Automatic Zonal Segmentation of the Prostate from 2D and 3D T2-weighted MRI and Evaluation for Clinical Use

Dimitri Hamzaoui, Sarah Montagne, Raphaele Renard-Penna, Nicholas Ayache, Hervé Delingette

To cite this version:

Dimitri Hamzaoui, Sarah Montagne, Raphaele Renard-Penna, Nicholas Ayache, Hervé Delingette. Automatic Zonal Segmentation of the Prostate from 2D and 3D T2-weighted MRI and Evaluation for Clinical Use. Journal of Medical Imaging, 2022, 9 (2), pp.024001. 10.1117/1.JMI.9.2.024001. hal-03587074v2

HAL Id: hal-03587074

https://hal.science/hal-03587074v2

Submitted on 16 Mar 2022

Automatic zonal segmentation of the prostate from 2D and 3D T2-weighted MRI and evaluation for clinical use

Dimitri Hamzaoui,[a,*] **Sarah Montagne**,[b] **Raphaële Renard-Penna**,[b]
Nicholas Ayache,[a] and **Hervé Delingette**[a]
[a]Université Côte d'Azur, Inria, Epione Project-Team, Sophia Antipolis, Valbonne, France
[b]Sorbonne Université, Radiology Department, CHU La Pitié Salpétrière/Tenon, Paris, France

Abstract

Purpose: An accurate zonal segmentation of the prostate is required for prostate cancer (PCa) management with MRI.

Approach: The aim of this work is to present UFNet, a deep learning-based method for automatic zonal segmentation of the prostate from T2-weighted (T2w) MRI. It takes into account the image anisotropy, includes both spatial and channelwise attention mechanisms and uses loss functions to enforce prostate partition. The method was applied on a private multicentric three-dimensional T2w MRI dataset and on the public two-dimensional T2w MRI dataset ProstateX. To assess the model performance, the structures segmented by the algorithm on the private dataset were compared with those obtained by seven radiologists of various experience levels.

Results: On the private dataset, we obtained a Dice score (DSC) of 93.90 ± 2.85 for the whole gland (WG), 91.00 ± 4.34 for the transition zone (TZ), and 79.08 ± 7.08 for the peripheral zone (PZ). Results were significantly better than other compared networks' (p-value < 0.05). On ProstateX, we obtained a DSC of 90.90 ± 2.94 for WG, 86.84 ± 4.33 for TZ, and 78.40 ± 7.31 for PZ. These results are similar to state-of-the art results and, on the private dataset, are coherent with those obtained by radiologists. Zonal locations and sectorial positions of lesions annotated by radiologists were also preserved.

Conclusions: Deep learning-based methods can provide an accurate zonal segmentation of the prostate leading to a consistent zonal location and sectorial position of lesions, and therefore can be used as a helping tool for PCa diagnosis.

© *2022 Society of Photo-Optical Instrumentation Engineers (SPIE)* [DOI: 10.1117/1.JMI.9.2.024001]

Keywords: prostate; segmentation; deep learning; lesion; magnetic resonance imaging; inter-rater variability.

Paper 21261R received Sep. 24, 2021; accepted for publication Feb. 23, 2022; published online Mar. 14, 2022.

1 Introduction

Prostate cancer (PCa) is the most frequent type of cancer affecting men in Europe and North America with more than 190,000 expected cases in the United States in 2021. It is estimated that 12% of American men will develop PCa during their life.[1] For years, the standard imaging modality used to guide biopsies was transrectal ultrasound, prone to underdetection of clinically significant PCa and to overestimation of benign lesions.[2] Now, its replacement by multiparametric-MRI (mp-MRI) is supported by several medical associations such as the European Association of Urology[3] and the American Urological Association.[4] Based on mp-MRI, the PI-RADS score[5] is designed to improve detection, localization, characterization, and risk stratification in patients with suspected PCa in treatment-naive prostate glands. It uses a five-point

---

[*] Address all correspondence to Dimitri Hamzaoui, dimitri.hamzaoui@inria.fr

2329-4302/2022/$28.00 © 2022 SPIE

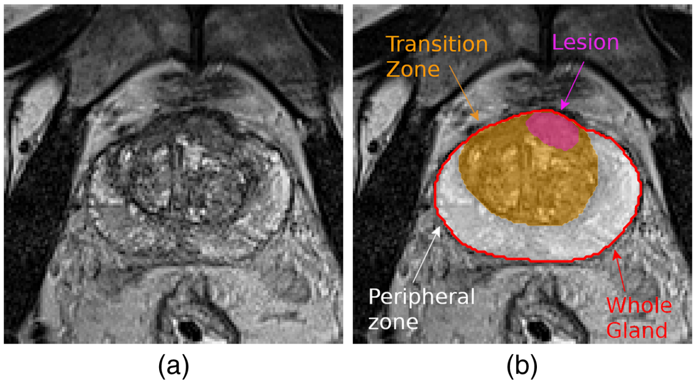


Fig. 1 (a) Axial view of the T2-weighted MR image of a prostate. (b) The corresponding zonal and lesion segmentation. The WG is the union of TZ and PZ.

scale based on the probability that a combination of mp-MRI findings on T2-weighted (T2w), diffusion-weighted (DWI), and dynamic contrast-enhanced sequences correlates with the presence of clinically significant PCa. Then, the patient will benefit from standard and targeted biopsies if a suspected lesion is detected.[2,6–9] PI-RADS defines a dominant sequence for each zone of the prostate: T2w for the transition zone (TZ) and DWI for the peripheral zone (PZ); so identification of the zonal location of a lesion is vital. Both zones are shown in Fig. 1.

In addition, to locate findings on MRI reports and to simplify discussions about biopsies and treatment, radiologists and urologists have defined sector maps that are based on those zones and on longitudinal, transverse, and anteroposterior directions. Since both PI-RADS scores and sectorial positions are subject to a high inter-rater variability,[10–12] there is a need for automated PCa diagnosis methods.

When required, the manual segmentation of prostate zones is commonly performed from T2w sequences. But several factors complicate this task and make it time-consuming even for a skilled physician.[13] First, boundaries of the prostatic gland and inner boundaries between TZ and PZ may be hard to detect. Second, the prostate is subject to an important intersubject variability due to physiological differences in terms of shape, size, and tissue intensities.[14,15] Finally, sequences acquired from different MRI machines increase the variability in appearance of the prostate in T2w imaging.

1.1 *Related Works*

Several authors proposed computerized methods for the automatic segmentation of the prostate from T2w MRI sequences. In 2012, the PROMISE12 challenge, dedicated to the segmentation of the whole prostatic gland (WG),[16] took place and was won by Vincent et al.[17] using active appearance models. Meanwhile, convolutional neural networks (CNN) began to provide promising results, especially in image classification.[18] Among the different architectures, UNet[19] appeared to be adapted to biomedical image segmentation. Furthermore, in 2016, Milletari et al.[20] presented V-net, a three-dimensional (3D) UNet variation with a Dice similarity coefficient (DSC)-based loss function dedicated to the automatic segmentation of the prostate with consistent results on PROMISE12 (mean DSC of $86.9 \pm 3.3\%$) and was the first of many works on WG segmentation using deep learning. For example, in 2017, Cheng et al.[21] used holistically nested networks[22] and coherence-enhancing diffusion filters[23] to perform this task, and in 2018 Tian et al.[24] studied the use of transfer learning from large-scale datasets. Now, CNN have become the most widely used methodology for automated segmentation of WG, with best mean DSC on PROMISE12 comprised between 91.5% and 93%.[25,26]

Although WG prostate segmentation is performed successfully in many cases, the zonal segmentation of the prostate is more difficult especially for the PZ. Indeed, in addition to having a croissant-like shape in axial views, this zone is subject to an important intersubject variability. For the zonal segmentation of the prostate, many authors used two-dimensional (2D) neural networks, as the large anisotropy of 2D T2w MRI sequences makes them closer to stacked

2D images than to real 3D volumes. To improve the generalization on previously unseen datasets, Rundo et al.[27] proposed a 2D UNet for the zonal segmentation with squeeze-and-excitation modules. Aldoj et al.[28] conceived a DenseNet-like network to perform a zonal segmentation, with a DSC of 92.1 ± 0.8% for WG and 89.5 ± 2% for the TZ. Cuocolo et al.[29] compared the classic 2D UNet with efficient neural network (ENet)[30] and efficient residual factorized ConvNet (ERFNet)[31] which aim to limit their number of parameters and computation times while keeping a high level of performance. All those articles used the public dataset ProstateX[32–34] consisting of 2D T2w MRI sequences with a slice thickness of 3 mm, originally dedicated to PCa diagnosis.[35] Several works also considered 3D neural networks to take into account the volumetric consistency between slices. Bardis et al.[36] used a combination of 3D UNets to respectively locate the whole prostate, segment the prostate in the image, and classify each voxel of the image as TZ or PZ. Meyer et al.[37] performed zonal segmentation with a 3D neural network including anisotropic MaxPooling and deconvolutions to perform a zonal segmentation of the prostate. Zavala-Romero et al.[38] used a multiplanar 3D neural network[39] to perform this zonal segmentation and studied in particular the impact of MRI vendors on generated segmentations, showing the importance of multicentric and multivendor datasets.

1.2 *Limitations*

All those works obtained good results regarding the zonal segmentation of the prostate. However, they did not tackle the issue of localizing prostate lesions within zones and sectors. This localization is important for grading those lesions in the PI-RADS standard.

Moreover, prior works were based on 2D T2w sequences since they are the most widely available modality. Yet, 3D T2w prostate MR images allow a shorter acquisition time[40,41] and simplify modalities fusion among other advantages,[42] while having similar performances in terms of diagnosis. For those reasons, they will probably become the new radiological standard in prostate imaging in the near future. Differences between 2D and 3D T2w sequences are shown in Fig. 2.

In addition, inter-rater variability in the segmentation of prostate zones and whole gland (WG) has not been widely considered.[28,37,43] The number of human raters was limited to three, and no consensus was built for the prostate zones.

1.3 *Contributions*

To cope with the previous limitations, we introduce in this paper the following contributions:

- To the best of our knowledge, we provide the first prostate zonal segmentation method on 3D T2w MR images in addition to 2D T2w MRI. The obtained results are similar to the state of the art.
- We propose a deep learning-based framework for the automatic zonal segmentation of the prostate (TZ and PZ), including a neural network architecture. This architecture takes into

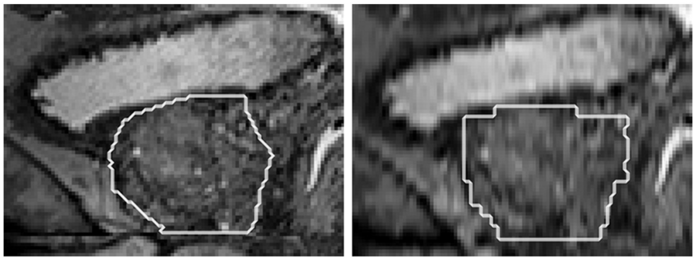


Fig. 2 (a) Sagittal view of a 3D T2w MR image of the prostate and its segmentation by a radiologist (slice thickness: 1 mm). (b) Sagittal view of a 2D T2w MR image of the same prostate and its segmentation from the axial views by the same radiologist, resampled to the resolution of the 3D T2w MRI (original slice thickness: 3.25 mm).

consideration the anisotropy of the data and includes dual attention mechanisms to improve the zonal segmentation. Partition loss functions were defined to enforce the partition of the prostate.

- We compare the generated segmentations with the ones supplied by seven radiologists of various experience levels, from which we derive a consensus segmentation. We show that our network performs similarly to the radiologists.
- Finally, we show that our method globally preserves both the zonal location and the sectorial position of lesions of the prostate, making it suitable as a helping tool for the detection and grading of lesions. Furthermore, we propose the first computerized method to generate a prostate's sector map from its zonal segmentation.

