On the Stability of Neural Segmentation in Radiology

Moritz Wolter[1], Lokesh Veeramacheneni[1], Bettina Baeßler[2],
Ulrike I. Attenberger[3] and Barbara D. Wichtmann[3,4] *

1- High-Performance Computing and Analytics Lab, University of Bonn, GER

2- Inst. of Diagnostic and Interventional Radiology, University Hospital Würzburg, GER

3- Dept. of Diagnostic and Interventional Radiology, University Hospital Bonn, GER

4- Dept. of Neuroradiology, University Hospital Bonn, GER

Abstract. Neural networks promise automated prostate segmentation for the development of precise and quantifiable image-based biomarkers in modern personalized oncology. Before clinical translation, however, their stability must be ensured. In this study, we train three-dimensional U-shaped convolutional neural networks to segment prostate magnetic resonance imaging (MRI) scans and evaluate different loss formulations to improve their performance. To evaluate generalizability and reproducibility of our networks, we compare their performance in a clinically acquired test/re-test MRI data set of 26 prostate cancer patients that was previously not seen by the networks. We find our networks to be generalizable with good reproducibility with a mean Intersection over Union of 0.88. While initial results are promising, anatomical accuracy remains to be evaluated in larger, multi-center data sets. To facilitate clinical applicability, we provide an easy to use toolbox online.

1 Introduction

Modern personalized oncology relies on precise and quantifiable data. Recent advances in Magnetic Resonance Imaging (mri) have enabled the development of image-based biomarkers that can be used for data-driven medicine [1]. Neural networks promise not only fast and reliable automated organ segmentation to generate quantifiable imaging biomarkers, but also the standardization of diagnostic processes [2, 3, 4]. Prior to the routine clinical application of such networks, however, their stability must be ensured on test/re-test data [5].

This study examines the effect of focal-loss cost scaling on a previously published three-dimensional UNet for prostate segmentation [6, 7]. Prostate cancer is the most common malignant tumor in men, arising in 70-75% from the peripheral zone (PZ). Reliable prostate segmentation can facilitate and accelerate diagnosis, including the determination of prostate-specific antigen density, zonal application of the PI-RADS v2.1 criteria and risk assessment, e.g., of invasive growth [8, 9]. It may also serve urologists as a roadmap for targeted biopsies.
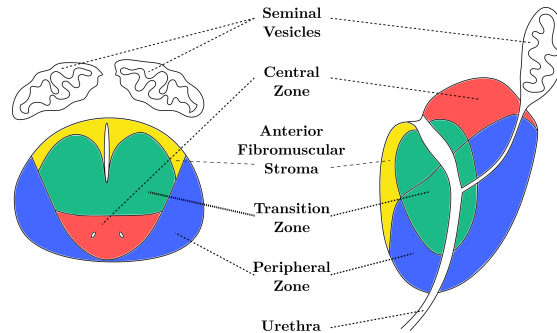
Fig. 1: Zonal anatomy of the prostate. Shown is an axial view (left) at the base of the prostate and a sagittal view (right) through the center (adapted from [8]).

To evaluate clinical generalizability and reproducibility, we compare the performance of our UNet in an mri test/re-test data set previously not seen by the network. To facilitate clinical transfer, our pipeline is available online [1].

2 Study data

With ethics committee approval, 26 prostate cancer patients (mean age 67.7 ± 4.9 years standard deviation) were enrolled in this study. Examinations were performed on a clinical 3T mri scanner (MAGNETOM Skyra; Siemens Healthcare, Erlangen, Germany). After sagittal and coronal high-resolution T2-weighted (T2w) imaging of the prostate, an axial high-resolution T2w sequence (voxel size 0.7 x 0.7 x 3 mm, TE/TR = 108/3350 ms, flip angle 160°, GRAPPA acceleration factor R = 2) of the prostate was acquired, replanned, and re-acquired to obtain a test/re-test data set.

