

A smart cropping pipeline to improve prostate's peripheral zone segmentation on MRI using Deep Learning

Dimitris Zaridis^{1,2}, Eugenia Mylona^{1,2}, Nikolaos Tachos^{1,2}, Kostas Marias³, Manolis Tsiknakis^{3,4}, Dimitrios I. Fotiadis^{1,2,*}

¹Dept. of Biomedical Research, FORTH-IMBB, Ioannina, Greece

²Unit of Medical Technology and Intelligent Information Systems, Dept. of Materials Science and Engineering, University of Ioannina, Ioannina, Greece

³Institute of Computer Science, FORTH, Heraklion, Greece

⁴Department of Electrical and Computer Engineering, Hellenic Mediterranean University, Heraklion, Crete, Greece

Abstract

INTRODUCTION: Although accurate segmentation of the prostatic subregions is a crucial step for prostate cancer diagnosis, it remains a challenge.

OBJECTIVES: To propose a deep learning (DL)-based cropping pipeline to improve the performance of DL networks for segmenting the prostate's peripheral zone.

METHODS: A U-net network was trained to crop the area around the peripheral zone on MRI in order to reduce the class imbalance between foreground and background pixels. The DL-cropping was compared with the standard center-cropping using three segmentation networks.

RESULTS: The DL-cropping improved significantly the segmentation performance in terms of Dice score, Sensitivity, Hausdorff Distance, and Average Surface Distance, for all three networks. The improvement in Dice Score was 34%, 13% and 16% for the U-net, Dense U-net and Bridged U-net, respectively.

CONCLUSION: For all the evaluated networks, the proposed DL-cropping technique outperformed the standard center-cropping.

Received on 08 December 2021; accepted on 21 February 2022; published on 24 February 2022

Keywords: deep learning, MRI segmentation, prostate cancer, peripheral zone, cropping

Copyright © 2022 Dimitris Zaridis *et al.*, licensed to EAI. This is an open access article distributed under the terms of the Creative Commons Attribution license, which permits unlimited use, distribution and reproduction in any medium so long as the original work is properly cited.

doi:10.4108/eai.24-2-2022.173546

1. Introduction

Prostate cancer is one of the most malignant tumors and the second cause of cancer-related death in males. Nevertheless, early detection and staging of the disease is associated with nearly 100% 5-year survival rate [1]. One of the critical steps for accurate prostate cancer diagnosis and efficient treatment is the precise delineation of the prostate gland and its sub-divisions. Today, the T2-weighted (T2w) magnetic resonance imaging (MRI) is considered the state-of-the-art imaging modality for prostate segmentation as it

provides superior resolution and contrast in soft tissues compared to other imaging techniques [2].

The prostate gland is subdivided into zones, namely the peripheral zone (PZ), the transitional zone (TZ), and the central zone (CZ), with tumor properties differing significantly depending on which zone they are found [3]. Since most commonly prostatic tumors are identified in the PZ (70%–80%) [4], it is imperative to segment this area with high accuracy.

In clinical setting, the contouring of the prostate and prostatic zones is performed manually. However this requires a significant amount of time and is highly dependent on the experience and expertise of the practitioner. Automating PZ delineation may allow overcoming the shortcomings of manual annotation.

*Corresponding author. Email: fotiadis@uoi.gr

Nonetheless, the wide range of prostate shape variation among patients and the heterogeneous pixel representation surrounding the PZ boundary, renders the automatic PZ segmentation a daunting task [5].

Over the past decade, the advances in Deep Learning, and particularly in convolutional neural network (CNN) based concepts, have significantly improved the performance of automatic prostate segmentation. In the field of medical imaging in general, a plethora of deep learning (DL) architectures have been proposed, with the original U-net [6] being a remarkable achievement that consists today the backbone of several more sophisticated models. Despite recent advances in DL-based segmentation methods, the performance of existing models for prostate, and particularly for PZ segmentation, is not considered sufficient to enable their transfer and deployment in clinical practice.