2 Material and Methods

2.1 *Dataset*

2.1.1 *MRI scans*

In this work approved by our joint institutional review boards, we used a private dataset of 131 3D T2w MR images from treatment-naive patients who underwent a prostate MRI before the first round of biopsy for clinical suspicion of PCa (linked to an elevated prostate-specific antigen, a positive digital rectal evaluation and a genetic susceptibility) between October 2013 and July 2019 from 3T Siemens scanners (Siemens Healthcare, Erlangen, Germany) on Pitié-Salpêtrière Hospital, Paris, France (100, 76.6%) and from 3T G.E. scanners (GE Healthcare; Chicago, Illinois) on Tenon Hospital, Paris, France (31, 23.4%). This dataset was built to have a diversity in terms of shapes, sizes, and volumes. The voxel dimensions are [0.36 to 0.78, 0.36 to 0.78, 0.5 to 1.0] mm. A random split of 91/40 (69%/31%) patients has been used between the training-validation set and the test set. In practice, the former was split into fivefolds in a cross-validation strategy where fourfolds served as a training set and the fifth one as a validation set. A minority (35.9%) had at least a clinically significant lesion, which was defined as a lesion with a PI-RADS score ≥ 3, of which 30 being on the training-validation set (33.0% of the training-validation set) and 17 on the test set (42.5%).

To assess the capacity of our network to provide a segmentation preserving not only the zonal location of the lesions but also their sectorial positions, we had access to an additional dataset of 33 3D T2w MRI sequences of prostates with 46 clinically significant lesions that we will call private lesion dataset. These sequences have been acquired on the same scanners than the private dataset between May 2017 and December 2019, with 24 from Tenon Hospital, Paris, France (73%, voxel dimensions: [0.547, 0.547, 0.5] mm) and nine from Pitié-Salpêtrière Hospital, Paris, France (27%, voxel dimensions: [0.36 to 0.78, 0.36 to 0.78, 0.5 to 1.0] mm).

In addition, we considered the public dataset ProstateX[32–34] to compare our method with prior works. It consists of 204 T2w MR images taken with 3T Siemens scanners on Radboud University Medical Center, with an in-plane dimension of $[0.375 \text{ to } 0.6]$ mm^2 and a slice thickness of [3 to 4.5] mm, the most frequent resolution being $0.5 \times 0.5 \times 3$ mm. We excluded three sequences for mismatches with their provided segmentation and randomly split the remaining data into a training-validation set of 141 sequences and a test set of 60 prostates, with a fivefold cross-validation strategy similar to the one used on the private dataset.

2.1.2 *Zonal segmentation*

The zonal segmentation of the private dataset consists of binary masks of the WG and the TZ. The segmentation of the training-validation set has been performed by a single expert radiologist, whereas on the test set seven radiologists of various levels of experience provided each a zonal segmentation for each prostate: three experts ($\geq$1000 prostate MRI interpreted), two seniors ($\approx$500 prostate MRI), and two juniors ($\leq$100 prostate MRI). This led to a total of 280 ($=40 \times 7$) zonal segmentations on the test set, for a rich comparison of performance with radiological experts. The radiologists were instructed to first segment WG and then TZ on the axial

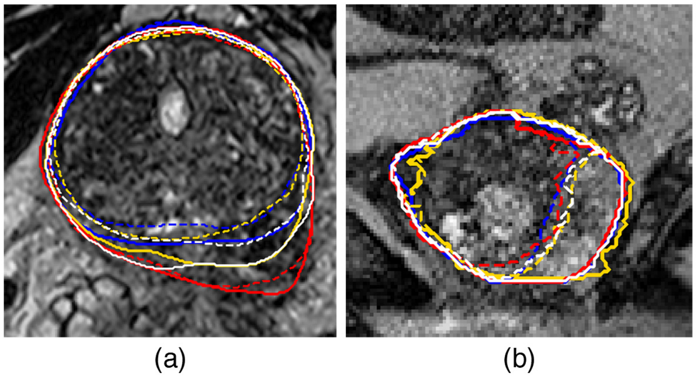

(a) (b)

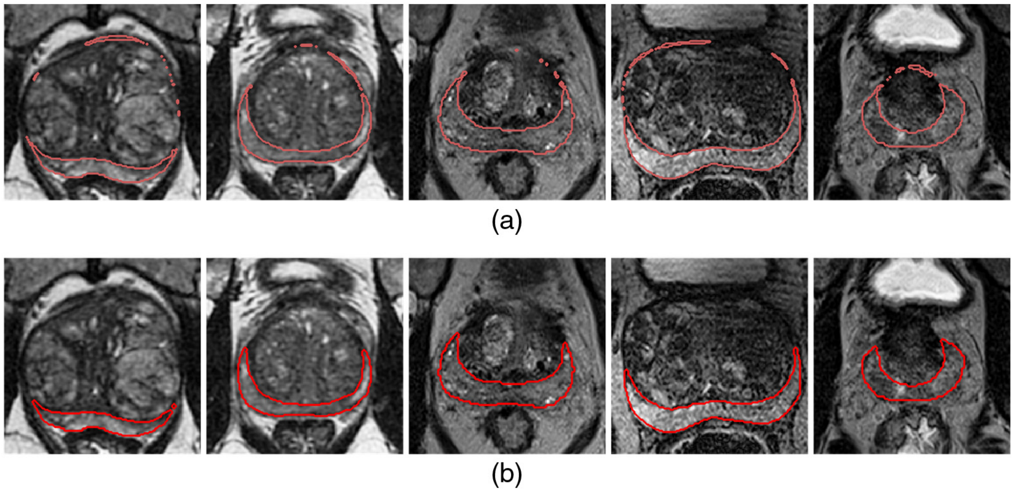
Fig. 3 Inter-rater variability, with segmentations from three of the seven raters (red, blue, yellow) and consensus segmentation from STAPLE using the seven raters (white), on (a) axial and (b) sagittal views. Solid line: WG, dashed line: TZ.

plane of the 3D T2w MRI sequence of our cohort. PZ was obtained by subtracting TZ to WG. Segmentation was performed using MedInria, an open-source software.[44]

In addition to these segmentations, we also generated a consensus segmentation for WG and TZ using the STAPLE algorithm,[45] which describes raters' binary segmentations by Bernoulli distributions and uses an expectation-maximization (EM) algorithm to produce a consensus of the segmentations. We binarized the obtained consensus by only keeping voxels with a probability ≥ 0.75. This threshold was chosen empirically, as its value varies following the different authors between 0.5 and 0.95.[46–48] PZ was then obtained by subtracting the consensus TZ to the consensus WG. An example with different raters segmentation and the consensus is shown in Fig. 3.

As the segmentations supplied by the radiologists have been prone to intrarater variability, leading to some gaps between TZ and WG border on the anterior part of the prostate which, after verification with a radiologist, do not belong to PZ, we applied on the initially determined PZ a slicewise 2D erosion, followed by a restriction to its largest connected component or the two largest components if the second component is at most three times smaller than the largest component, then a slicewise 2D dilation. Examples of the impact of those corrections are shown in Fig. 4.

For ProstateX, we used the zonal segmentation provided by Cuocolo et al.[29,49] No zonal segmentation has been done on the private lesion dataset.


(a)

(b)

Fig. 4 (a) Examples of segmentations of the PZ (= WG − TZ) before correction. (b) Same segmentations after correction.

2.1.3 *Lesion placement*

On the private dataset, each expert or senior radiologist (five of the seven raters) provided a segmentation of the lesions, from which we derived a consensus using the STAPLE algorithm as explained in Sec. 2.1.2. On the private lesion dataset, a radiologist provided for each lesion its sectorial position according to the 27 regions of interest sector map defined in Dickinson et al.[50]—with for some lesions two or three sectors indicated, as well as their size, their PI-RADS score, and their Likert score.[51]

2.2 *Objectives and Architecture of the Networks*

For the zonal segmentation of the prostate, we chose a framework with two cascaded UNet-based neural networks. A sum-up of the whole framework is shown in Fig. 5.

The objective of the first network, that we will call global location network, is to roughly segment WG to generate a bounding box around the prostate with a fixed size of 8 cm according to transverse and anteroposterior directions, and a margin of 10 cm above and below that segmentation. It takes as inputs patches of size $192 \times 192 \times 32$ voxels at a resolution of $1 \times 1 \times 3$ mm and uses as a loss function the generalized Dice loss function.[52] For image resampling, we used the python module SimpleITK,[53] with BSpline interpolation for images and nearest-neighbors interpolation for masks.

The second network, or zonal segmentation network, operates at a higher resolution of $0.5 \times 0.5 \times 1$ mm and takes input patches of $96 \times 96 \times 48$ voxels, which appeared to be a good compromise between the quantity of information brought to the network and the available GPU memory. Resampled and rescaled images had a size of $160 \times 160 \times [57 \text{ to } 110]$ voxels. The loss functions used for the training of the zonal segmentation network are defined in Sec. 2.4. Detailed architecture of the networks is shown in Fig. 6. In the final framework, we used UFNet for both the global location network and the zonal segmentation network, but with fewer parameters for the former. We combined in UFNet two methods to take into account the existing anisotropy of the data. As in Ref. 54, we used anisotropic MaxPooling and 3D deconvolutions with varying kernel sizes, and we replaced the classic $k \times k \times k$ kernels by a combination of $k \times k \times 1$ and $1 \times 1 \times k$ as presented in Fig. 6(d) inspired by Liu et al.[55] We used the activation function LeakyReLU with a parameter $\alpha = 0.1$ except for the last layer which uses sigmoid activation. In UFNet, we used deep supervision, which consists of introducing upscaled versions of intermediate results from the decoder into the final result as a form of regularization. We also performed Dropout[56] and instance normalization[57] to fight against overfitting and improve stability of our network, and we used attention modules which are presented in Sec. 2.3.

2.3 *Attention Mechanisms*

Attention in deep learning consists of encouraging the network to focus on some specific parts of the data, deemed with particularly relevant information for its task, and to downplay the importance of the rest of the data. The information can be highlighted based on its spatial location (spatial attention) or on the characteristics of the feature maps that contains it (channel attention). Here, we combined both channel attention and spatial attention through two different methods: squeeze-and-excitation modules and attention gates, respectively.[58,59]

2.3.1 *Squeeze-and-excitation modules*

The objective of squeeze-and-excitation modules[58] is to put more focus on the feature maps that provides useful information for the segmentation task. Given $\mathbf{x}$ the input of the module, of size $W \times H \times D \times C$, we define $\tilde{\mathbf{x}} = [\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \dots, \tilde{\mathbf{x}}_C]$ (size: $W \times H \times D \times C$) the output of the module as

$$\mathbf{f}_{\mathrm{sq}} = \mathrm{ReLU}(\mathbf{W}_1^T \mathrm{GAP}(\mathbf{x})), \tag{1}$$

$$\mathbf{f}_{\mathrm{ex}} = \sigma(\mathbf{W}_2^T \mathbf{f}_{\mathrm{sq}}), \tag{2}$$

$$\forall\, c \in [1, C], \quad \tilde{\mathbf{x}}_c = (\mathbf{f}_{\mathrm{ex}})_c \cdot \mathbf{x}_c, \tag{3}$$