The prostate is divided into four zones: transition zone (TZ), central zone (CZ), peripheral zone (PZ), and anterior fibromuscular stroma (AFS) (1). For our initial training of zonal segmentation, we rely on the ProstateX dataset comprising 346 subjects [10] and the annotations provided by Meyer et al. [7], including 78 annotations for training and 20 for testing purposes. Meyer et al. [7] combine the TZ and CZ to the central gland. During training, our networks only see the Prostate X data set, so we study their stability in the above described clinical test/re-test data set of 26 patients.

3 Computing the Region of Interest

Before we can fit a network to expert segmentation, we must locate the Region of Interest (RoI), which later serves as network input. After resampling the scan volumes to 0.5 x 0.5 x 3 mm, we compute the RoI's bounding box by using the intersection of at least two scans in axial as well as sagittal and/or

---

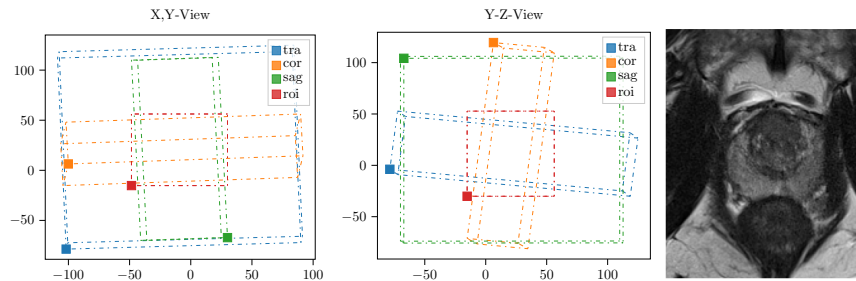[1] https://github.com/Krebs-Busters/zone-segmentation

Fig. 2: The X-Y, Y-Z bounding boxes on the left. The prostate is centred in the extracted RoI.

coronal orientation. Modern scanners store image slices as volumes alongside a displacement vector $\mathbf{o} \in \mathbb{R}^{3,1}$ and a rotation matrix $\mathbf{R} \in \mathbb{R}^{3,3}$. To compute the RoI we translate the array indices from local into global coordinate systems and back,

$$\mathbf{Rx} + \mathbf{o} = \mathbf{g}. \tag{1}$$

The equation above is computed for limiting points of a unit box at the origin. Our original inputs in $\mathbf{x}$ are the 3D-Tensor indices. The equation above yields displaced and rotated lines $\mathbf{g} \in \mathbb{R}^{3,1}$ in the device coordinate system. We find the intersection via a combination of maximum and minimum operations. Assuming the scans overlap, we compute the lower end of each box by choosing the largest value of the minima along the three axes. Similarly, we find the upper end by picking the smallest element from the maxima of the boxes for every axe. The first two plots of Figure 2 illustrate the process. Using the lower and upper bounds, we construct the bounding box of our RoI and transform it back to the tensor-index coordinate system via,

$$\mathbf{R}^{-1}\mathbf{b}_{\text{roi}} - \mathbf{o} = \mathbf{x}_{\text{roi}}. \tag{2}$$

With the inverse of the rotation matrix $\mathbf{R}^{-1}$. $\mathbf{x}_{\text{roi}} \in \mathbb{R}^3$ is a point on the boundary of the local RoI-box we seek. We discard all but the axial T2w scans, figure 2 shows a sample RoI. Once the region of interest is computed, we train a three-dimensional neural network as described by [7] to perform the zonal segmentation of the prostate. To achieve the best possible fit, we additionally introduce a focal loss [11] that, to our knowledge, has not previously been reported in the context of zonal segmentation of the prostate.

4   Cost functions

Lin et al. [11] introduce a sigmoidal focal-loss formulation. Focal loss is a variant of the cross entropy loss. The cross-entropy loss [12] serves as the starting point. To emphasize rare, often misclassified classes, Lin et al. [11] include a

	TZ		PZ		AFS	
cost	max	$\mu \pm \sigma$	max	$\mu \pm \sigma$	max	$\mu \pm \sigma$
focal-loss (ours)	0.92	0.905 ± 0.010	0.793	0.770 ± 0.017	0.534	0.396 ± 0.129
cross-entropy (ours)	0.92	0.901 ± 0.013	0.793	0.761 ± 0.029	0.417	0.327 ± 0.078
DSC-loss [7]	0.88	-	0.798	-	-	-
cross-entropy [14]	0.85	-	0.60	-	-	-