Novel image preprocessing approaches are commonly applied to boost networks' performance and to enhance segmentation accuracy. A common but critical issue that may hamper model's performance is the presence of imbalanced data that characterize the image labels [7]. Specifically in medical imaging, class imbalance refers to the situation where the number of pixels in the region of interest (ROI) is significantly smaller than the number of pixels in the background. The presence of class imbalance in the dataset used for model training might result to inconsistent segmentation networks. A workaround to overcome this problem is to crop the image around the ROI before training the DL network. To automate this process, conventional center cropping is commonly used under the premise that the ROI is located at image center [8]. While this practice is best suited for large and consistent ROIs, for PZ segmentation this risks producing faulty segmentation results [9].

In this paper we propose a smart DL-cropping pipeline that permits cropping around the PZ of the prostate on T2w MR images with the objective to improve DL models' segmentation accuracy. The efficacy of DL-cropping for improving segmentation performance was assessed on three state-of-the-art segmentation networks and was compared to the conventional center-cropping approach.

2. Materials and Methods

2.1. Dataset

For the purpose of this study, the publicly available Cancer Imaging Archive (TCIA) PROSTATEx dataset was used [10]. This includes T2w MRI images of 98 patients along with their annotations, manually delineated by experts. The main MRI vendors were Siemens (MAGNETOM Trio and Skyra models) with a magnetic field 3 Tesla. In total, the number of annotated 2D frames on the PZ is 1319, the slice thickness was 3.6mm and the number of slices ranged from 15-22.

The frames were 384X384 pixels in size before being resized to 256X256 to match the models' specifications. To increase model variability and generalizability, data augmentation was used to apply a set of affine transformations to the original image, including (i) image rotation in varying predetermined degrees (-20, -10, -5, 5, 10, 20), and (ii) image shifting in any direction by a factor of 0.5.

2.2. DL cropping pipeline

In this paper, we suggest a deep learning cropping strategy for reducing the pixels' class imbalance between the PZ of the prostate and the background pixels in the frame. This can be considered as a pre-processing step for PZ segmentation where a bounding box enclosing the ROI is created on each frame. The ROI box is then extended by 40 pixels either horizontally as well as vertically from the original mask, and the training frames are clipped around the bounding box region [12]. A U-net network [6] was trained to trim the region surrounding the PZ in the testing dataset for this purpose, resulting in a more equal mixture of foreground and background pixels. Fig. 1 shows the pipeline that was used to define the bounding box. A sample from the center-cropped images and the related annotation are presented in steps (i) & (ii). These frames are sent into the U-net model (step (iii)), which was previously trained to detect the region surrounding the PZ using the larger bounding boxes. The predictions are taken from the initial frames in step (iv), approximately defining the region of interest. Furthermore, the bounding boxes are constructed in step (v) utilizing the amorphous masks' minimum and maximum coordinates on the x and y axes from step (iv). The original annotations are always included in the final cropped image thanks to this approach. In step (vi), the frames are resampled to 256X256 pixels in order for the input frames to meet the network's criteria. Finally, the cropped frames and annotations are displayed in steps (vii) and (viii), and the data resulted from the previous stages may be utilized to train the networks.

2.3. Deep Learning segmentation networks

For the comparison of the DL-cropping technique to the traditional center-cropping method, the original versions of three state-of-the-art segmentation networks were used. The first is the U-net model [6], which employs an encoder-decoder layer combination, with the layers coupled in serial and parallel to increase the network's capacity to learn spatial features. Furthermore, Dense U-net [13] is an encoder-decoder network in which dense blocks [14] are used to propagate information from previous layers forward, while transitional blocks flatten the feature maps and keep

