multi-centric datasets generally outperforms training on each individual dataset during testing, allowing for both intra-/inter-dataset generalization—this is valuble in medical imaging, as clinical applications involve such multi-centric settings. In general, U-Net outperforms the other methods, especially when testing is performed on samples of the datasets used during training.

1. Introduction

Prostate cancer (PCa) is expected to be the most common cancer among American males during 2018 [1]. Several imaging modalities can be used for PCa diagnosis—such as Transrectal Ultrasound (TRUS), Computed Tomography (CT), and Magnetic Resonance Imaging (MRI)—according to the clinical context. As a matter of fact, conventional structural T1-weighted (T1w) and T2-weighted (T2w) MRI sequences can be combined with the functional information conveyed by Dynamic Contrast Enhanced MRI (DCE-MRI), Diffusion Weighted Imaging (DWI), and Magnetic Resonance Spectroscopic Imaging (MRSI) [3]. Therefore, MRI plays a decisive role in PCa diagnosis, revealing the internal prostatic anatomy, prostatic margins, and extent PCa [2]. The prostate Whole Gland (WG) is subdivided into the Central Gland (CG) and Peripheral Zone (PZ). In prostate imaging, T2w MRI serves as the principal sequence, thanks to its high resolution that enables to differentiate the hyper-intense PZ and hypo-intense CG in young male subjects.

Besides manual detection and delineation of the WG and PCa on MR images, distinguishing between the CG and PZ is clinically essential, since the frequency and severity of tumors differ in these regions [4]; the PZ harbors $70-80\%$ of PCa and represents a target for prostate biopsy. Therefore, radiologists must conduct zonal partition first to assess PCa patients' multi-parametric MRI. However, better PCa diagnosis requires a reliable and automatic zonal segmentation method, since such manual delineation is time-consuming and operator-dependent [9]. Moreover, in clinical practice, generalization abilities—between multi-centric prostate MRI datasets—is essential due to large anatomical inter-subject variability.

So, how can we extract the CG and PZ from the WG on different MRI datasets? In this work, we automatically segment the prostate zones on two multi-centric T2w MRI datasets to evaluate the generalization ability of Convolutional Neural Network (CNN)-based architectures: SegNet [6], U-Net [7], and pix2pix [8]. However, this is challenging since multi-centric datasets are characterized by different contrast, visual consistency, and heterogeneous

---

[1]   Graduate School of Information Science and Technology, The University of Tokyo, Tokyo, Japan
[2]   Department of Informatics, Systems and Communication, University of Milano-Bicocca, Milan, Italy
a)   han@nlab.ci.i.u-tokyo.ac.jp
b)   leonardo.rundo@disco.unimib.it
c)   zhangjin@nlab.ci.i.u-tokyo.ac.jp
d)   hataya@nlab.ci.i.u-tokyo.ac.jp
e)   nagano@nlab.ci.i.u-tokyo.ac.jp
f)   mauri@disco.unimib.it
g)   nakayama@ci.i.u-tokyo.ac.jp

image characteristics. Therefore, this study adopted a mixed scheme by (*i*) training on either each dataset or both datasets and (*ii*) testing on both datasets. In such a context, we compared the segmentation performances achieved with/without pre-training.

Contributions. Our main contributions are:

- **Prostate Zonal Segmentation:** This reseach shows that U-Net achieves accurate CG and PZ segmentation results.

- **Cross-dataset Generalization:** This first prostate cross-dataset study reveals that CNN training on multi-centric datasets generally outperforms training on each dataset during testing.

2. Materials and Methods

Towards better diagnosis in a clinical routine, we segment CG and PZ from the WG using three different CNN-based architectures in 4-fold cross-validation.

2.1 MRI Datasets

This paper exploits two multi-parametric prostate MR datasets:

- Cannizzaro Hospital (Catania, Italy) dataset with 21 patients/193 images (acquired using a 3.0 Tesla Philips scanner) [9];

- Initiative for Collaborative Computer Vision Benchmarking (I2CVB) dataset with 19 patients/503 images (acquired using a 3.0 Tesla Siemens scanner) [3].

For clinical feasibility, we analyzed only T2w images—the most commonly used sequence for prostate zonal segmentation. Fig. 1 shows two example T2w MR images of the analyzed two datasets.

