PROper-Net: A Deep-Learning Approach for Prostate's Peripheral Zone Segmentation based on MR imaging

Eugenia Mylona FORTH Ioannina, Greece mylona.eugenia@gmail.com

Dimitris Zaridis FORTH Ioannina, Greece dimzaridis@gmail.com

Manolis Tsiknakis

Department of Electrical and Computer Engineering Hellenic Mediterranean University Heraklion, Greece tsiknaki@ics.forth.gr

Nikolaos Tachos Kostas Marias Biomedical Research Institute Biomedical Research Institute Biomedical Research Institute Institute of Computer Science FORTH FORTH Ioannina, Greece Heraklion, Greece ntachos@gmail.com kmarias@ics.forth.gr

> Dimitrios I. Fotiadis Unit of Medical Technology and Intelligent Information Systems University of Ioannina **Biomedical Research Institute** FORTH Ioannina. Greece fotiadis@uoi.gr

Abstract-Prostate cancer detection and characterization on Magnetic Resonance Images (MRI) requires accurate segmentation of the prostate gland and the prostatic sub-regions. With the majority of tumoral lesions located in the prostate's peripheral zone, a precise segmentation of this region is imperative for tumor characterization. Despite the growing success of Convolution neural networks (CNN) in the task of prostate gland segmentation, there a is a knowledge gap in the performance of such networks for segmenting prostatic subregions. In the present work, we propose an novel Deep Learning (DL) approach, named PROper-Net, for segmenting the prostate's peripheral zone on T2weighted (T2w) MR images. Our network was compared against four state-of-the-art encoder-decoder CNNs: the original Unet, and its extensions Unet++, Unet3+, and Bridged Unet. Overlapand distance-based metrics were computed to assess models' performance and to quantify the superiority of the proposed segmentation approach. The results show that the proposed network successfully outperforms existing networks for the peripheral zone segmentation task, yielding a median performance of 0.74 in terms of Dice Score and 0.88 in terms of balanced accuracy. The improvement in segmentation performance was significant (p-value < 0.05) with respect to Unet, Unet++, Unet3+ for all the evaluation metrics while for Bridged Unet significant improvement was achieved for Dice Score, Balanced Accuracy, Sensitivity, and Rand Error Index.

Index Terms-deep learning, segmentation, U-Net, prostate, MRI

I. INTRODUCTION

Diagnosis and treatment of prostate cancer at early stages requires a precise and efficient segmentation of the prostatic gland and its subregions on Magnetic Resonance (MR) Images. Within the prostate gland, tumoral lesions are most commonly identified in the transitional and peripheral zones. Specifically, 70%-80% of prostate cancers are located in the peripheral zone (PZ) [1]. Given the significantly worse

prognosis of PZ cancer, compared to lesions arising in the transitional zone (TZ), it is crucial to accurately segment this region and enable the accurate subsequent characterization of the tumor [2].

The MR imaging is one of the most reliable and effective medical imaging methods for prostate cancer diagnosis [3]. Despite the excellent soft-tissue contrast of MR imaging, manual prostate segmentation is a daunting and time-consuming task which requires certain clinical experience [4]. However, many traditional segmentation methods tend to underperform because of the morphological and positional properties of the prostate. Particularly the varying and irregular pixel representation of the peripheral zone's boundaries, make its automatic segmentation a challenging task [5].

Over the last decade, DL techniques and, particularly, those based on convolutional neural networks (CNN), have seen accelerating growth in pattern recognition, computer vision and medical image processing. Particularly for image segmentation tasks, the U-net has become the mainstream DL architecture [6]. In general, these models implement a variable number of hidden layers between the input and output and the various nodes are connected to others with different weights.