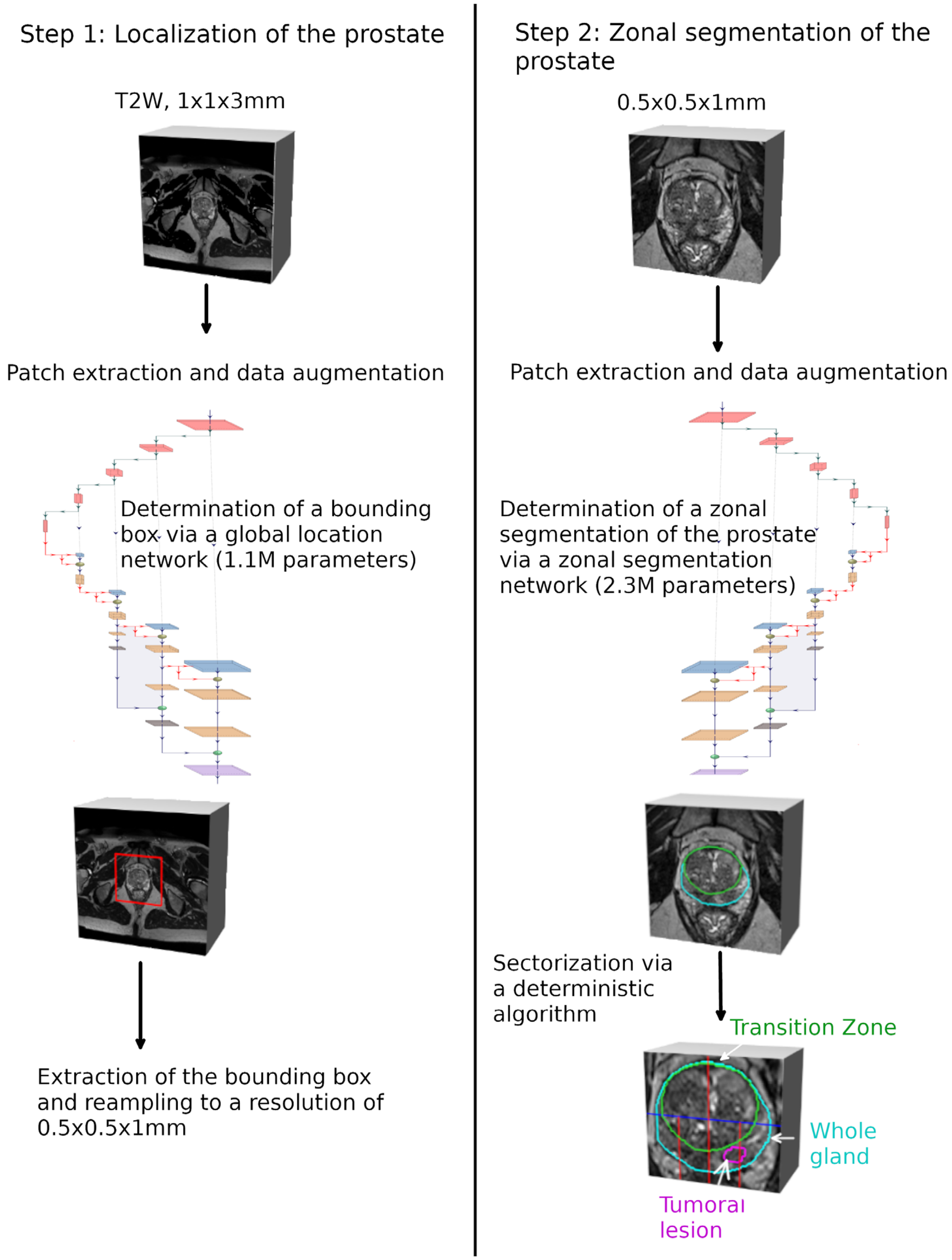


Fig. 5 Framework for the zonal segmentation of the prostate. The global location network extracts from a T2w MRI sequence a bounding box, which serves as input to the zonal segmentation network, generating the zonal segmentation of the WG (cyan), the TZ (green), and the PZ. Finally, a sector map is constructed from the zonal segmentation to provide information about the location of the lesion (magenta).

GAP() is a 3D global average pooling module, $\sigma()$ is the sigmoid function, and $\mathbf{W}_1$, $\mathbf{W}_2$ are convolutional kernels [see Fig. 6(b)]. Both $\mathbf{f}_{\mathrm{sq}}$ and $\mathbf{f}_{\mathrm{ex}}$ have a size of $1 \times 1 \times 1 \times C$. We used squeeze-and-excitation modules on the encoder part of our network.

2.3.2 *Attention gates*

Introduced in biomedical segmentation by Oktay et al.,[59] the principle of attention gates is to highlight in skip connections spatial zones with the more informative content. More precisely, if

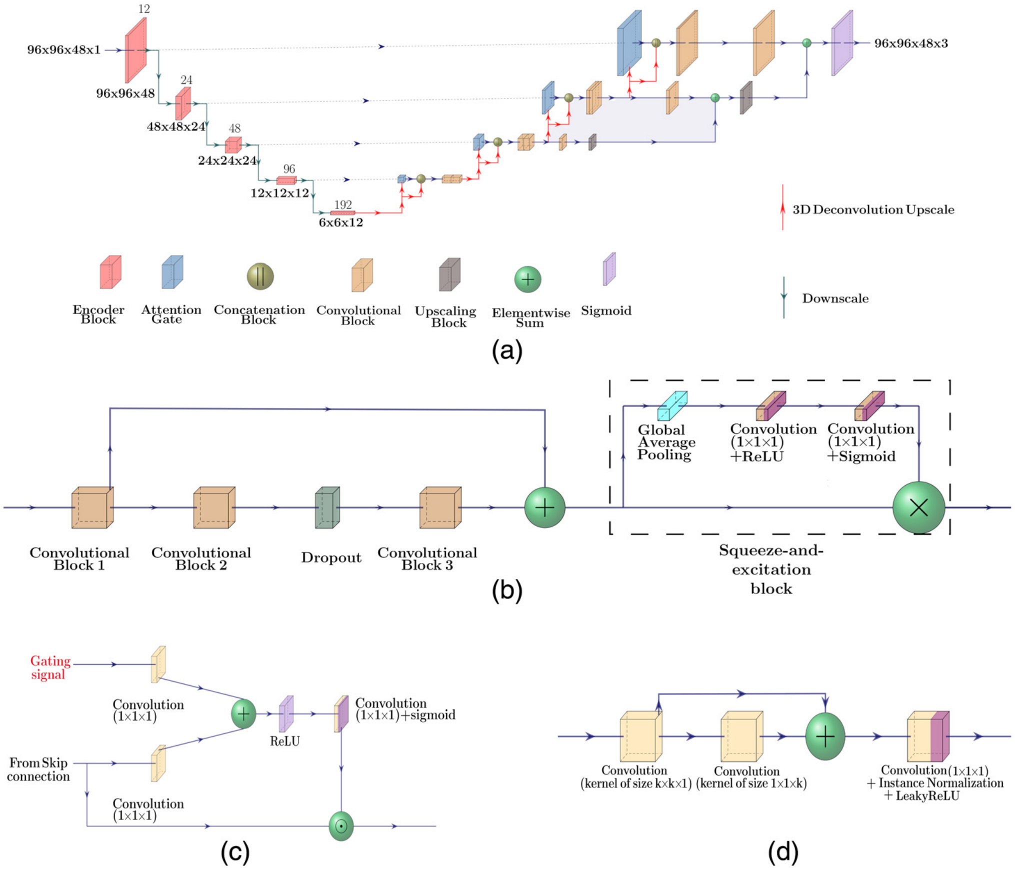


Fig. 6 (a) Architecture of the zonal segmentation network and its components: (b) the encoder block, (c) the attention gate, and (d) the used convolutional block. Values above layers in (a) correspond to the number of output filters in the convolutions performed in this layer. Architecture of global location network is similar but with a lower number of parameters.

$\mathbf{x}$ and $\mathbf{g}$ are the feature maps from skip connection and the gating signal with coarser information, respectively, we define

$$\mathbf{q}_{\text{att}} = \psi^T(\text{ReLU}(\mathbf{W}_g^T\mathbf{g} + \mathbf{W}_x^T\mathbf{x} + \mathbf{b}_g)) + \mathbf{b}_\psi, \tag{4}$$

$$\alpha = \sigma(\mathbf{q}_{\text{att}}); \quad \hat{\mathbf{x}} = \alpha \odot \mathbf{x}, \tag{5}$$

where $\mathbf{W}_g$ and $\mathbf{W}_x$ are the convolutional kernels, ψ is a $1 \times 1 \times 1$ convolutional kernel, $\mathbf{b}_\psi$ and $\mathbf{b}_g$ are biases and $\odot$ the Hadamard product. A visualization of the architecture is available in Fig. 6(c). We used attention gates on each layer of our network.

Combination of both attention methods has already been used for PCa detection[60,61] but to the best of our knowledge, it is the first time it is applied to prostate zonal segmentation.

2.4 *Loss Functions for the Zonal Segmentation Network*

We used as the main loss function the mean of 1-DSC for WG, for TZ, and for PZ.

Several approaches have been proposed in prior works to enforce the partition of the prostate such that $\text{WG} = \text{TZ} \cup \text{PZ}$ and $\text{TZ} \cap \text{PZ} = \emptyset$. First, one can only segment WG and TZ and build PZ by subtracting TZ to WG.[27] Second, one can learn to segment WG and both zones and rely on postprocessing to enforce the partition.[54] Another approach is to segment WG and to classify its voxels as either TZ or PZ.[36] In this paper, we propose another approach only based on segmentation, with partition loss functions dedicated to obtain a partition of the prostate. If we consider $p^{\text{WG}}, p^{\text{TZ}}, p^{\text{PZ}} \in [0,1]^N$ the probabilistic segmentation by the zonal segmentation network of WG, TZ, and PZ, respectively, then we define the two losses:

$$\mathcal{L}^{1}_{\text{aux}}(p^{\text{WG}}, p^{\text{TZ}}, p^{\text{PZ}}) = \frac{1}{N}\sum_{i=1}^{N}(p_i^{\text{WG}} - p_i^{\text{TZ}} - p_i^{\text{PZ}})^2, \tag{6}$$

and

$$\mathcal{L}^{2}_{\text{aux}}(p^{\text{WG}}, p^{\text{TZ}}, p^{\text{PZ}}) = \frac{\sum_{i=1}^{N}(p_i^{\text{TZ}}.p_i^{\text{PZ}})}{\left(\sum_{i=1}^{N} p_i^{\text{WG}}\right) + \varepsilon}. \tag{7}$$

The objective of $\mathcal{L}^{1}_{\text{aux}}$ is to ensure that the segmentation obtained by the network is coherent, i.e., that the segmentations for TZ and for PZ are within the limits of the segmentation of WG, and that they totally cover it. For this reason, we chose to penalize not only segmented voxels outside the WG segmentation but also the voxels that are included in both the TZ and PZ segmentations. The objective of $\mathcal{L}^{2}_{\text{aux}}$ is to enforce this lack of intersection.

2.5 *Sector Map Construction*

Another objective of our work is also to estimate the efficiency of our method to correctly assess the sectorial position of lesions. To this end, we designed an algorithm taking as input the zonal segmentation of a prostate and constructing the associated sector map. We chose to base our sector map on the 27 regions of interest sector map defined by Dickinson et al.[50] We defined the limits between sectors as follows:

- According to the longitudinal axis: we split the prostate on three equal-sized parts corresponding to the apex, to the midgland, and to the base, taking as extremes points the lowest and highest positions of WG segmentation. Following axes are computed separately for each third of the prostate.
- According to the anteroposterior axis: we split each part across its median sagittal slice x_{mid}.
- According to the transverse axis: for each part, we take the mean of the extreme positions of their slices (their leftest and rightest positions) according to the transverse axis x_{left} and x_{right}, and we define the positions of the inner subdivisions as: $x_{\text{midleft}} = 0.4(x_{\text{mid}} - x_{\text{left}}) + x_{\text{left}}$ and $x_{\text{midright}} = 0.6(x_{\text{right}} - x_{\text{mid}}) + x_{\text{mid}}$.