We conducted three experiments in 4-fold cross-validation to confirm the generalization effect under different training/testing conditions:

- individual dataset (Cannizzaro), training on Cannizzaro (16 patients) alone, testing on Cannizzaro (5 patients) and I2CVB (19 patients) separately;

- individual dataset (I2CVB), training on I2CVB (15 patients) alone, testing on Cannizzaro (21 patients) and I2CVB (4 patients) separately;

- mixed dataset, training on Cannizzaro (16 patients) and I2CVB (15 patients) together, testing on Cannizzaro (5 patients) and I2CVB (4 patients) separately.

2.2 Pre-processing

To fit the image resolution of the dataset #1, we center-cropped the images of the dataset #2 and resized them to

| (a) | (b) |
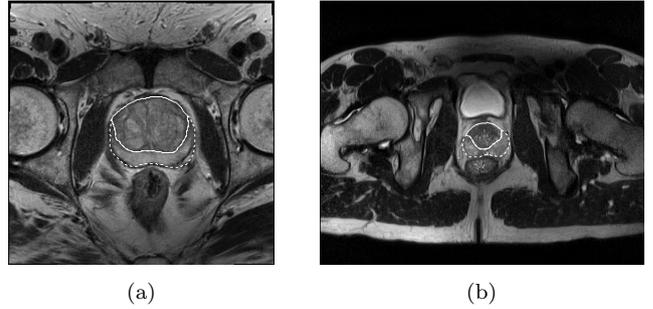
Fig. 1 Example input prostate T2w MR axial slices in their original image ratio: (a) dataset #1; (b) dataset #2. The CG and PZ are highlighted with solid and dashed white lines, respectively.

288×288 pixels. For better training, we randomly cropped the input images from 288×288 to 256×256 pixels and horizontally flipped them.

2.3 Post-processing

Two efficient morphological operations were applied on the obtained CG binary masks to smooth boundaries and avoid disconnected regions:

- a hole filling algorithm on the segmented $\mathcal{R}_{CG}$ to remove possible holes in a predicted map;

- a small area removal operation dealing with connected components smaller than $\lfloor |\mathcal{R}_{WG}|/8 \rfloor$ pixels, where $|\mathcal{R}_{WG}|$ denotes the number of pixels contained in WG segmentation.

2.4 Prostate Zonal Segmentation

This work adopts a selective two-step delineation approach to focus on pathological regions in the CG and PZ denoted with $\mathcal{R}_{CG}$ and $\mathcal{R}_{PZ}$, respectively. Relying on [10], the PZ was obtained by subtracting the CG from the WG ($\mathcal{R}_{WG}$) meeting the constraints: $\mathcal{R}_{WG} = \mathcal{R}_{CG} \cup \mathcal{R}_{PZ}$ and $\mathcal{R}_{CG} \cap \mathcal{R}_{PZ} = \emptyset$.

All the investigated CNN-based architectures were trained using the $\mathcal{L}_{DSC}$ loss function through the classifying pixels N:

$$\mathcal{L}_{DSC} = -\frac{2\sum_{i=1}^{N} s_i \cdot r_i}{\sum_{i=1}^{N} s_i + \sum_{i=1}^{N} r_i}, \tag{1}$$

where s_i and r_i refer to the continuous values in $[0,1]$ of the prediction map and ground truth at i-th pixel, respectively.

2.4.1 SegNet

SegNet[6] is a CNN architecture for semantic pixel-wise segmentation. During training, we used the Stochastic Gradient Descent (SGD) with a learning rate of 0.01, momentum of 0.9, weight decay of 5×10^{-4}, and batch size

of 8. It was trained for 50 epochs and the learning rate was multiplied by 0.2 at the 20-th and 40-th epochs.

2.4.2 U-Net

U-Net [7] is a fully CNN capable of stable training with reduced samples, combining encoders-decoders with skip connections between them. Using four scaling operations, it wwas implemented. We used SGD with a learning rate of 0.01, momentum of 0.9, weight decay of 5×10^{-4}, and batch size of 4. Training was executed for 50 epochs, multiplying the learning rate by 0.2 at the 20-th and 40-th epochs.

2.4.3 pix2pix

pix2pix [8]—an image-to-image translation method with conditional adversarial networks—was used to translate the original image into the segmented one. The generator (U-Net) and discriminator include eight and five scaling operations, respectively. Adam ? was used as an optimizer with a learning rate of 0.01 for the generator—which was multiplied by 0.1 every 20 epochs—and 0.0002 for the discriminator. It was trained for 50 epochs with a batch size of 12.