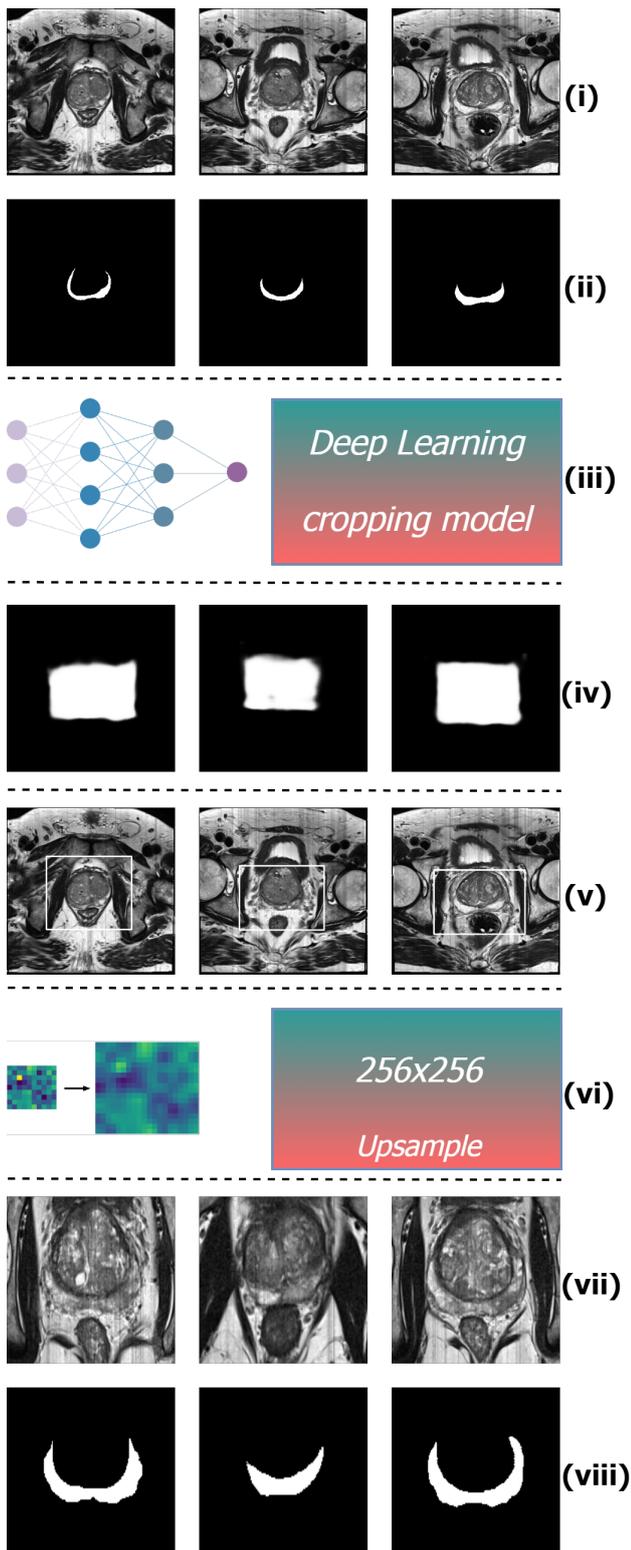


Figure 1. Cropping pipeline based on DL approach.

the most important features, lowering the network's computational cost. The third network, Bridged U-net [11], is comprised of two inter-connected U-nets, with

cross linkages across levels from the first U-net to the second, enabling the models to interact and collaborate further on feature extraction.

2.4. Network training

Two pipelines were used to train and test the network designs. The networks were trained using the frames after conventional center-cropping in the first (Fig.1, steps i, ii) while the suggested DL cropping technique was utilized to crop the frames into variable size slices ranging from 90 to 140 pixels for the second pipeline, and the generated images (Fig.1, steps vi, vii) were used to train the networks.