The main difference between our constructed sector map and the 27 regions of interest sector map[50] is the absence of the three sectors related to the anterior fibromuscular stroma, as we did not segment this particular zone but included it into TZ. For this reason, we included lesions located in the anterior fibromuscular stroma among the TZ lesions. An illustration is shown in Fig. 7

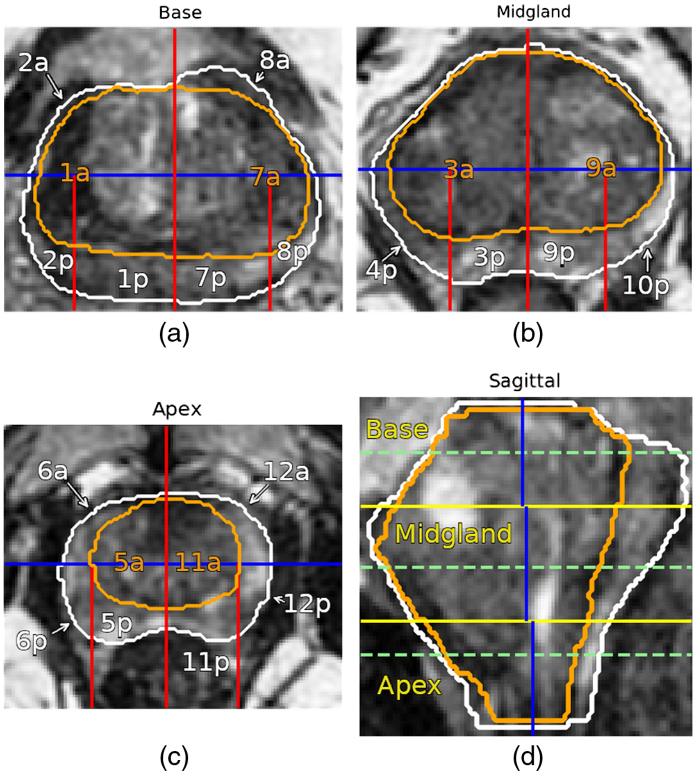
The zonal location of a lesion in the prostate is defined as the zone (PZ or TZ) with the highest proportion of lesions' voxels, and the sectorial position of the lesion as the sector within the considered zone with the highest proportion of lesion's voxels.

3 Experimental Design

3.1 *Training of the Network*

Training on the private dataset was performed using an Intel(R) Xeon(R) Gold 6246R CPU and a NVIDIA Tesla V100 SXM2 32 GB GPU, a NVIDIA Tesla T4 16 GB GPU having also been used for the ProstateX dataset. We used Keras[62] and Tensor Flow 2.4.1[63] as a deep learning framework.

The training of the global location network has been done with a batch size of 4, using RMSProp as a gradient optimizer with an initial learning rate of 0.0005. The training of the zonal segmentation network has been done with a batch size of 8, using Adam as a gradient optimizer with an initial learning rate of 0.0005, and attributing loss weights of 1, 0.1, and 0.01, respectively, to $\mathcal{L}$, $\mathcal{L}^{1}_{\text{aux}}$ and $\mathcal{L}^{2}_{\text{aux}}$. Several values of parameters (learning rate, batch size…) were tested before choosing those values.



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

Fig. 7 Sector map of the prostate according to axial views in (a) the base, (b) the midgland, and (c) the apex; (d) the sagittal view. White: WG; orange: TZ. The blue axis separates the anterior from the posterior of the prostate, the red axes are left-right based separations and the yellow axes represent the separations between (a) the base, (b) the midgland, and (c) the apex. The green dashed lines on the sagittal view indicate the location of the different axial views.

For both networks, we adopted a maximal number of epochs of 500, with a policy of early stopping if the validation loss did not improve for 70 epochs. We also adopted a policy of reduction of the learning rate with a multiplication by 0.2 in case of stagnation of the validation loss for 30 epochs. The dropout rate was set to 0.3.

To improve the performance of the network, we artificially increased the number of images used during the training due to data augmentation through the Python module batchgenerators.[64] This module allowed us to apply several transformations such as rotation according to the longitudinal axis, mirror transform along the anteroposterior axis, elastic transform or intensity transforms such as gamma transform. The intensity of each sequence has also been normalized via the subtraction of its voxels' mean value and a division by their standard deviation.

3.2 *Test and Postprocessing*

We used test-time augmentation,[65,66] which consists of applying different transformations to an image during the test procedure, using the network on each of these transformed images and then reverting those transforms and combining the obtained results into one final prediction by taking their mean, to improve the final segmentation and the robustness of the process. Transformations applied were all combinations of flipping along the anteroposterior axis with a rotation of ±10 deg, for a total of six images.

In postprocessing, to apply our method on the whole image, we applied a sliding window strategy, where patches of size $96 \times 96 \times 48$ were extracted with steps of (24, 24, 12) voxels according to each dimension and where the contribution of each patch to a specific voxel is divided by the number of patches contributing to this voxel. Finally, after reconstruction of the segmentation from the patches, we applied a threshold of 0.5 to obtain WG segmentation that we restrained to its largest connected component according to the longitudinal direction. Within WG segmentation, we defined TZ segmentation as the voxels for which the probability to be in TZ was higher than the probability to be in PZ, and conversely for PZ.

4 Results

The main metrics we used to estimate the performance of our network were DSC and 95% Hausdorff distance (HD95%). We mainly compared three networks for the zonal segmentation network: a 3D UNet as presented by Isensee et al.,[67] introducing context and localization modules and which served as a basis for other networks, the network UNetV2 with elements presented in Sec. 2.2 to take into account the image anisotropy and UFNet which adds deep supervision, attention modules, and the partition losses. These three networks have 2.15M, 2.28M, and 2.32M parameters, respectively. A UFNet with 1.03M parameters was used as the global location network.

To improve the segmentations and their stability, inspired by Isensee et al.,[25] we combined the results of five neural networks obtained through cross-validation to provide a final, more precise ensembled segmentation by taking the mean of their prediction before postprocessing, including test-time augmentation. These networks are named with the suffix -E.

Table 1 Comparison between our method and UNet on our private dataset after correction.

Model	Zone	DSC (in %)	*p*-value/UNet-E	HD95% (in mm)	*p*-value/UNet-E
UFNet-E (5 × 2.32M parameters)	WG	**93.90 ± 2.85**	***	**2.59 ± 0.98**	**
	TZ	**91.00 ± 4.34**	***	**3.20 ± 0.90**	***
	PZ	**79.08 ± 7.08**	***	**3.87 ± 1.65**	*
UNetV2-E (5 × 2.28M parameters)	WG	93.65 ± 2.46	*	2.73 ± 0.80	*
	TZ	90.73 ± 4.09	*	3.42 ± 0.94	*
	PZ	78.11 ± 7.58	*	4.26 ± 2.36	>.05
UNet-E (5 × 2.15M parameters)	WG	93.48 ± 2.54	—	2.81 ± 0.80	—
	TZ	90.47 ± 4.28	—	3.53 ± 0.97	—
	PZ	77.51 ± 7.95	—	4.09 ± 1.93	—
Model	Zone	DSC (in %)	*p*-value/UNet	HD95% (in mm)	*p*-value/UNet
UFNet	WG	93.45 ± 2.89	***	2.81 ± 0.93	**
	TZ	90.45 4.51	***	3.41 ± 0.93	***
	PZ	77.80 ± 7.79	***	4.15 ± 1.92	*
UNetV2	WG	93.16 ± 2.51	>0.05	2.91 ± 0.80	>.05
	TZ	90.24 ± 4.16	*	3.52 ± 0.93	>0.05
	PZ	76.52 ± 8.03	*	4.37 ± 2.09	>0.05
UNet	WG	92.87 ± 2.79	—	3.21 ± 1.32	—
	TZ	89.85 ± 4.56	—	3.83 ± 1.17	—
	PZ	75.95 ± 8.47	—	4.38 ± 1.91	—

Note: Top: Results for the ensembling of five networks of three different types, obtained from cross-validation. Best results for each considered metric are in bold. Bottom: Mean of the results from the five networks used in the ensemble version. Signed-rank Wilcoxon test with Bonferroni–Holm correction has been used to assess statistically significant differences and to compute *p*-values for ensembled networks and for mean of networks on each fold. Significant differences are indicated.

**p*-value ≤ 0.05.

***p*-value ≤ 0.01.

****p*-value ≤ 0.001.

4.1 *Results for Private Dataset*

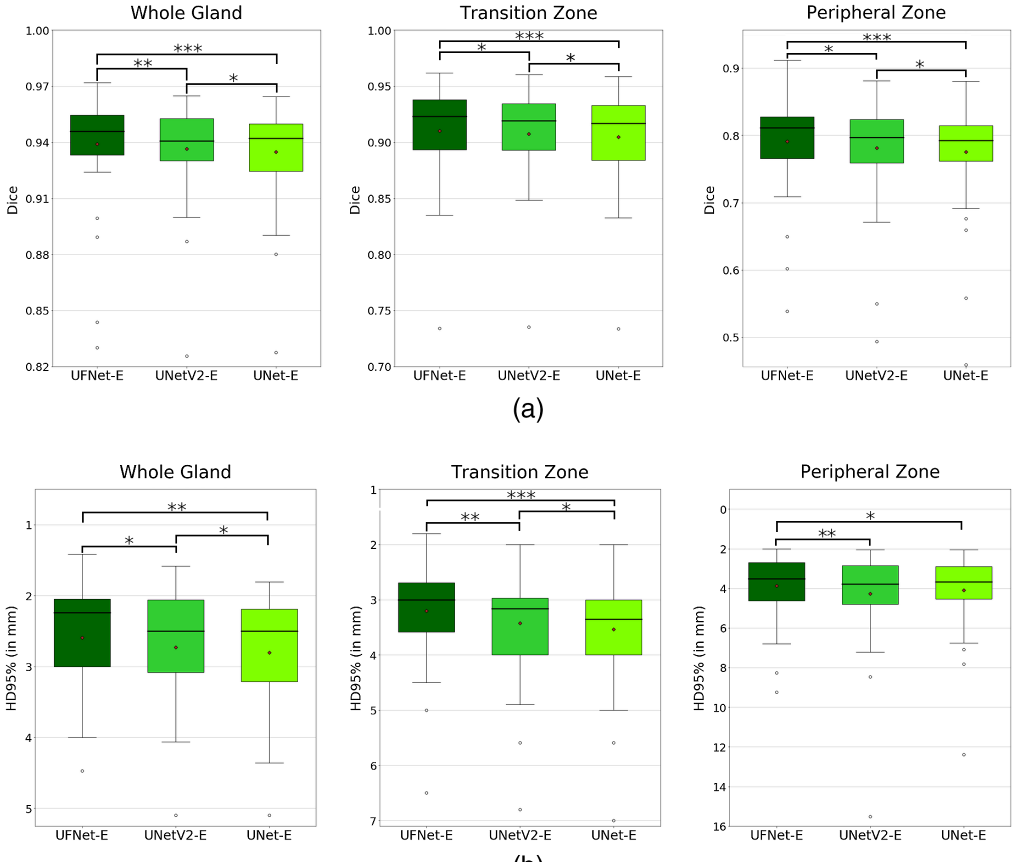
For this dataset, we compared the outputs of the networks to the consensus obtained from the seven radiologists and corrected as described in Sec. 2.1.2 (i.e., slicewise restriction of PZ to its largest components). The same correction was applied to the outputs of the network.

4.1.1 *Segmentation results*

Results obtained for the metrics on WG, TZ, and PZ are given in Table 1. They are shown in Fig. 8, along with statistical differences between the performance of the networks. The global location network provided an adequate bounding box of the prostate, i.e., which surrounds the prostate without crossing it, for all images. Some examples of zonal segmentation are provided in Figs. 9 and 10. The mean time to process one patch was 0.8 s, and the mean time to process a sequence on this dataset was 4.5 s.