For prostate segmentation, many U-Net-based architectures have emerged recently, demonstrating promising results and outperforming traditional segmentation techniques. Although they often achieve a remarkable performance when trained and tested on images obtained by the same clinical center or under the same acquisition protocol, their generalizability on external, independent datasets is far from being established. The optimistic performances often reported may be datasetspecific and therefore such models are still immature for clinical routine applications [7]. This indicates the necessity to both develop more robust and reliable DL segmentation networks and to extend the evaluation of existing delineation approaches in a multi-center environment. Furthermore, only a few studies have evaluated the performance of these architectures for specifically segmenting the PZ of the prostate.

To account for these issues, in the present work we propose a novel U-Net-based model, named PROper-Net, for improving the segmentation performance of prostate's PZ on MR imaging. The performance of our network, trained and tested on different imaging datasets, is compared against four state-of-the-art DL-based segmentation networks using several evaluation metrics.

II. METHODOLOGY

A. Dataset

Two publicly available datasets were used for model training and testing. Model training was performed on 204 T2-Weighted MR patient images from the Prostate-X dataset, consisting of 3206 frames, along with the corresponding masks of the peripheral zone [8]. The vendor used was Siemens, with TrioTim and Skyra models. The prostate-X dataset partitioned into training and validation sets with 85% and 15% of the available data, respectively, while the partitioning remained the same for all the experiments. The performance of the segmentation model was tested on the Prostate 3T dataset [9]. It consist of 30 T2-Weighted MR patients comprising a total of 421 frames, with the corresponding annotations of the peripheral zone. The vendor used was Siemens, Skyra model.

B. Segmentation Networks

1) U-net: The U-net [6] is a long-established architecture in the biomedical imaging domain for segmentation applications. The network consists of an encoder-decoder combination, where the encoder captures the semantic features regarding the area of interest while the decoder upsamples that information assisted by the skipped connections.

2) U-net++: The U-net++ [10] is another encoder-decoder model based on the original U-net. Its novelty arises from the ensembled networks inside the outer encoder-decoder. Specifically U-net++ has three encoder-decoder networks with each stage being more swallow than the previous one. The outputs from each individual encoder-decoder network are connected to produce the final ensembled prediction.

3) U-net3+: The U-net3+ architecture [11] also belongs to the U-net family of models. Its composition consists of fully connected stages between each layer. Specifically, each block from the encoder structure is concatenated into each block of the decoder structure, transferring the semantic information from simple substructures that composes the region of interest to more complex ones. Furthermore, in the decoder path all the decoder are communicating with each other, minimizing the effect of vanishing gradient problem [12].

4) Bridged U-net: The Bridged U-Net [13], on the other hand, combines two interconnected U-net substructures, forming a W-shaped network, to transfer the information and further analyze the image into more complex patches. Feature

maps are fused together using concatenation, thereby improving the performance of the network.

5) *PROper-Net:* The network proposed in this study is shown in Fig.1. Residual connections [14] have been used in order to improve the information transfer and tackle the vanishing gradient issue (Eq. 1), and a Batch normalization layer has been used for standardizing each batch of data and improve model's robustness (Eq. 2):

$$O_{i,j,fm} = H(I_{i,j,fm}) + I_{i,j,fm},$$
 (1)

where i, j, fm are width, height and the number of feature maps respectively, $I_{i,j,fm}$ is each layer's input, $O_{i,j,fm}$ is each layer's output and $H(I_{i,j,fm})$ is the output of convolution operations.

$$BNi, j, fm = \frac{I_{i,j,fm} - I_{fm}}{\sigma_{I_fm}},$$
(2)

where i, j, fm are width, height and the number of feature maps respectively, $I_{i,j,fm}$ is each layer's input, BNi, j, fmis the normalized batch and σ_{I_fm} is the standard deviation of the input. Furthermore, four layers of encoder and decoder paths were utilized to identify complex patterns. Specifically, the bottleneck block has 1024 feature maps to recognize more complex features. While most architectures come with a bottleneck layer of 512 feature maps due to frames' size, for our network 1024 were used due to their capability to recognize complex structures such as the prostate's peripheral zone. On the other hand, due to the complexity and the workforce load of the model, max pooling operations were utilized for dimensionality reduction and feature selection while dropout layers were used to neglect non-important neurons and further improve the information flow through the model.