Regarding the training parameters in the DL models, the cost functions employed were the training accuracy and binary cross-entropy loss. The optimization method used was the Adam algorithm [16] instead of the Stochastic gradient descent [17] since it has been shown that the former converges faster. For all architectures, the model was trained for 120 epochs. To decrease calculation time, a checkpoint technique and early stopping were utilized, as well as tensorflow's tensorboard for monitoring the training and validation process.

Training was performed in 5-fold cross validation pipeline to properly evaluate the generalizability of the models. The partition of the images within each fold was done patient-wised to ensure unbiased model training by preserving the intra-individual size and shape variations of prostate's PZ. For all models and both DL-crop and center-crop techniques, the patients were segregated in the folds in identical way. A total of 78 patients were used in each training and validation set, with the remaining 20 patients being used for testing. In terms of frames' distribution (2D slices), each training fold had 891 slices prior data augmentation and 1692 slices post data augmentation, with an image size of 256x256 pixels. There were roughly 152 slices in each validation set and 276 slices in each testing fold.

2.5. Performance evaluation

Several metrics were used to assess the segmentation performance, including the Dice Score coefficient, the Balanced Accuracy, the Hausdorff distance, the Average Surface Distance, the Rand Error index, the Sensitivity and the Specificity [18]. The performance of the trained models was calculated by averaging the 5-fold cross validation results over the test-sets. The non-parametric Wilcoxon matched-pairs signed rank test was used to compare the center-crop and DL-crop for each metric and architecture.

3. Results

The performance of the three segmentation algorithms was evaluated for both the proposed DL-cropping and

the standard center-cropping using seven performance metrics. The average values and standard deviation are shown in Table 1. Comparing the scores in Table 1, the DL-cropping resulted to better performance for all the metrics, except specificity. Besides, the proposed pipeline has the lowest standard deviation values for all the metrics, except specificity. It is worth mentioning that the increased specificity achieved with center-cropping is directly associated with the higher class imbalance in the resulting images in favor of the background pixels leading to a considerably larger number of true negative predictions.

Overall, the proposed pipeline outperformed the center-cropping method with significant improvements for all architectures ($p < 0.001$) in terms of Dice score, Sensitivity, Hausdorff Distance and Average Surface Distance. The improvement in mean Dice Score was 34%, 13% and 16% for the U-net, Dense U-net and Bridged U-net, respectively. Regarding the Hausdorff Distance, there was an improvement of 53%, 76% and 39% for U-net, Dense U-net and Bridged U-net, respectively. The corresponding boxplots of the seven metrics for the three segmentation networks using center-cropping and DL-cropping are shown in Fig. 2.

In Fig. 3 an indicative example of the segmentation performance for the three DL networks after DL-cropping and center-cropping is provided. The blue contours originate from the original PZ mask, while the predicted mask is depicted with orange. The qualitative assessment through visual inspection of the segmented regions using the three networks, also confirmed that image preprocessing with DL cropping instead of center cropping, improved the performance of the algorithms.

Metric	Cropping technique	Segmentation Networks		
		Bridged Unet	Dense Unet	Unet
Dice Score	center-crop	0.51±0.31	0.53±0.28	0.41±0.28
	DL-crop	0.60±0.25	0.61±0.26	0.58±0.25
	<i>p-value</i>	<0.001	<0.001	<0.001
Balanced accuracy	center-crop	0.78±0.17	0.80±0.16	0.71±0.16
	DL-crop	0.81±0.14	0.83±0.14	0.80±0.13
	<i>p-value</i>	<0.405	0.012	0.001
Sensitivity	center-crop	0.61±0.32	0.56±0.34	0.43±0.32
	DL-crop	0.64±0.28	0.68±0.27	0.64±0.27
	<i>p-value</i>	<0.001	0.001	<0.001
Specificity	center-crop	0.99±0.006	0.99±0.004	0.99±0.004
	DL-crop	0.97±0.02	0.98±0.02	0.97±0.02
	<i>p-value</i>	1	1	1
Rand error index	center-crop	0.27±0.26	0.30±0.29	0.34±0.30
	DL-crop	0.22±0.17	0.22±0.19	0.24±0.18
	<i>p-value</i>	<0.001	0.15	<0.001
Hausdorff Distance (mm)	center-crop	9.39±7.57	14.35±13.45	11.60±7.97
	DL-crop	6.45±5.15	6.35±4.40	6.75±4.78
	<i>p-value</i>	<0.001	<0.001	<0.001
Average surface distance (mm)	center-crop	2.72±3.45	3.89±5.60	3.51±3.52
	DL-crop	1.80±1.84	1.79±2.34	1.98±2.03
	<i>p-value</i>	<0.001	<0.001	<0.001