4.1.2 *Comparison with radiologists*

We compared these results to the segmentations provided by the seven radiologists on the same dataset, as shown in Fig. 11. While not being as good as the best radiologists, i.e., those with



(a)

(b)

Fig. 8 Metrics between network segmentation and consensus segmentation on private dataset for all three networks. (a) Dice for WG, TZ, and PZ.(b) Hausdorff distance for WG, TZ, and PZ. Black line and red point are median and mean, respectively. Signed-rank Wilcoxon test with Bonferroni–Holm correction has been used to assess statistically significant differences and to compute *p*-values. Significant differences are indicated (* : p-value ≤ 0.05; **: p-value ≤ 0.01; ***: p-value ≤ 0.001).

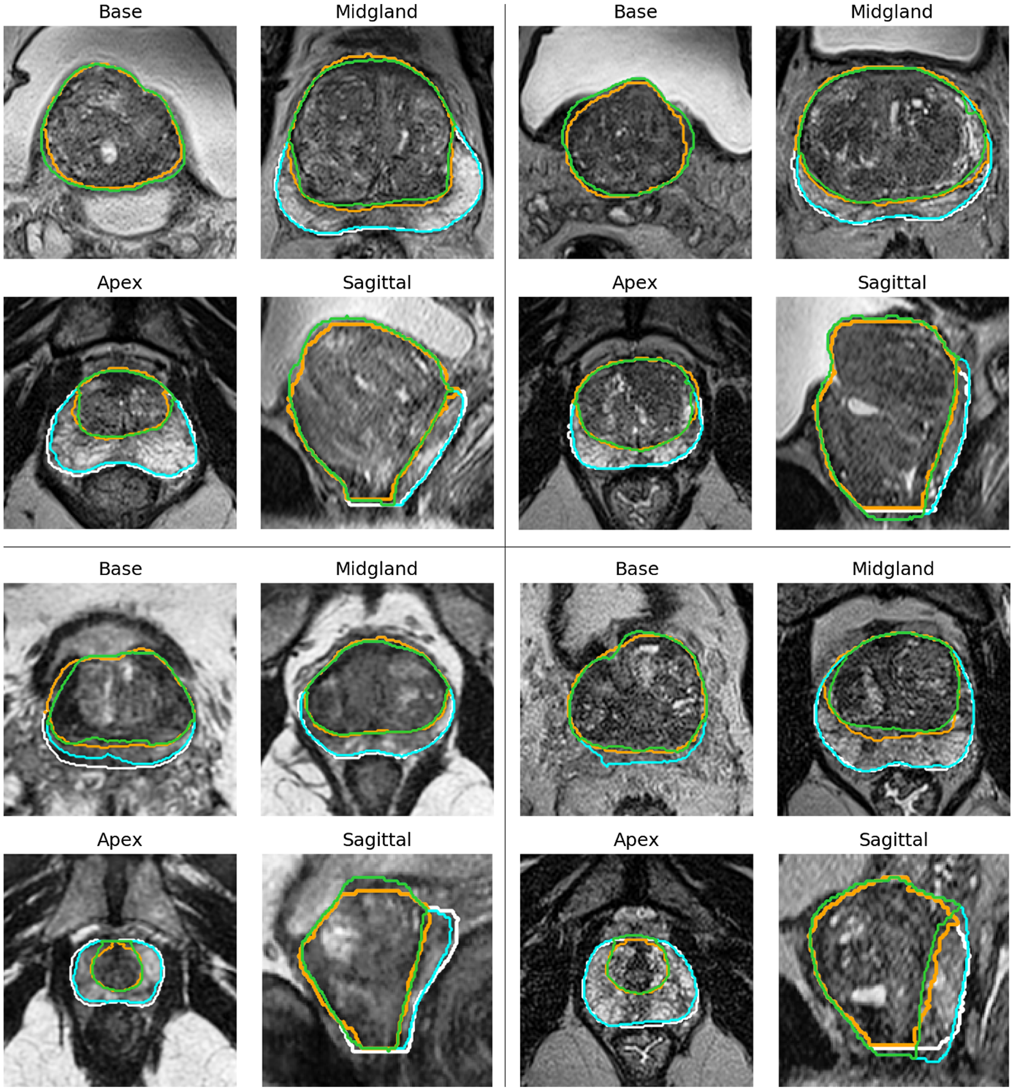


Fig. 9 Good segmentations on the private dataset, with axial views of (top left) the base, (top right) the midgland, and (bottom left) the apex, and (bottom right) sagittal view. White: ground truth WG; orange: ground truth TZ; cyan: network-segmented WG; green: network-segmented TZ.

the closest segmentations to the consensus, UFNet-E obtained results similar to radiologists'. If we rank UFNet-E and all radiologists according to how close they are to the consensus for all metrics according to their means, UFNet-E is ranked between the third place and the seventh place, with a global fifth place.

4.2 *Lesion Positions*

4.2.1 *Test set*

On the 17 prostates of the test set with a lesion, we applied our sector map construction algorithm and determined the location of the lesion using the sector map derived from the ground truth as the true sector map. On the zonal location of the lesion, we obtained a 100% accuracy, whereas on the sectorial position we obtained an accuracy of 88% (15 out 17 cases). Observed lesion position errors are due to differences on the delineations of the apex, midgland, and base between the ground truth segmentation and the segmentation from the network. Examples of lesion locations are shown in Fig. 12.

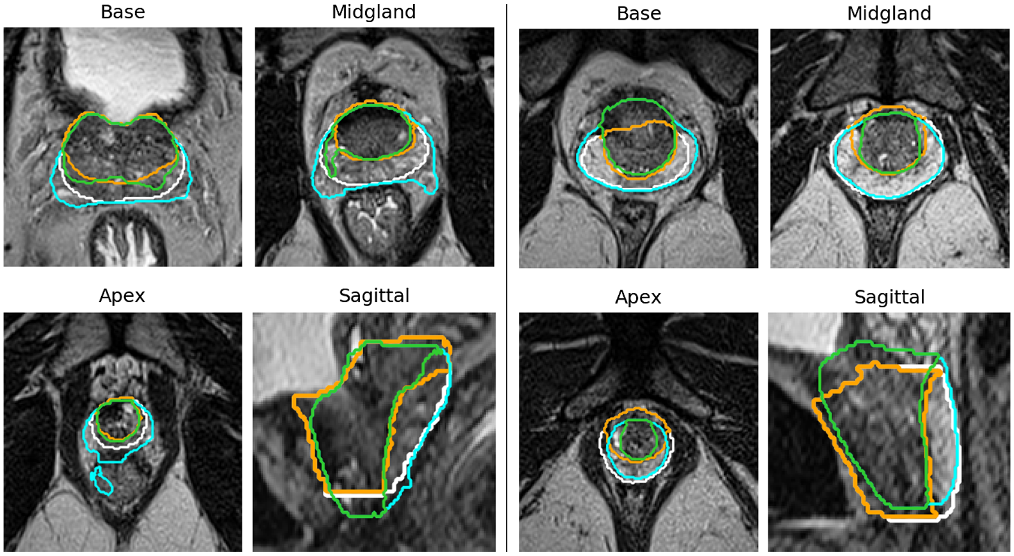
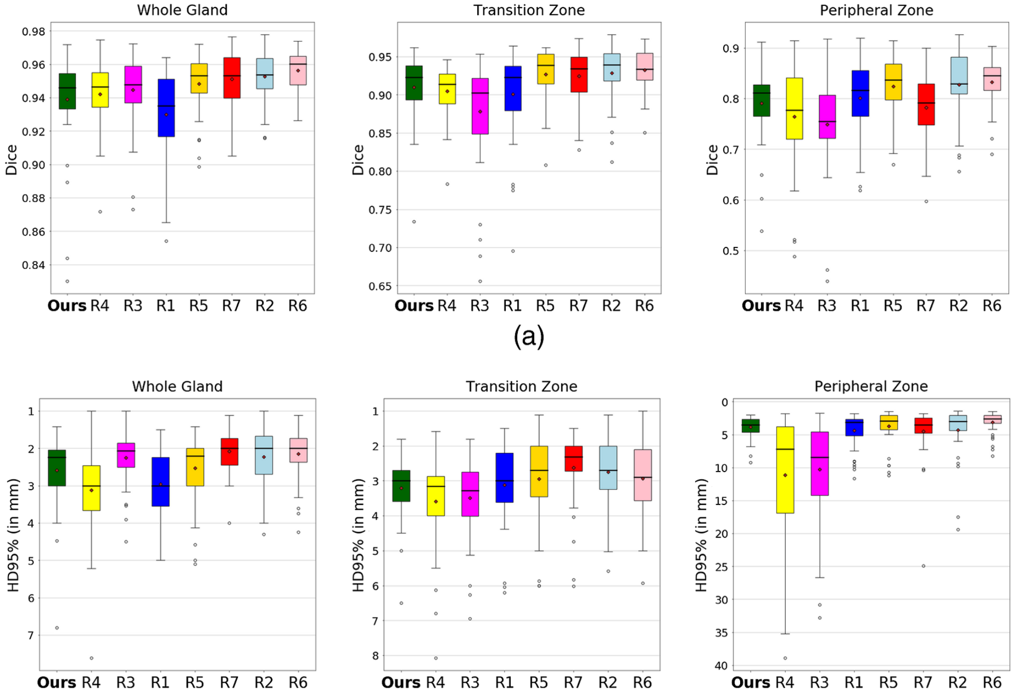


Fig. 10 Poor segmentations on the private dataset, with axial views of (top left) the base, (top right) the midgland and (bottom left) the apex, and (bottom right) sagittal view. White: ground truth WG; orange: ground truth TZ; cyan: network-segmented WG; green: network-segmented TZ.



(a)

(b)

Fig. 11 Metrics between UFNet-E (ours) segmentation and the consensus segmentation on private dataset, side to side with the metrics for raters' segmentation ranked in increasing order of performance (the rater *k* being written as Rk). (a) Dice for WG, TZ and PZ. (b) 95% Hausdorff distance for WG, TZ, and PZ. Black line and red point are median and mean values, respectively.

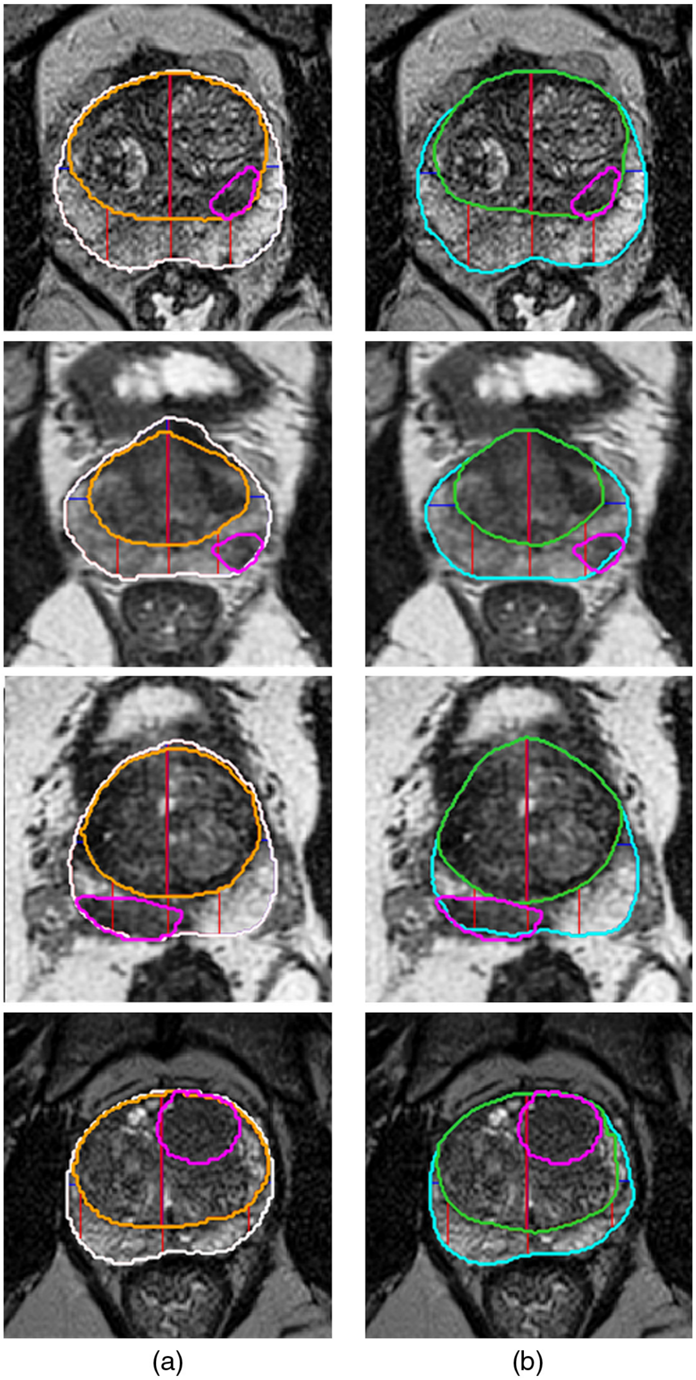


Fig. 12 Examples of correct lesion placement on the test set. (a) Placement derived from the true segmentation (white: WG; orange: TZ). (b) Placement on the sector map computed from the network segmentation (cyan: WG; green: TZ). Separations between sectors are in red (anteroposterior direction) and in blue (transverse direction). Lesions are in magenta (consensus segmentation from five radiologists).