Fig. 1. The proposed PROper-Net architecture. W,H and C represent the width height and number of channels in the input and output

C. Model specification

Model consists of 2D convolutional layers with kernel size 3x3, dilation rate 1 and stride 1. Batch normalization used L2 regularization technique for gamma and beta regularizers with a value of 10^{-4} . The max pooling layer has stride value of 2x2 while the dropout rate was 0.5. Furthermore, batch size was kept at 16, the number of epochs set to 120 while

 TABLE I

 MEDIAN VALUES AND INTERQUARTILE RANGE OF PZ SEGMENTATION PERFORMANCE ON THE TEST SET FOR THE FIVE DL NETWORKS.

Models	Sensitivity	Balanced accuracy	Dice score	Hausdorff Distance (mm)	Average surface distance (mm)	Rand error index
PROper-Net	0.77 [0.53-0.86]	0.88 [0.76-0.93]	0.75 [0.58-0.85]	7.42 [3.65-14.82]	1.50 [0.95-2.52]	0.20 [0.13-0.30]
Bridged U-net	0.61 [0.42-0.76]	0.81 [0.71-0.88]	0.70 [0.55-0.83]	6.84 [3.47-12.4]	1.46 [0.86-2.33]	0.26 [0.19-0.32]
U-net	0.60 [0.33-0.70]	0.80 [0.67-0.85]	0.68 [0.46-0.80]	9.02 [5.18-16.11]	1.83 [1.17-3.35]	0.27 [0.18-0.31]
U-net++	0.52 [0.4-0.66]	0.76 [0.70-0.83]	0.60 [0.44-0.71]	7.98 [5.18-13.16]	2.15 [1.52-2.92]	0.31 [0.24-0.33]
U-net3+	0.52 [0.27-0.67]	0.76 [0.63-0.83]	0.62 [0.39-0.76]	9.12 [4.97-14.83]	2 [1.29-3.24]	0.28 [2-0.32]

early stopping used to halt the training process when the model didn't improve further at the validation set. Binary crossentropy used as a loss function and Adam optimizer employed for model's weights updating while the learning rate set to 10^{-4} . The GPU used for the experiments is the NVIDIA Quadro P6000, the drivers are of version 441.66. The python packages utilized for this experiment are scikit-image=0.18.3, numpy=1.21.2, keras-unet-collection=0.1.11, scipy=1.7.1, tensorflow=2.2.0 and tensorflow-addons=0.11.2.

D. Evaluation metrics

For the evaluation of the results six metrics have been used. They have been carefully selected as each one serves a specific purpose. For instance, Dice Score and Rand Error Index are overlap metrics while Average surface distance and Hausdorff distance are distance metrics.

Sensitivity is a measure of model's ability to identify pixels that belong to the object of interest, based on true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN), and it is given as:

$$Sensitivity = \frac{TP}{TP + FN}.$$
(3)

Balanced accuracy(BA) on the other hand is used when there is class imbalance and background pixels are prevalent to foreground and it is given as:

$$BA = \frac{Sensitivity + \frac{TN}{TN + FP}}{2}.$$
 (4)

Dice score is a well established metric for segmentation applications and its form is given as:

$$Dice \ score = \frac{2 * TP}{2 * TP + FP + FN}.$$
(5)

Rand error index (REI) measures the error of intensity correlation between two clusters of intensity values and it is given by as:

$$REI = 1 - \frac{TP + TN}{TP + TN + FP + FN}.$$
(6)

Hausdorff Distance (HD) measures how far are two data points, each of them originated by ground truth and predicted boundaries. In this work, the 95% HD has been used to avoid extreme values of the metric as they may not be representative of the actual model performance. The expression that defines HD is given as:

$$HD(G,P) = max(h(G,P),h(P,G)),$$
(7)

where G, P are the ground truth and predicted boundaries, h(G, P) is the closest distance from ground truth points to predicted and h(P, G) are the closest distance from predicted points to ground truth.