3. Results

We quantitatively evaluated the segmentation results $\mathcal{S}$ against the gold standard mannual segmentation $\mathcal{G}$, using DSC:

$$DSC = 2 \times \frac{|\mathcal{S} \cap \mathcal{G}|}{|\mathcal{S}| + |\mathcal{G}|} \times 100(\%). \qquad (2)$$

Table 1 shows the 4-fold cross-validation results under different training/testing conditions. U-Net generally outperforms the other methods on both CG and PZ segmentation, thanks to its good generalization ability when testing is performed on samples of the datasets used during training; however, it experiences problems when trained on dataset #1 and tested on dataset #2, due to the lack of training images. In such a case, pix2pix generalizes better thanks to the internal generative model. Moreover, SegNet reveals rather unstable results, especially when trained on a limited amount of data.

The results also show that training on multi-centric datasets generally outperforms training on each individual dataset during testing, allowing for both intra-/inter-dataset generalization; therefore, we may train CNNs on multiple datasets in medical imaging, as clinical applications involve such multi-centric settings. Comparing the CG and PZ segmentation, when tested on dataset #1, the results on the PZ are generally more accurate, while the tendency is opposite for the dataset #2;

this is probably due to their diffrent MRI scanners.

For a visual assessment, two examples (one for each dataset) are shown in Fig. 2. It can be seen that U-Net generally achieves more accurate results compared with SegNet and pix2pix, confirming the trend revealed by the *DSC* values in Table 1.

4. Conclusion and Future Work

Our results show that CNN-based architectures, especially U-Net, can achieve accurate prostate zonal segmentation on two different datasets, leading to valuable clinical insights; CNNs suffer when trained/tested on different MRI datasets with different devices/protocols, but it can be mitigated by training the CNNs on multiple datasets, generalizing excellent and robust.

As future developments, we may further improve the results by refining the predicted binary masks for better smoothness and continuity, avoiding disconnected segments; furthermore, we should enhance the output delineations considering the 3D spatial information among slices. Finally, for better cross-dataset generalization, additional prostate zonal datasets and domain adaptation *via* transfer learning by maximizing distribution similarity; in this context, Generative Adversarial Networks and Variational Auto-Encoders could be useful.

References

[1] R. L., Siegel, K. D., Miller, A., Jemal, Cancer statistics, 2018. CA Cancer J. Clin., 68(1), 7–30, 2018.

[2] G. M., Villeirs, G. O., De Meerleer, Magnetic resonance imaging (MRI) anatomy of the prostate and application of MRI in radiotherapy planning, Eur. J. Radiol., 63(3), 361–368, 2007.

[3] G., Lemaître, R., Martí, J., Freixenet, et al., Computer-aided detection and diagnosis for prostate cancer based on mono and multi-parametric MRI: A review, Comput. Biol. Med., 60, 8–31, 2015.

[4] E., Niaf, O., Rouvière, F., Mège-Lechevallier, et al., Computer-aided diagnosis of prostate cancer in the peripheral zone using multiparametric MRI, Phys. Med Biol., 57(12), 3833, 2012.

[5] T., Clark, J., Zhang, S., Baig, et al., Fully automated segmentation of prostate whole gland and transition zone in diffusion-weighted MRI using convolutional neural networks, J. Med. Imaging, 4(4), 2017. 041307

[6] V., Badrinarayanan, A., Kendall, R., Cipolla, Segnet: A deep convolutional encoder-decoder architecture for image segmentation, IEEE Trans. Pattern Anal. Mach Intell., 39(12), 2481–2495, 2017.

[7] O., Ronneberger, P., Fischer, T., Brox, U-Net: Convolutional networks for biomedical image segmentation, in Medical Image Computing and Computer-Assisted Intervention (MICCAI), LNCS Vol. 9351, pp. 234–241, Springer, 2015.

[8] P., Isola, J. Y., Zhu, T., Zhou, A. A., Efros, Image-to-

Table. 1 Prostate zonal segmentation results of the three CNN-based architectures in 4-fold cross-validation assessed by *DSC* (presented as the average and standard deviation). The experimental results are calculated on the different conditions of (*i*) training on either each dataset or both datasets and (*ii*) testing on both datasets. Numbers in bold indicate the highest *DSC* values for each prostate region (i.e., CG and PZ) among all the architectures.