4.2.2 *Private lesion set*

On the private lesion set, we obtained an accuracy for the zonal location of lesions of 91% (42 out of 46) and of 74% for their sectorial positions (34 out of 46).

All four cases where the lesion has been placed in the wrong zone are cases with a configuration similar to the configuration presented in Fig. 13(b), i.e., with a lesion in the anterior part of PZ near the border with TZ and with a voxel intensity closer to TZ's than to PZ's. Cases with a

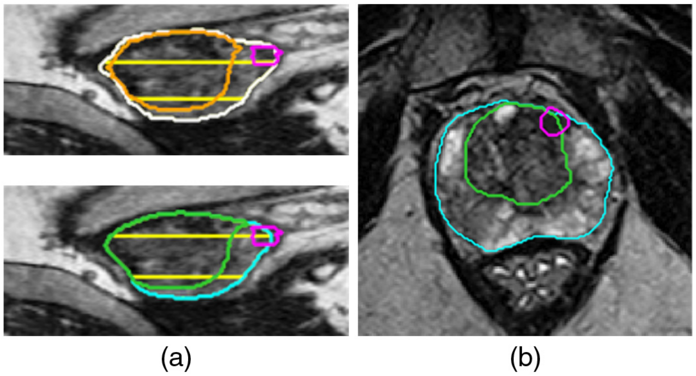


Fig. 13 (a) Example of wrong lesion placement on the test set. Top: placement derived from the true segmentation (white: WG; orange: TZ), in the base. Bottom: placement on the sector map computed from the network segmentation, in the midgland (cyan: WG; green: TZ). Limits between base/midgland/apex are in yellow, the lesion is in magenta.(b) Example from the private lesion set with the lesion (in magenta) located in TZ by the network whereas the radiologists located it in PZ.

sectorial position error correspond to either lesions located according to our algorithm across two adjacent sectors—including the true sectorial position—but with a majority of voxels in the wrong sector (four cases), or a lesion well located according to transverse and anteroposterior directions but not according to longitudinal direction, for example, lesions located in the midgland whereas the true position is in the base (four cases).

4.3 *Results on ProstateX*

4.3.1 *Segmentation results*

Results obtained on ProstateX are given on Table 2 and are shown in Fig. 16. A boxplot graph illustrating the performances according to the different metrics is available in Fig. 14. The global location network provided an adequate bounding box of the prostate for all images. The mean inference time of the network for each patch was 0.09 s, adding up to 3.5 s on the whole sequence.

4.3.2 *Evolution of Dice score during training*

We also estimated on ProstateX the impact of the different architectures on the evolution of DSC during the training. These results are shown in Fig. 15. It appeared that taking into account the anisotropy does not have an important impact on the course of the training, contrary to the introduction of attention modules which help to speed up the training.

5 Discussion

In this work, we used a framework composed of two successive neural networks to generate a zonal segmentation of the prostate on both 2D and 3D T2w MRI sequences. The global location network has been able to provide an adapted bounding box of the prostate, and the zonal segmentation network has been able to generate an accurate zonal segmentation of the prostate. This framework not only allows us to detect the prostate despite its relatively small size (around 5% of the image on the private dataset) but also to take advantage of the better resolution on 3D T2w MRI sequences while only requiring a reasonable amount of memory.

The results for the segmentation of the different zones of the prostate are comparable with the results of the state of the art, as for example, Bardis et al.[36] obtained a mean Dice of 94.0% for the whole prostate, 91.4% for the TZ, and 77.6% for the OZ. On ProstateX, we obtained DSC similar or superior to Cuocolo et al.[29] with segmentations from the same dataset, showing the efficiency of our method. Those differences could be explained by the use of 3D networks (compared to 2D

Table 2 Comparison between our method and UNet on the ProstateX dataset.

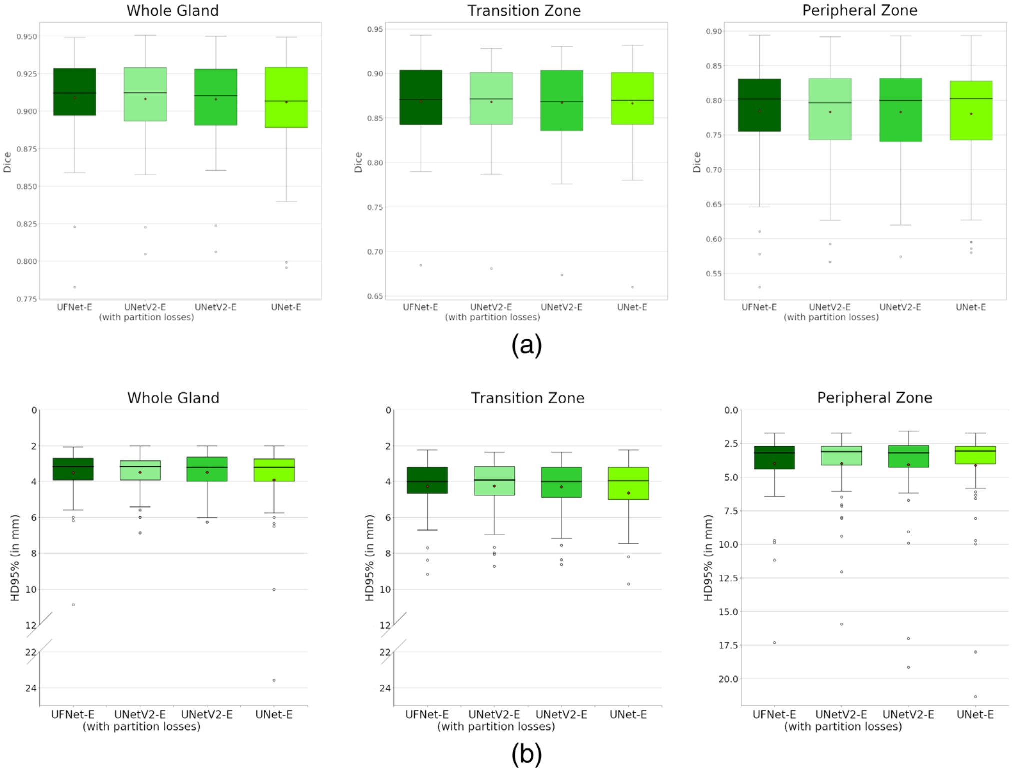
Model	Zone	DSC (in %)	HD95% (in mm)
UFNet-E	WG	**90.90 ± 2.94**	3.50 ± 1.36
	TZ	**86.84 ± 4.33**	4.27 ± 1.40
	PZ	**78.40 ± 7.31**	**4.00 ± 2.54**
UNetV2-E (with partition losses)	WG	90.83 ± 2.81	**3.48 ± 1.12**
	TZ	86.82 ± 4.53	**4.25 ± 1.42**
	PZ	78.29 ± 7.14	4.01 ± 2.48
UNetV2-E (without partition losses)	WG	90.81 ± 2.82	3.48 ± 1.11
	TZ	86.73 ± 4.31	4.29 ± 1.43
	PZ	78.31 ± 7.12	4.08 ± 3.04
UNet-E	WG	90.59 ± 3.09	3.91 ± 2.89
	TZ	86.66 ± 4.56	4.64 ± 3.02
	PZ	78.04 ± 7.60	4.14 ± 3.34
UFNet	WG	90.62 ± 2.92	3.58 ± 1.15
	TZ	86.45 ±4.45	4.38 ± 1.41
	PZ	77.81 ± 7.35	4.06 ± 2.38
UNetV2 (with partition losses)	WG	90.48 ± 2.86	3.59 ± 1.11
	TZ	86.38 ± 4.39	4.37 ± 1.37
	PZ	77.45 ± 7.42	4.26 ± 3.46
UNetV2 (without partition losses)	WG	90.47 ± 2.88	3.66 ± 1.19
	TZ	86.31 ± 4.58	4.44 ± 1.44
	PZ	77.53 ± 7.54	4.32 ± 3.94
UNet	WG	90.25 ± 3.04	3.86 ± 1.93
	TZ	86.19 ± 4.58	4.60 ± 2.10
	PZ	77.27 ± 7.75	4.24 ± 3.26

Top: The -E signals the use of an ensemble of five networks. Best results for each considered metric are in bold. No statistical differences were found between the ensembles of networks.
Bottom: Mean of the results from the five networks used in the ensemble version.

used by Cuocolo et al.),[29] of attention modules, and by the difference in training set sizes and number of training epochs.

The use of attention modules and anisotropy-adapted modules appeared to have a positive impact, with fewer epochs for training with attention modules and overall better results for both modifications, especially on 3D T2w sequences with p-values < 0.05 between the classic UNet, the version with the anisotropy modules and the version with both additional modules on almost all considered metrics. On the other hand, the use of partition losses seemed to have a positive but moderate impact on the Hausdorff Distance and on TZ segmentation, whereas slightly deteriorated results for PZ.

The apex and the base of the prostate are the most difficult regions to segment, which is in accordance with previous publications, involving human[14,15] or automatic raters[28,43] (see Figs. 9, 10, and 16). The reasons behind this complexity include, but are not limited to, the heterogeneity of the tissues on these regions, the possible ambiguities of the borders, or exotic shapes of both WG and TZ at their top and bottom slices.



(a)

(b)

Fig. 14 Metrics between network segmentation and ground truth segmentation on ProstateX dataset for all three networks. (a) Dice for WG, TZ, and PZ. (b) 95% Hausdorff distance for WG, TZ, and PZ. Black line and red point are median and mean values, respectively. Signed-rank Wilcoxon test with Bonferroni–Holm correction has been used to assess statistically significant differences, but no differences were found.