Average surface distance (AVSD) is an one dimensional metric that measures the average distance which is the euclidian norm from one point of the ground truth boundary to the closest point of the predicted boundary and it is given as:

$$AVSD = \frac{\int_A \min||pnt(g) - pnt(p)|| \, dg \, dp}{A},\tag{8}$$

where A is the total number of points, pnt(g) is point in ground truth boundary and pnt(p) is point in predicted boundary.

In order to assess whether the performance of the proposed segmentation network is significantly different than the other networks in comparison, the non-parametric two-sided Wilcoxon signed-rank test was used.

III. RESULTS

In Table I, the median values of the six performance metrics computed on the testing dataset for the proposed segmentation algorithm along with the four state-of-the-art networks, are presented. PROper-Net outperformed U-net, U-net++ and Unet3+ for all the evaluation metrics considered. Compared to Bridged U-net model, PROper-Net performed better in terms of Dice Score, BA, Sensitivity and REI but scores were slightly lower for HD and AVSD metrics. The corresponding p-values computed by the 2-sided Wilcoxon signed-rank test, are provided in the form of heatmap in Fig. 2. Non-significant differences were identified only for HD an AVSD metrics.



Fig. 2. Heatmap of p-values for the evaluation metrics between PROper-NET and all the other segmentation networks.



Fig. 3. Four examples of the predicted contours for each network.

In Fig. 3 four examples of the derived PZ delineations are shown for the different DL segmentation networks. The predicted segmentations are depicted with blue colours while the ground truth is shown in orange. As it is shown in Fig. 3, cases 1 and 3 are some examples of difficult cases for the segmentation task compared to cases 2 and 4 were all the algorithms perform relatively well.

IV. DISCUSSION AND CONCLUSION

A new DL-based approach was presented, for improving the segmentation of prostate's PZ on T2w MR images, evaluated on an external population data and compared against state-of-the-art segmentation networks. As it was shown, the PROper-Net algorithm outperformed all the other networks. The achieved overlap-based indices, prove that segmentation accuracy in terms the algorithm's capability to correctly detect true labels while not detecting wrong labels withing the segmented PZ, were superior compared to other networks. A similar trend is apparent for spatial distance-based metrics, with the exception of Bridged-Unet, which performs marginally better.

It is well known that prostate gland segmentation is not trivial but prostate zonal segmentation is even more challenging [4]. Specifically for automatic PZ segmentation, there are several factors that hamper models performance, namely the ambiguity of the PZ boundary, tissue heterogeneity, partial volume effects, as well as the presence of artifacts and the large differences between the number of foreground and background pixels in the image. For instance, by applying a smart cropping technique on prostate MR images [15] to counterbalance background and foreground pixels, was able to significantly improve PZ segmentation performance of different DL algorithms [16].

From a clinical perspective, the precise identification of prostate cancers' location has important diagnostic and prognostic implications. In general, tumors located in the PZ result to poor clinical outcomes and worse biochemical recurrence-free survival compared to cancers of the TZ. Furthermore, some authors have identified differences in biological and molecular characteristics between TZ and PZ cancer [17].

Over the last decades, several types of segmentation techniques have been developed [18], including among others, graph cuts [19], [20], deformable segmentation [21], active appearance models [22], [23], random walks [24], atlas matching [25], watersheds [26] and c-means clustering [27], [28]. After the success of CNN in computer vision and object detection, it recently attracted remarkable attention for medical image segmentation tasks as novel DL-based algorithms consistently outperformed more conventional approaches.