Table 1. For three DL models, Performance of segmentation using center-crop and DL-crop

4. Discussion

The current study presents a novel preprocessing technique for increasing the efficiency and effectiveness of established PZ segmentation DL architectures on T2w MR images. To address the problem of class imbalance between background and foreground pixels in the image, a DL-based framework for image cropping is presented. As it was shown, the suggested DL cropping approach outperformed the traditional center cropping for all of the prostate segmentation networks considered in this work.

The prostate gland, particularly the PZ, consist just a small component of the typical pelvic MRI. At the same time, it is well-documented that when machine- and deep-learning algorithms are trained on unbalanced data, may suffer from restricted prediction accuracy [19]. In the case of imbalanced representation between two classes in the training image (i.e. foreground and background pixels), then the most frequently occurring class will be favored during training. A potential solution is to tackle this issue during model training by opting for a loss functions able to compensate for the presence of class imbalance. With the weighted cross-entropy loss, for example, class weights inversely related to the incidence of each class are assigned, thereby penalizing the most frequently occurring class. Nonetheless, the choose of the most efficient weighting function is cumbersome and application-specific [20]. In [21], the authors compared the effects on PZ segmentation of some of the most popular loss functions.

Some authors have leveraged object detection approaches to improve segmentation performance in various medical imaging applications. Jaeger et al. [22], proposed the Retina Unet, which combines the RetinaNet's one-stage detector with the standard U-Net architecture. Training both segmentation and detection tasks simultaneously was able to improve detection rate, compared to U-net-like models. Based on the 3D Retina Unet, recently, a fully automatic DL-based model was proposed to perform at the same time prostate cancer detection, segmentation and Gleason Grade estimation, achieving a state-of-the-art performance level [23]. Nevertheless, the original U-Net architecture has 30 million parameters and the Resnet model, which is the backbone of RetinaNet, has 25 million parameters. Conversely, our proposed smart-crop U-Net, with only 1.94 million parameters, maintains the same performance to crop efficiently the area of interest. Apart from two-stage detectors, some works have also implemented one-stage detectors to directly localize ROIs without requiring candidate regions to be proposed [24]. These detectors have shown to be more flexible, straightforward, and computationally efficient. Additionally, novel object detection approaches, such