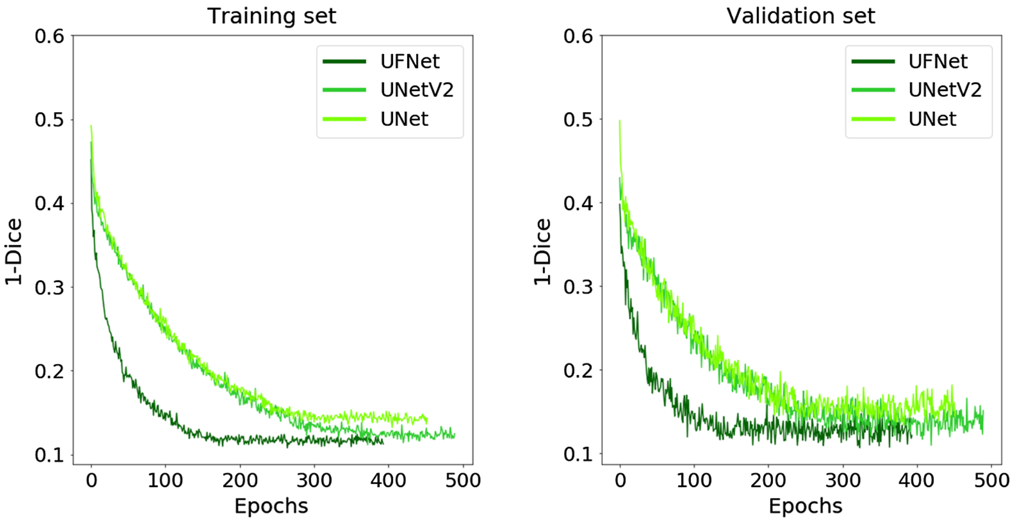
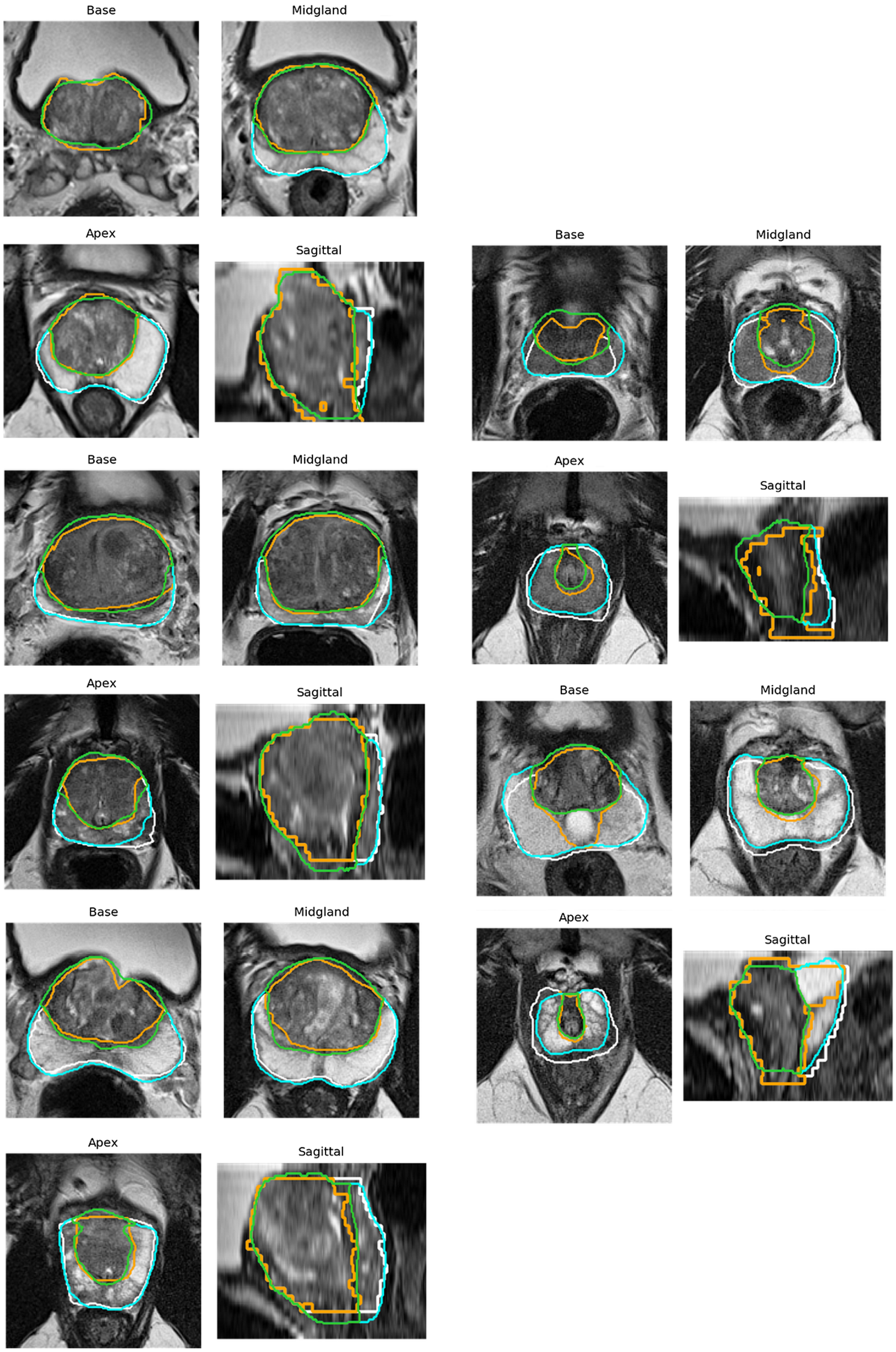


Fig. 15 Evolution of the Dice coefficient on both training and validation sets during training (on onefold of ProstateX).

Results obtained by the network appeared to be on the same level of performance than those of the radiologists, who can be separated in three categories: the experts (raters 3, 6, and 7), the seniors (raters 1 and 2), and the juniors (raters 4 and 5). The network obtained results comparable to those of radiologists in the middle of the pack and significantly better results than the rater 3 (an expert radiologist) for both TZ and PZ (p-value < 0.05).



(a) (b)

Fig. 16 Segmentations on the 2D dataset ProstateX, with axial views of (top left) the base, (top right) the midgland, and (bottom left) the apex, and (bottom right) sagittal view. (a) Good segmentations and (b) poor segmentations. White: ground truth WG; orange: ground truth TZ; cyan: network-segmented WG; green: network-segmented TZ.

We have observed on the private lesion dataset that in more than 90% of the cases, the automatic segmentation preserved the zone in which the lesions were located. The four cases of zonal location errors correspond to a very specific situation that was not present in the training-validation set. Also, we compare favorably with the inter-rater variability on sectorial positions observed by Greer et al.[12] with 74% of agreement between the radiologists and our method, even though we rely on a slightly simpler sector map. Moreover, the observed errors on sectorial positions would only have a minor clinical impact, as the global position of the lesion (left/right and anterior/posterior) has been preserved in each case. It may also be more appropriate for clinical use to provide the top two sectors on which each lesion lies, since in almost all cases the true sector is among those two. Those results validate the use of neural network segmentation as a second reader for automatic PCa diagnosis to detect in which zone the suspected lesion is located. This is important as it determines which sequence to use to determine the PI-RADS score, as the importance of the different sequences depends on which zone the suspected lesion is located in.

Our study faces several limitations. First the quality of the ProstateX segmentations is uneven, as shown in Fig. 16. Nevertheless, we chose to work with this annotation set for the sake of comparing ourselves with other methods. On the private dataset, we had to correct PZ masks computed as WG-TZ with an ad hoc method to avoid the occurrence of thin isolated lines or voxels (see Fig. 4). Indeed, without this processing, PZ-associated metrics and especially PZ HD95% were affected. However, the impact was the same on the three compared networks, and it did not modify their relative performances. This processing was not required on the ProstateX segmentation.

In addition, the algorithm to create the prostate sector map is based on debatable hypotheses. In particular, boundaries were determined partwise and not slicewise, and we arbitrarily defined inner anteroposterior boundaries positions since there exists no formal definition to define them. Those choices have a direct impact on sectorial positions, since peripheral sectors on the apex (especially sectors 6p and 12p) can have a very small area with our method. Moreover, the sector map used in this study, inspired by the 27 region of interest sector map defined by Dickinson et al.,[50] does not correspond to the PI-RADS 2.1 standard sector map[5] which is based on 39 regions of interest. Nevertheless, in practice, the differences between the two sector maps have little to no practical effects.

In Fig. 10, we can see that some of the segmentations may have very unusual shapes, with for instance “outgrowths” on the generated segmentation or WG split into two parts. To enforce coherent shapes of WG and TZ, it may be possible to project the prostate on a restricted learned shape space[68,69] to correct aberrant segmentations.

6 Conclusion

We proposed a deep learning-based method for the zonal segmentation of the prostate from T2w MRI sequences, taking into account the anisotropy of these sequences, with attention modules and enforcing the partition of the prostate. This method can be applied on both 2D and 3D T2w MRI sequences. The obtained results not only are similar to the results from the state of the art but are also coherent with the results obtained by radiologists and globally preserving zonal locations and sectorial positions of the lesions, making our method suitable as a first step tool for an automated system dedicated to diagnosis and grading of PCa, as done by Hosseinzadeh et al.[70] or by Mehta et al.[71]

Disclosures

The authors have no conflicts of interest to declare.

Acknowledgments

We thank Julien Castelneau, software engineer at Inria, for his help in the development of MedInria Software (MedInria—Medical image visualization and processing software by

Inria[44]). The authors are grateful to the OPAL infrastructure from the Université Côte d'Azur for providing resources and support. We also thank Alexandre Allera, Malek Ezziane, Anna Luzurier, Raphaelle Quint, and Mehdi Kalai for providing prostate segmentations. This work has been supported by the French government, through the 3IA Côte d'Azur and UCA DS4H Investments in the Future project managed by the National Research Agency (ANR) with the reference numbers ANR-19-P3IA-0002 and ANR-17-EURE-0004, and by the Health Data Center of the AP-HP (Assistance Publique-Hôpitaux de Paris). Private data were extracted from the Clinical Data Warehouse of the Greater Paris University Hospitals (Assistance Publique-Hôpitaux de Paris).

References

1. R. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2020," *CA Cancer J. Clin.* **70**(1), 7–30 (2020).
2. A. Stabile et al., "Multiparametric MRI for prostate cancer diagnosis: current status and future directions," *Nat. Rev. Urol.* **17**, 41–61 (2019).
3. N. Mottet et al., "EAU-EANM-ESTRO-ESUR-SIOG guidelines on prostate cancer—2020 update. Part 1: screening, diagnosis, and local treatment with curative intent," *Eur. Urol.* **79**(2), 243–262 (2021).
4. 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.* **203**(4), 706–712 (2020).
5. B. Turkbey et al., "Prostate imaging reporting and data system version 2.1: 2019 update of prostate imaging reporting and data system version 2," *Eur. Urol.* **76**(3), 340–351 (2019).
6. V. Kasivisvanathan et al., "MRI-targeted or standard biopsy for prostate-cancer diagnosis," *New Engl. J. Med.* **378**(19), 1767–1777 (2018).
7. O. Rouvière et al., "Use of prostate systematic and targeted biopsy on the basis of multiparametric MRI in biopsy-naive patients (MRI-FIRST): a prospective, multicentre, paired diagnostic study," *Lancet Oncol.* **20**(1), 100–109 (2019).
8. V. Kasivisvanathan et al., "Magnetic resonance imaging-targeted biopsy versus systematic biopsy in the detection of prostate cancer: a systematic review and meta-analysis," *Eur. Urol.* **76**(3), 284–303 (2019).
9. F. Elkhoury et al., "Comparison of targeted vs systematic prostate biopsy in men who are biopsy naive: the prospective assessment of image registration in the diagnosis of prostate cancer (PAIREDCAP) study," *JAMA Surg.* **154**(9), 811–818 (2019).
10. A. C. Westphalen et al., "Variability of the positive predictive value of PI-RADS for prostate MRI across 26 centers: experience of the society of abdominal radiology prostate Cancer disease-focused panel," *Radiology* **296**(1), 76–84 (2020).
11. M. Gatti et al., "Prostate cancer detection with biparametric magnetic resonance imaging (bpMRI) by readers with different experience: performance and comparison with multiparametric (mpMRI)," *Abdom. Radiol.* **44**(5), 1883–1893 (2019).
12. M. D. Greer et al., "All over the map: an interobserver agreement study of tumor location based on the PI-RADSv2 sector map," *J. Magn. Reson. Imaging* **48**(2), 482–490 (2018).
13. S. Sara Mahdavi et al., "Semi-automatic segmentation for prostate interventions," *Med. Image Anal.* **15**(2), 226–237 (2011).
14. A. Becker et al., "Variability of manual segmentation of the prostate in axial T2-weighted MRI: a multi-reader study," *Eur. J. Radiol.* **121**, 108716 (2019).
15. S. Montagne et al., "Challenge of prostate MRI segmentation on T2-weighted images: inter-observer variability and impact of prostate morphology," *Insights Imaging* **12**, 71 (2021).
16. G. Litjens et al., "Evaluation of prostate segmentation algorithms for MRI: the PROMISE12 challenge," *Med. Image Anal.* **18**(2), 359–373 (2014).
17. G. Vincent, G. Guillard, and M. Bowes, "Fully automatic segmentation of the prostate using active appearance models," in *MICCAI Grand Challenge: Prostate MR Image Segmentation 2012* (2012).

18. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Adv. Neural Inf. Process. Syst.*, F. Pereira et al., Eds., Curran Associates, Inc., Vol. 25 (2012).
19. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," *Lect. Notes Comput. Sci.* **9351**, 234–241 (2015).
20. F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: fully convolutional neural networks for volumetric medical image segmentation," in *Fourth Int. Conf. 3D Vision (3DV)*, pp. 565–571 (2016).
21. R. Cheng et al., "Automatic magnetic resonance prostate segmentation by deep learning with holistically nested networks," *J. Med. Imaging* **4**, 041302 (2017).
22. S. Xie and Z. Tu, "Holistically-nested edge detection," in *IEEE Int. Conf. Comput. Vision (ICCV)*, pp. 1395–1403 (2015).
23. J. Weickert, "Coherence-enhancing diffusion filtering," *Int. J. Comput. Vision* **31**(2), 111–127 (1999).
24. Z. Tian et al., "PSNet: prostate segmentation on MRI based on a convolutional neural network," *J. Med. Imaging* **5**, 021208 (2018).
25. F. Isensee et al., "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation," *Nat. Methods* **18**, 203–211 (2021).
26. H. Jia et al., "HD-Net: hybrid discriminative network for prostate segmentation in MR images," *Lect. Notes Comput. Sci.* **11765**, 110–118 (2019).
27. L. Rundo et al., "USE-Net: incorporating squeeze-and-excitation blocks into U-Net for prostate zonal segmentation of multi-institutional MRI datasets," *Neurocomputing* **365**, 31–43 (2019).
28. N. Aldoj et al., "Automatic prostate and prostate zones segmentation of magnetic resonance images using DenseNet-like U-net," *Sci. Rep.* **10**, 14315 (2020).
29. R. Cuocolo et al., "Deep learning whole-gland and zonal prostate segmentation on a public MRI dataset," *J. Magn. Reson. Imaging* **54**(2), 452–459 (2021).
30. A. Paszke et al., "ENet: a deep neural network architecture for real-time semantic segmentation, arXiv:1606.02147 (2016).
31. E. Romera et al., "ERFNet: efficient residual factorized ConvNet for real-time semantic segmentation," *IEEE Trans. Intell. Transp. Syst.* **19**(1), 263–272 (2018).
32. G. Litjens et al., "ProstateX challenge data," The Cancer Imaging Archive (TCIA), 2017, https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=23691656.
33. G. Litjens et al., "Computer-aided detection of prostate cancer in MRI," *IEEE Trans. Med. Imaging* **33**, 1083–1092 (2014).
34. K. Clark et al., "The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository," *J. Digital Imaging* **26**(6), 1045–1057 (2013).
35. S. G. Armato et al., "PROSTATEx challenges for computerized classification of prostate lesions from multiparametric magnetic resonance image," *J. Med. Imaging* **5**, 044501 (2018).
36. M. Bardis et al., "Segmentation of the prostate transition zone and peripheral zone on MR images with deep learning," *Radiol. Imaging Cancer* **3**(3) (2021).
37. A. Meyer et al., "Towards patient-individual PI-Rads v2 sector map: CNN for automatic segmentation of prostatic zones from T2-weighted MRI," in *IEEE 16th Int. Symp. Biomed. Imaging (ISBI 2019)*, pp. 696–700 (2019).
38. O. Zavala-Romero et al., "Segmentation of prostate and prostate zones using deep learning: a multi-MRI vendor analysis," *Strahlenther Onkol.* **196**, 932–942 (2020).
39. A. Meyer et al., "Automatic high resolution segmentation of the prostate from multi-planar MRI," in *IEEE 15th Int. Symp. Biomed. Imaging (ISBI 2018)*, IEEE, pp. 177–181 (2018).
40. A. Rosenkrantz et al., "Prostate cancer: comparison of 3D T2-weighted with conventional 2D T2-weighted imaging for image quality and tumor detection," *Am. J. Roentgenol.* **194**(2), 446–452 (2010).
41. S. Polanec et al., "3D T2-weighted imaging to shorten multiparametric prostate MRI protocols," *Eur. Radiol.* **28**(4), 1634–1641 (2017).
42. T. K. Bathala et al., "Quality comparison between three-dimensional T2-weighted SPACE and two-dimensional T2-weighted turbo spin echo magnetic resonance images for the

brachytherapy planning evaluation of prostate and periprostatic anatomy," *Brachytherapy* **19**(4), 484–490 (2020).
43. M. Shahedi et al., "Accuracy validation of an automated method for prostate segmentation in magnetic resonance imaging," *J. Digit. Imaging* **30**(6), 782–795 (2017).
44. N. Toussaint, J.-C. Souplet, and P. Fillard, "MedINRIA: medical image navigation and research tool by INRIA," in *Proc. of MICCAI'07 Workshop on Interaction in Medical Image Analysis and Visualization* (2007).
45. S. Warfield, K. Zou, and W. Wells, "Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation," *IEEE Trans. Med. Imaging* **23**, 903–921 (2004).
46. A. Suinesiaputra et al., "A collaborative resource to build consensus for automated left ventricular segmentation of cardiac MR images," *Med. Image Anal.* **18**(1), 50–62 (2014).
47. B. W. Cox et al., "International spine radiosurgery consortium consensus guidelines for target volume definition in spinal stereotactic radiosurgery," *Int. J. Radiat. Oncol. Biol. Phys.* **83**(5), e597–e605 (2012).
48. T. Popa et al., "Tumor volume measurement and volume measurement comparison plug-ins for VolView using ITK," *Proc. SPIE* **6141**, 395–402 (2006).
49. R. Cuocolo et al., "Quality control and whole-gland, zonal and lesion annotations for the PROSTATEx challenge public dataset," *Eur. J. Radiol.* **138**, 109647 (2021)
50. L. Dickinson et al., "Magnetic resonance imaging for the detection, localisation, and characterisation of prostate cancer: recommendations from a European consensus meeting," *Eur. Urol.* **59**(4), 477–494 (2011).
51. A. B. Rosenkrantz et al., "Prostate cancer localization using multiparametric MR imaging: comparison of prostate imaging reporting and data system (PI-RADS) and Likert scales," *Radiology* **269**(2), 482–492 (2013).
52. C. H. Sudre et al., "Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations," *Lect. Notes Comput. Sci.* **2017**, 240–248 (2017).
53. B. Lowekamp et al., "The design of SimpleITK," *Front. Neurosci.* **7**, 45 (2013).
54. A. Meyer et al., "Anisotropic 3D multi-stream CNN for accurate prostate segmentation from multi-planar MRI," *Comput. Methods Programs Biomed.* **200**, 105821 (2021).
55. S. Liu et al., "3D anisotropic hybrid network: transferring convolutional features from 2D images to 3D anisotropic volumes," *Lect. Notes Comput. Sci.* **11071**, 851–858 (2018).
56. N. Srivastava et al., "Dropout: a simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.* **15**(1), 1929–1958 (2014).
57. D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Instance normalization: the missing ingredient for fast stylization," arXiv:1607.08022 (2017).
58. J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *IEEE/CVF Conf. Comput. Vision and Pattern Recognit.*, pp. 7132–7141 (2018).
59. O. Oktay et al., "Attention U-net: learning where to look for the pancreas," arXiv:1804.03999 (2018).
60. G. Zhang et al., "A bi-attention adversarial network for prostate cancer segmentation," *IEEE Access* **7**, 131448–131458 (2019).
61. A. Saha, M. Hosseinzadeh, and H. Huisman, "End-to-end prostate cancer detection in bpMRI via 3D CNNs: effects of attention mechanisms, clinical priori and decoupled false positive reduction," *Med. Image Anal.* **73**, 102155 (2021).
62. F. Chollet et al., "Keras," 2015, https://github.com/fchollet/keras.
63. M. Abadi et al., "TensorFlow: large-scale machine learning on heterogeneous systems," 2015, arXiv:1603.04467, https://www.tensorflow.org/.
64. F. Isensee et al., "Batchgenerators: a Python framework for data augmentation," 2020, https://github.com/MIC-DKFZ/batchgenerators.
65. G. Wang et al., "Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks," *Neurocomputing* **338**, 34–45 (2019).
66. D. Shanmugam et al., "When and why test-time augmentation works," arXiv:2011.11156 (2020).

67. F. Isensee et al., "Brain tumor segmentation and radiomics survival prediction: contribution to the BRATS 2017 challenge," *Lect. Notes Comput. Sci.* **10670**, 287–297 (2018).
68. D. Karimi et al., "Prostate segmentation in MRI using a convolutional neural network architecture and training strategy based on statistical shape models," *Int. J. Comput. Assist. Radiol. Surg.* **13**(8), 1211–1219 (2018).
69. Z. Tang et al., "An augmentation strategy for medical image processing based on statistical shape model and 3D thin plate spline for deep learning," *IEEE Access* **7**, 133111–133121 (2019).
70. M. Hosseinzadeh et al., "Deep learning–assisted prostate cancer detection on bi-parametric MRI: minimum training data size requirements and effect of prior knowledge," *Eur. Radiol.* **32** (2021).
71. P. Mehta et al., "Computer-aided diagnosis of prostate cancer using multiparametric MRI and clinical features: a patient-level classification framework," *Med. Image Anal.* **73**, 102153 (2021).

Dimitri Hamzaoui is a PhD student at Inria Sophia Antipolis, France. His thesis focuses on prostate segmentation and detection of prostate tumors from multiparametric MRI. He received his engineering degree from Télécom Paris, Paris, France.

Sarah Montagne has a medical degree in medical imaging from Sorbonne University. She works in Pitié-Salpêtrière and Tenon Hospitals, Paris, and she specializes in uro-radiology.

Raphaële Renard-Penna is a professor of radiology at Sorbonne Université and the chairman of the Uro-radiology Department of the Pitié-Salpêtrière Hospital. Her research focuses on the development of automatic methods for the diagnosis and treatment of prostate cancer.

Nicholas Ayache is a research director at Inria in Sophia Antipolis, France, head of the Epione research team dedicated to e-patients for e-medicine, and scientific director of the AI institute 3IA Côte d'Azur where he holds a research chair. He is a member of the French Academy of Sciences and Academy of Surgery. His current research focuses on AI methods to improve diagnosis, prognosis, and therapy from medical images, and clinical, biological, behavioral, and/or environmental data available on the patient.

Hervé Delingette is a research director at Inria, scientific director at the Université Côte d'Azur, a fellow of the MICCAI Society, and he holds a chair at the AI institute 3IA Côte d'Azur. He received his engineering degree and PhD from the École Centrale Paris. His research focuses on various aspects of artificial intelligence in medical image analysis, computational physiology, and surgery simulation.