Computer Methods and Programs in Biomedicine xxx (xxxx) xxx

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

Anisotropic 3D Multi-Stream CNN for Accurate Prostate Segmentation from Multi-Planar MRI

Anneke Meyer[a,1,*], Grzegorz Chlebus[b,c,1], Marko Rak[a], Daniel Schindele[d], Martin Schostak[d], Bram van Ginneken[c,b], Andrea Schenk[b], Hans Meine[e,b], Horst K. Hahn[b], Andreas Schreiber[b], Christian Hansen[a]

[a] *Faculty of Computer Science and Research Campus STIMULATE, University of Magdeburg, Germany*
[b] *Fraunhofer Institute for Digital Medicine MEVIS, Bremen, Germany*
[c] *Radboud University Medical Center, Nijmegen, The Netherlands*
[d] *Clinic of Urology and Pediatric Urology, University Hospital Magdeburg, Germany*
[e] *University of Bremen, Medical Image Computing Group, Bremen, Germany*

ARTICLE INFO

Article history:
Received 31 January 2020
Accepted 26 October 2020
Available online xxx

Keywords:
MRI
Prostate Segmentation
Multi-Stream-CNN
Anisotropic CNN
Hyperparameter Optimization

ABSTRACT

Background and Objective: Accurate and reliable segmentation of the prostate gland in MR images can support the clinical assessment of prostate cancer, as well as the planning and monitoring of focal and loco-regional therapeutic interventions. Despite the availability of multi-planar MR scans due to standardized protocols, the majority of segmentation approaches presented in the literature consider the axial scans only. In this work, we investigate whether a neural network processing anisotropic multi-planar images could work in the context of a semantic segmentation task, and if so, how this additional information would improve the segmentation quality. *Methods:* We propose an anisotropic 3D multi-stream CNN architecture, which processes additional scan directions to produce a high-resolution isotropic prostate segmentation. We investigate two variants of our architecture, which work on two (dual-plane) and three (triple-plane) image orientations, respectively. The influence of additional information used by these models is evaluated by comparing them with a single-plane baseline processing only axial images. To realize a fair comparison, we employ a hyperparameter optimization strategy to select optimal configurations for the individual approaches. *Results:* Training and evaluation on two datasets spanning multiple sites show statistical significant improvement over the plain axial segmentation ($p < 0.05$ on the Dice similarity coefficient). The improvement can be observed especially at the base (0.898 single-plane vs. 0.906 triple-plane) and apex (0.888 single-plane vs. 0.901 dual-plane). *Conclusion:* This study indicates that models employing two or three scan directions are superior to plain axial segmentation. The knowledge of precise boundaries of the prostate is crucial for the conservation of risk structures. Thus, the proposed models have the potential to improve the outcome of prostate cancer diagnosis and therapies.

1. Introduction

Prostate cancer is the most prevalent type of cancer among men accounting for over 164 thousand new cases and more than 29 thousand deaths in the US in 2018 [1]. Clinical workflows of prostate cancer patients commonly involve MR imaging, which, thanks to the high soft-tissue contrast, can be employed for diagnosis, staging, and therapy planning. Prostate segmentation in MRI is a time-consuming task, requiring expert knowledge and suffering from inter-observer variability. Knowledge of the gland size and shape, which can be derived from the segmentation mask, is often utilized in clinical and research applications. For instance, Shah *et al.* [2] has shown that MRI findings can be correlated with the prostatectomy specimen by employing the prostate segmentation. Moreover, it is often used to facilitate radiotherapy planning [3] and targeted biopsy with MRI-TRUS (transrectal ultrasound) fusion [4,5]. Because neighboring structures as seminal vesicles, bladder, neurovascular bundles, and the external sphincter are essential for the erectile function and urine continence of men, the segmentation should be as precise as possible for the planning of prostate cancer therapy.