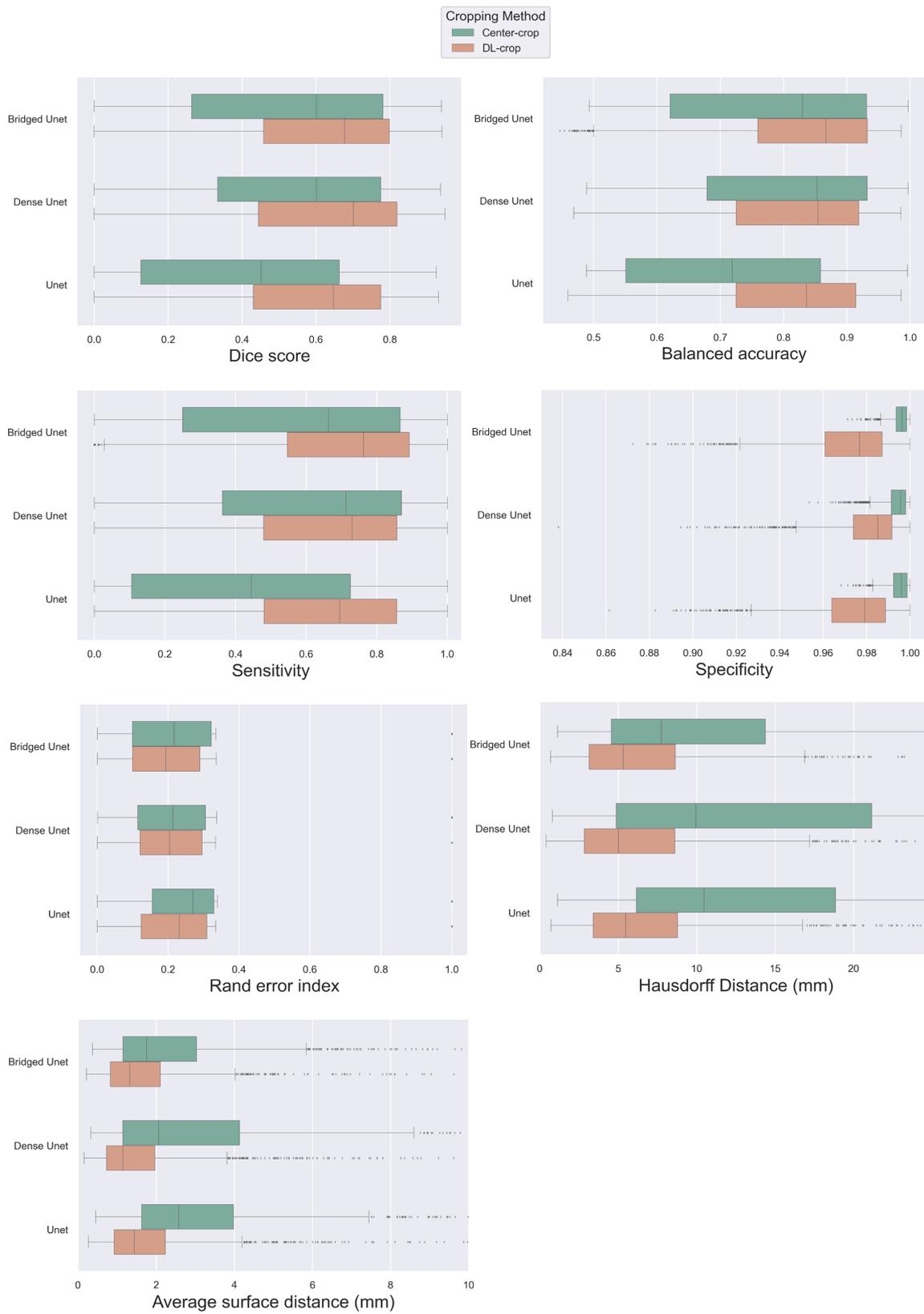


Figure 2. Boxplots of the PZ segmentation performance for center-cropping and DL-cropping using three networks.