Training	Network architecture	Region	Testing on Cannizzaro		Testing on I2CVB	
			Average	*Std. Dev.*	*Average*	*Std. Dev.*
Cannizzaro	SegNet	CG	80.20	3.28	74.48	5.82
		PZ	80.66	11.51	59.57	12.68
	U-Net	CG	**84.33**	2.37	74.18	3.77
		PZ	**88.98**	2.98	66.63	1.93
	pix2pix	CG	82.35	2.09	**76.61**	2.17
		PZ	87.09	2.72	**73.20**	2.62
I2CVB	SegNet	CG	76.04	2.05	87.07	2.41
		PZ	**77.25**	3.09	82.45	1.77
	U-Net	CG	**78.88**	0.88	**88.21**	2.10
		PZ	74.52	1.85	**83.03**	2.46
	pix2pix	CG	77.90	0.73	86.95	2.93
		PZ	66.09	3.07	81.33	0.90
Mixed	SegNet	CG	84.28	3.12	87.92	2.80
		PZ	87.74	1.66	82.21	0.79
	U-Net	CG	**86.34**	2.10	**88.12**	2.34
		PZ	**90.74**	2.40	**83.04**	2.30
	pix2pix	CG	83.07	3.39	86.39	3.16
		PZ	83.53	2.36	80.40	1.80

| (a) | (b) | (c) | (d) |

| (e) | (f) | (g) | (h) |
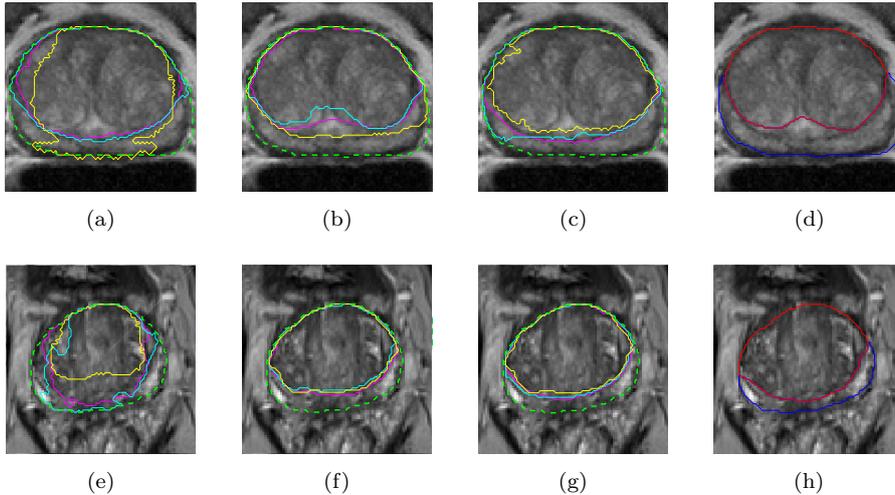
Fig. 2 Examples of prostate zonal segmentation. The first row concerns testing on dataset #1, trained on: (a) dataset #1; (b) dataset #2; (c) mixed dataset. The second row concerns testing on dataset #2, trained on: (e) dataset #1; (f) dataset #2; (g) mixed dataset. The $\mathcal{R}_{CG}$ segmentation results are represented with magenta, cyan, and yellow solid contours for SegNet, U-Net, and pix2pix, respectively. The dashed green line denotes the $\mathcal{R}_{WG}$ boundary. The last column (sub-figures (d) and (h)) shows the gold standard for $\mathcal{R}_{CG}$ and $\mathcal{R}_{PZ}$ with red and blue lines, respectively. The images are zoomed with a $4\times$ factor.

image translation with conditional adversarial networks, arXiv preprint arXiv:1611.07004, 2016.

[9] L., Rundo, C., Militello, G., Russo, G., et al., Automated prostate gland segmentation based on an unsupervised fuzzy c-means clustering technique using multispectral T1w and T2w MR imaging, Information, 8(2), 49, 2017.

[10] W., Qiu, J., Yuan, E., Ukwatta, et al. Dual optimization based prostate zonal segmentation in 3D MR images,

Med. Image Anal., 18(4), 660–673, 2014.

[11] K., Simonyan, A., Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556, 2015.