Table 1: Dice Similarity Coefficient (DSC)-test score comparison for our networks alongside results from the literature. We observe improved AFS segmentation performance without performance loss in the PZ and central gland for the focal loss term. Mean μ and standard deviation σ values are computed for four runs using the seeds $1, 2, 3, 4$.

scaling term into the formulation. This formulation works well when classes are not mutually exclusive. Entries of the network output tensor $\mathbf{y} \in [0,1]$ are normalized by a sigmoidal function. In our case, however, prostate parts can only ever be in a single zone. Consequently, we formulate a softmax version of the focal-loss. Again, we start with the cross-entropy loss formulation for this case [12]

$$\mathcal{L}_{\mathrm{mce}}(\mathbf{t}, \mathbf{y}) = -\sum_{n=1}^{N} \sum_{k=1}^{K} \mathbf{t}_{kn} \ln(\mathbf{y}_{kn}), \tag{3}$$

and add the focal scaling [13],

$$\mathcal{L}_{\mathrm{soft\text{-}fl}}(\mathbf{t}, \mathbf{y}) = -\sum_{n=1}^{N} \sum_{k=1}^{K} (1 - \mathbf{y}_{kn})^{\gamma} \mathbf{t}_{kn} \ln(\mathbf{y}_{kn}). \tag{4}$$

This formulation differs from [11]. In the following section, we will evaluate its utility for our prostate segmentation problem.

5 Zonal segmentation of the Prostate-X data set

The AFS does not consist of glandular tissue, but primarily of fibrous and smooth muscular components, leading to a relatively low signal intensity on T2w imaging. In the apical half striated muscles blend into the gland and the muscle of the pelvic diaphragm, while laterally and posteriorly the AFS thins to form the fibrous prostate capsule, which renders its segmentation challenging [15]. This section studies the effect of focal re-weighting on this problem. We choose to set $\gamma = 1.5$, for all focal losses. Table 1 lists network accuracies. Detection of the AFS significantly benefits from focal re-weighting of the cross entropy cost, both subjectively as well as in terms of a higher mean Dice Similarity Coefficient (DSC). Figure 3 illustrates the validation accuracy during training, since the sigmoidal formulations did not converge well, we omitted these from Table 5. This result is intuitive since every voxel can only be part of a single class. The softmax captures this setting better than the sigmoidal formulation.
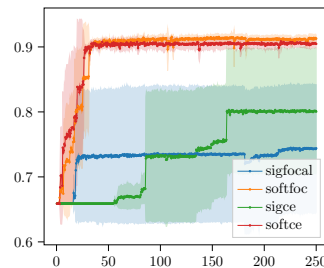
Fig. 3: Mean validation Intersection over Union (IoU) for training runs with seeds 1,2,3,4. The softmax-focal loss formulation delivers slightly better mean performance than unweighted softmax-cross-entropy on the ProstateX data.
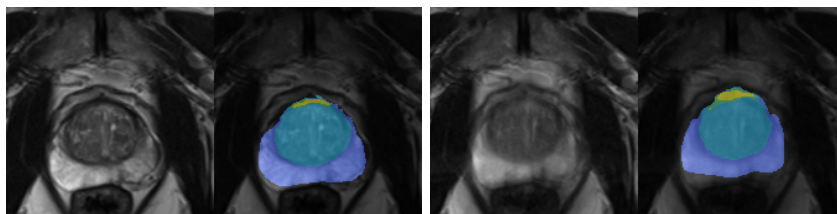

Fig. 4: Segmentation performance of the UNet in a previously unseen test/re-test prostate mri data set. Only subtle differences are notable. We observe robustness despite the slight motion corruption in the scan on the right. Zones are colored according to figure 1.