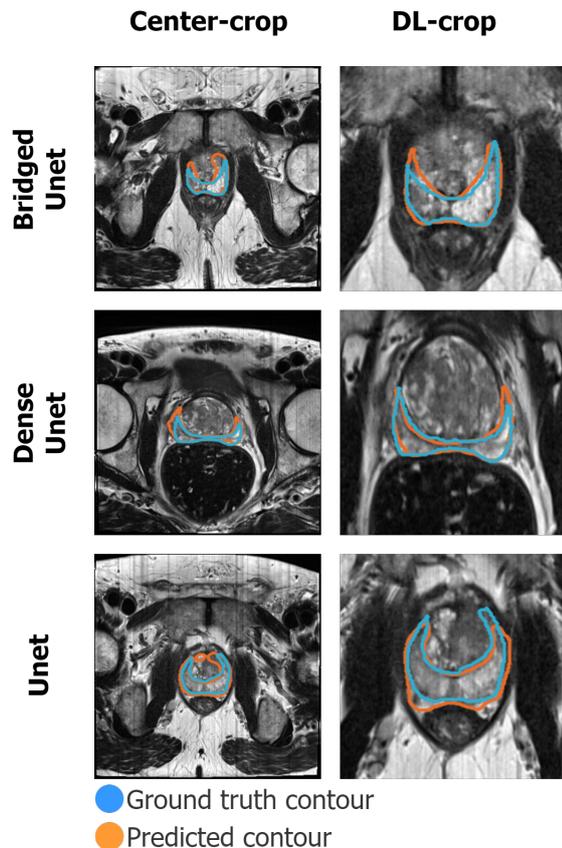


Figure 3. PZ outlines for three segmentation networks using center-cropping (left) and DL-cropping (right) approaches.

as the CaraNet [25], are particularly attractive for the segmentation of small objects in medical images with some recent works, demonstrating promising results. In future studies, it would be of particular interest to compare emerging object detection methods with the proposed DL-based smart cropping for addressing challenging task of prostate cancer segmentation.

The present work has some limitations. First, we have not been able to reach the performance scores reported in the literature for the different DL segmentation networks, possibly due to the lower number of patients included in our study. For instance, using a training sample of 141 patients for PZ segmentation, the U-net and Dense U-net models have achieved 75% and 78% Dice scores, respectively [21]. Herein, models' performance was maximum 61% but model training was performed on a dataset of 78 patients. These differences in model performance can also be attributed to the fact than in [21], the authors have only estimated segmentation performance on the mid-gland region of the prostate where segmentation models tend to perform better than the apex, since the later may

present important shape and size variations. Despite all, additional evidence is required to prove that the suggested approach is superior to conventional methods like center-cropping.

5. Conclusion

A preprocessing technique is proposed to effectively overcome the class imbalance problem in prostate MRI segmentation tasks. The improvement in PZ segmentation performance of DL networks was significant when the proposed method was employed in comparison with the conventional center-cropping method. In the future, the generalizability of the proposed pipeline needs to be demonstrated on independent populations through external validation including images acquired by different MRI vendors, field properties and acquisition protocols.

Acknowledgement. This work is supported by the ProCancer-I project, funded by the European Union's Horizon 2020 research and innovation program under grant agreement No 952159. It reflects only the author's view. The Commission is not responsible for any use that may be made of the information it contains.

References

- [1] Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2021. *CA: A Cancer Journal for Clinicians*. 2021;71(1):7–33.
- [2] de Rooij M, Israël B, Tummers M, Ahmed HU, Barrett T, Giganti F, et al. ESUR/ESUI consensus statements on multi-parametric MRI for the detection of clinically significant prostate cancer: quality requirements for image acquisition, interpretation and radiologists' training. *European Radiology*. 2020;30(10):5404–5416.
- [3] Viswanath SE, Bloch NB, Chappelw JC, Toth R, Rofsky NM, Genega EM, et al. Central gland and peripheral zone prostate tumors have significantly different quantitative imaging signatures on 3 tesla endorectal, in vivo T2-weighted MR imagery. *Journal of Magnetic Resonance Imaging*. 2012;36(1):213–224.
- [4] McNeal JE, Redwine EA, Freiha FS, Stamey TA. Zonal distribution of prostatic adenocarcinoma. Correlation with histologic pattern and direction of spread. *The American Journal of Surgical Pathology*. 1988;12(12):897–906.
- [5] Hassanzadeh T, Hamey L, Ho-Shon K. Convolutional Neural Networks for Prostate Magnetic Resonance Image Segmentation. *IEEE Access*. 2019;7:36748–36760.
- [6] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2015;9351, p.234–241.
- [7] Goceri E. Challenges and Recent Solutions for Image Segmentation in the Era of Deep Learning. *2019 9th International Conference on Image Processing Theory, Tools and Applications, IPTA 2019*. 2019.