To date, apart form whole gland segmentation, only a few studies have tackled the problem of PZ segmentation. In Chilali et al., prostate and zonal segmentation was performed using atlas-based techniques and c-means clustering [28]. Karimi et al. introduced statistical shape models to a CNN algorithm [31] demonstrating superior performance to conventional CNN techniques. Rundo et al. describe an architecture to segment prostatic zones where a squeeze-and-excitation module has been incorporated into the U-net, outperforming other state-of-the-art algorithms [29]. More recently, Aldoj et al. proposed different combinations of Unet and DenseNet algorithms to segment the prostate gland and its subregions achieving a performance superior to the standard U-net [4].

Future works should focus on more selective segmentation approaches, to simultaneously delineate the whole gland and multiple prostatic zones. This can also be extended to segment prostate cancer using a two-step approach, focusing on pathological regions in the PZ or other regions of clinical interest. Additionally, 3D segmentation networks are emerging as a promising alternative for further augmenting DL performance, by taking into account the spatial information among slices. Nevertheless, more research is necessary to establish their superiority, particularly for zonal segmentation of the prostate.

ACKNOWLEDGMENTS

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

- McNeal J., Redwine E., Freiha F., Stamey T., Zonal distribution of prostatic adenocarcinoma. Correlation with histologic pattern and direction of spread. *The American Journal of Surgical Pathology*. **12** pp. 897-906 (1988)
- [2] Viswanath S., Bloch N., Chappelow J., Toth R., Rofsky N., Genega E., 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 : JMRI*. 36 pp. 213-224 (2012)
- [3] De Rooij M., Israël B., Tummers M., Ahmed H., 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*. **30** pp. 5404-5416 (2020)
- [4] Aldoj 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.* **10**, 14315 (2020)
- [5] Qian Y., ProSegNet: A New Network of Prostate Segmentation Based on MR Images. *IEEE Access.* 9 pp. 106293-106302 (2021)
- [6] Ronneberger O., Fischer P., Brox T., U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing And Computer-Assisted Intervention – MICCAI 2015*, pp. 234-241 (2015)
- [7] Bluemke D., Moy L., Bredella M., Ertl-Wagner B., Fowler K., Goh V., et al., Assessing Radiology Research on Artificial Intelligence: A Brief Guide for Authors, Reviewers, and Readers—From the Radiology Editorial Board. *Radiology*. 294. pp. 487-489 (2020)
- [8] Cuocolo R., Stanzione A., Castaldo A., De Lucia D., Imbriaco M., Quality control and whole-gland, zonal and lesion annotations for the PROSTATEx challenge public dataset. *European Journal Of Radiology*. 138 (2021)