6 Clinical generalizability and reproduciblity

Generalizability and reproducibility are fundamental for clinical applicability. To this end, we analyze the performance of our UNet using a series of mri test-retest scans previously not seen by the network. Using the focal-loss trained network with seed 1, we segment scan/re-scan image pairs and compute the Intersection over Union (IoU) between every pair. Overall, we obtain a mean IoU of 0.88 calculated over all image pairs in the dataset, indicating a significant overlap of the test-retest-scan pair segmentations. Figure 4 illustrates the subtle differences in a test/re-test scan pair. Patient movement creates the blur visible on the right. Consequently, this is a difficult test-retest pair.

7 Conclusion

Focal-loss cost rescaling improved UNet-segmentation performance of the prostate, i.e., the AFS without performance loss in the PZ or central gland. Our neural network was generalizable to a previously unseen clinically acquired test/re-test mri scan data set with good reproducibility. While initial results are promising, anatomical accuracy remains to be evaluated in larger, multi-center data sets.

References

[1] Ulrike I Attenberger, Stephan Biber, and Barbara D Wichtmann. Technological advances of magnetic resonance imaging in today's health care environment. *Investigative Radiology*, 55(9), 2020.

[2] Pablo Cesar Quihui-Rubio, Gilberto Ochoa-Ruiz, Miguel Gonzalez-Mendoza, Gerardo Rodriguez-Hernandez, and Christian Mata. Comparison of automatic prostate zones segmentation models in mri images using u-net-like architectures. In *Mexican International Conference on Artificial Intelligence*. Springer, 2022.

[3] Multi-Institutional Target Delineation in Oncology Group. Human–computer interaction in radiotherapy target volume delineation: a prospective, multi-institutional comparison of user input devices. *Journal of digital imaging*, 24, 2011.

[4] Leon Lenchik, Laura Heacock, Ashley A Weaver, Robert D Boutin, Tessa S Cook, Jason Itri, Christopher G Filippi, Rao P Gullapalli, James Lee, Marianna Zagurovskaya, et al. Automated segmentation of tissues using ct and mri: a systematic review. *Academic radiology*, 26(12), 2019.

[5] Barbara D Wichtmann, Steffen Albert, Wenzhao Zhao, Angelika Maurer, Claus Rödel, Ralf-Dieter Hofheinz, Jürgen Hesser, Frank G Zöllner, and Ulrike I Attenberger. Are we there yet? the value of deep learning in a multicenter setting for response prediction of locally advanced rectal cancer to neoadjuvant chemoradiotherapy. *Diagnostics*, 2022.

[6] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*. Springer, 2015.

[7] Anneke Meyer, Marko Rakr, Daniel Schindele, Simon Blaschke, Martin Schostak, Andriy Fedorov, and Christian Hansen. Towards patient-individual pi-rads v2 sector map: Cnn for automatic segmentation of prostatic zones from t2-weighted mri. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE, 2019.

[8] American College of Radiology. Prostate imaging - reporting and data system, pi-rads v2.1. *https://www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/PI-RADS [accessed April 30, 2024]*, 2019.

[9] Barbara Daria Wichtmann, Frank Gerrit Zöllner, Ulrike Irmgard Attenberger, and Stefan O Schönberg. Multiparametric mri in the diagnosis of prostate cancer: Physical foundations, limitations, and prospective advances of diffusion-weighted mri. In *RöFo-Fortschritte auf dem Gebiet der Röntgenstrahlen und der bildgebenden Verfahren*, volume 193. Georg Thieme Verlag KG, 2021.

[10] Geert Litjens, Oscar Debats, Jelle Barentsz, Nico Karssemeijer, and Henkjan Huisman. Prostatex challenge data. *The cancer imaging archive*, 10:K9TCIA, 2017.

[11] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, 2017.

[12] Christopher M. Bishop. *Pattern recognition and machine learning, 5th Edition*. Information science and statistics. Springer, 2007.

[13] Guancheng Chen and Huabiao Qin. Class-discriminative focal loss for extreme imbalanced multiclass object detection towards autonomous driving. *The Visual Computer*, 2022.

[14] Germonda Mooij, Ines Bagulho, and Henkjan J. Huisman. Automatic segmentation of prostate zones. *CoRR*, abs/1806.07146, 2018.

[15] Anil Bhavsar, Sadhna Verma, et al. Anatomic imaging of the prostate. *BioMed research international*, 2014, 2014.