- [8] Feng X, Qing K, Tustison NJ, Meyer CH, Chen Q., Deep convolutional neural network for segmentation of thoracic organs-at-risk using cropped 3D images. *Medical Physics*. 2019;46(5):2169–2180.
- [9] To MNM, Vu DQ, Turkbey B, Choyke PL, Kwak JT. Deep dense multi-path neural network for prostate segmentation in magnetic resonance imaging. *International Journal of Computer Assisted Radiology and Surgery*. 2018;13(11):1687–1696.
- [10] Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, et al., The cancer imaging archive (TCIA): Maintaining and operating a public information repository. *Journal of Digital Imaging*. 2013;26(6):1045–1057.
- [11] Chen W, Zhang Y, He J, Qiao Y, Chen Y, Shi H, et al. Prostate segmentation using 2D bridged U-net. *arXiv*. 2018.
- [12] Zaridis DG, Mylona E, Tachos NS, Marias K, Papanikolaou N, Tsiknakis M, et al. A new smart-cropping pipeline for prostate segmentation using deep learning networks. *arXiv*. 2018.
- [13] Guan S, Khan L, Sikdar S et al., Fully dense UNet for 2D sparse photoacoustic tomography artifact removal. *arXiv*. 2018.
- [14] Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely connected convolutional networks. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*. 2017. p.2261–2269.
- [15] Zou KH, Warfield SK, Bharatha A, Tempany CM, Kaus MR, Haker SJ, et al. Statistical validation of image segmentation quality based on a spatial overlap index. *Academic radiology*. 2004;11(2):178–189.
- [16] Kingma D, Ba JL. Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. 2015. p.1–15.
- [17] Bottou L. Large-scale machine learning with stochastic gradient descent. In: *Lechevallier Y., Saporta G. (eds) Proceedings of COMPSTAT'2010*. Physica-Verlag HD. 2010.
- [18] Yeghiazaryan V, Voiculescu I. Family of boundary overlap metrics for the evaluation of medical image segmentation. *J Med Imaging (Bellingham)*. 2018; 5(1):15006.
- [19] Taghanaki SA, Abhishek K, Cohen JP, Cohen-Adad J, Hamarneh G. Deep semantic segmentation of natural and medical images: a review. *Artificial Intelligence Review*. 2021; 54(1):137–178.
- [20] Long J, Shelhamer E, Darrell T. Fully Convolutional Networks for Semantic Segmentation. *arXiv*. 2015.
- [21] Aldoij N, Biavati F, Michallek F, Stober S, Dewey M. Automatic prostate and prostate zones segmentation of magnetic resonance images using DenseNet-like U-net. *Scientific Reports*. 2020;10(14315).
- [22] Jaeger PF, Kohl SAA, Bickelhaupt S, Isensee F, Kuder TA, et al. Retina U-Net: Embarrassingly Simple Exploitation of Segmentation Supervision for Medical Object Detection. *Proceedings of the Machine Learning for Health NeurIPS Workshop*. 2020; 116:171–183.
- [23] Pellicer-Valero OJ, Marenco Jiménez JL, Gonzalez-Perez V, Casanova Ramón-Borja JL. Deep Learning for fully automatic detection, segmentation, and Gleason Grade estimation of prostate cancer in multiparametric Magnetic Resonance Images. *arXiv*. 2022.
- [24] Zlocha M, Dou Q, Glocker B. Improving RetinaNet for CT Lesion Detection with Dense Masks from Weak RECIST Labels. *arXiv*. 2019.
- [25] Lou A, Guan S, Ko H, Loew M. CaraNet: Context Axial Reverse Attention Network for Segmentation of Small Medical Objects *arXiv*. 2022.