- [9] Litjens G., Futterer J., Huisman H., Data From Prostate-3T. The Cancer Imaging Archive. (2020)
- [10] Zhou Z., Siddiquee M., Tajbakhsh N., Liang, J., UNet++: A Nested U-Net Architecture for Medical Image Segmentation. (2018)
- [11] Huang H., Lin L., Tong R., Hu H., Zhang Q., Iwamoto Y., et al., UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation. (2020)
- [12] Hochreiter, S. The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions. *International Journal Of Uncertainty, Fuzziness And Knowledge-Based Systems.* 6 pp. 107-116 (1998)
- [13] Chen W., Zhang Y., He J., Qiao Y., Chen Y., Shi H., Tang X., W-net: Bridged U-net for 2D Medical Image Segmentation. ArXiv. (2018)
- [14] He K., Zhang X., Ren S., Sun J., Deep Residual Learning for Image Recognition. (2015)
- [15] Zaridis D., Mylona E., Marias K., Papanikolaou N., Tachos N., Fotiadis D., A new smart-cropping pipeline for prostate segmentation using deep learning networks. ArXiv. (2021)
- [16] Zaridis D., Mylona E., Tachos N., Marias K., Tsiknakis M., Fotiadis D., A Deep Learning-based cropping technique to improve segmentation of prostate's peripheral zone. 2021 IEEE 21st International Conference On Bioinformatics And Bioengineering (BIBE). pp. 1-4 (2021)
- [17] Lee J., Thomas I., Nolley R., Ferrari M., Brooks J., Leppert J., A. Biologic differences between peripheral and transition zone prostate cancer. *Prostate*. **75** (2015)
- [18] Comelli A., Dahiya N., Stefano A., Vernuccio F., Portoghese M., Cutaia G., et al., A. Deep Learning-Based Methods for Prostate Segmentation in Magnetic Resonance Imaging. *Applied Sciences.* 11 (2021)
- [19] Korsager A., Fortunati V., Lijn F., Carl J., Niessen W., Østergaard L., et al., The use of atlas registration and graph cuts for prostate segmentation in magnetic resonance images. *Medical Physics*. 42 pp. 1614-1624 (2015)
- [20] Tian Z., Liu L., Zhang Z., Fei B., Superpixel-Based Segmentation for 3D Prostate MR Images. *IEEE Transactions On Medical Imaging*. 35 pp. 791-801 (2016)
- [21] Guo Y., Gao Y., Shen D., Deformable MR Prostate Segmentation via Deep Feature Learning and Sparse Patch Matching. *IEEE Transactions* On Medical Imaging. 35 pp. 1077-1089 (2016)
- [22] Toth R., Madabhushi A., Multifeature Landmark-Free Active Appearance Models: Application to Prostate MRI Segmentation. *IEEE Transactions On Medical Imaging*. **31** pp. 1638-1650 (2012)
- [23] Vincent G., Guillard G., Bowes M., Fully Automatic Segmentation of the Prostate using Active Appearance Models. (2012)
- [24] Li A., Li C., Wang X., Eberl S., Feng D., Fulham M., Automated Segmentation of Prostate MR Images Using Prior Knowledge Enhanced Random Walker. 2013 International Conference On Digital Image Computing: Techniques And Applications (DICTA). pp. 1-7 (2013)
- [25] Klein S., Heide U., Lips I., Vulpen M., Staring M., Pluim J., Automatic segmentation of the prostate in 3D MR images by atlas matching using localized mutual information. *Medical Physics*. 35 pp. 1407-1417 (2008)
- [26] Huang Z., Jiang S., Yang Z., Ding Y., Wang W., Yu Y., Automatic multi-organ segmentation of prostate magnetic resonance images using watershed and nonsubsampled contourlet transform. *Biomed. Signal Process. Control.*. 25 pp. 53-61 (2016)
- [27] Rundo L., Militello C., Russo G., Garufi A., Vitabile S., Gilardi M., et al., Automated Prostate Gland Segmentation Based on an Unsupervised Fuzzy C-Means Clustering Technique Using Multispectral T1w and T2w MR Imaging. *Information.* 8 (2017)
- [28] Chilali O., Puech P., Lakroum S., Diaf M., Mordon S., Betrouni N., Gland and Zonal Segmentation of Prostate on T2W MR Images. *Journal Of Digital Imaging*. **29** pp. 730-736 (2016)
- [29] Rundo L., Han C., Nagano Y., Zhang J., Hataya R., Militello C., et al., P. USE-Net: Incorporating Squeeze-and-Excitation blocks into U-Net for prostate zonal segmentation of multi-institutional MRI datasets. *Neurocomputing*. **365** pp. 31-43 (2019)
- [30] Rundo L., Han C., Zhang J., Hataya R., Nagano Y., Militello C., et al., Mauri, G., CNN-based Prostate Zonal Segmentation on T2-weighted MR Images: A Cross-dataset Study. arXiv. (2019)
- [31] Karimi D., Samei G., Kesch C., Nir G., Salcudean S., Prostate segmentation in MRI using a convolutional neural network architecture and training strategy based on statistical shape models. *International Journal Of Computer Assisted Radiology And Surgery*. 13 pp. 1211-1219 (2018)