* Corresponding author. Postal Address: Universitaetsplatz 2, 39106 Magdeburg, Germany.
E-mail address: anneke.meyer@ovgu.de (A. Meyer).
[1] Anneke Meyer and Grzegorz Chlebus contributed equally to this work.

https://doi.org/10.1016/j.cmpb.2020.105821
0169-2607/

1.1. Related Work

Before the advance of deep learning, prostate segmentation was mainly performed with atlas-based segmentation or deformable models based on hand-crafted features. A comprehensive summary of those methods is given in [6]. Early approaches incorporating deep learning used voxel-wise classification to yield a segmentation mask. For instance, Liao *et al.* [7] learned deep features with a stacked independent subspace analysis network in an unsupervised fashion and perform segmentation with label propagation from atlases. Guo *et al.* [8] also used deep features but generated by a supervised stacked sparse autoencoder, yielding a prostate likelihood map, which is then segmented by a deformable model. Jia *et al.* [9] performed patch-based prediction with ensemble deep convolutional neural networks (CNNs).

CNNs are gaining attention in the medical image processing field thanks to state-of-the-art results on numerous classification and segmentation tasks. Various CNN architectures for segmentation problems were proposed. Long *et al.* proposed a fully convolutional neural network (FCN), which can be applied to arbitrarily sized images [10]. The U-Net model by Ronneberger *et al.* following the encoder/decoder design with long skip connections to retain the locality information was successfully used for different image segmentation problems [11]. Established CNN architectures, as well as their modified versions, have been introduced for prostate segmentation on T2-weighted MRI. For instance, Tian *et al.* fined-tuned a FCN model for prostate segmentation [12]. Yan *et al.* [13] adopted a FCN to embed superpixel information as low-level features in combination with high-level deep features. Another modification strategy to improve network segmentation is to add deep supervision [14–17].

Learning and segmentation performance can benefit from different aspects regarding network design to retain fine-detailed information and alleviate the vanishing gradient problem. While the U-Net architecture employs skip connections from the encoder to the decoder part of the network, Yu *et al.* [18] analyzed the effect of short and long residual connections and showed that a combination of both is beneficial in a 3D CNN for segmentation. Wang *et al.* [16] observed improvements with residual connections between neighboring blocks in combination with strided convolutions. Hossain *et al.* [19] adapted the VGG19 architecture [20] into an FCN and added short and long residual connections. A ResNet [21] encoder was extended with a decoder with 3D global convolutional block and boundary refinement blocks in [22]. The authors combined this network with an adversarial network for higher-order consistent predictions. In the whole model, anisotropic convolutions are employed to reflect the high slice thickness of the MR input volumes. The authors furthermore suggested using the ResNet encoder in combination with an anisotropic decoder and multi-level pyramid convolutional skip connections as well as adversarial training [23].

The use of dense connections that enhance feature reuse and propagation has been shown in the last two years to improve performance additionally. Hassanzadeh *et al.* [24] evaluated the use of various residual and dense connections. Yuan *et al.* [25] made use of densely connected blocks in encoder and decoder and trained with a joint loss function that incorporates the Dice similarity coefficient and the reconstruction error of dense block outputs. Also, Zhu *et al.* [26,27], To *et al.* [28] and Liu *et al.* [29] incorporated, amongst others, dense blocks into their architectures. Brosch *et al.* [30] formulated the segmentation as a regression task. They combined a 3D shape model with a convolutional regression network, where the network is used to obtain the distance from the surface mesh to the corresponding boundary point of the prostate.

The above-mentioned methods use only the axial T2-weighted scan as input, which is suboptimal as MR images acquired in a typical prostate imaging protocol are highly anisotropic (in-plane to out-of-plane resolution ratio of 6-10), see Fig. 1. This leads to substantial partial volume artifacts, making it difficult to precisely identify prostate boundaries, especially in the apex and base regions. In addition, segmentations created only on axial volumes suffer from step artifacts due to large slice spacing. However, in prostate cancer imaging protocols as in Weinreb *et al.* [31], it is mandatory to acquire at least an additional scan direction (sagittal or coronal) and in multiple clinical routines, all three scan directions are acquired for better interpretation. These additional scans could be used to improve the prostate segmentation quality, especially in the areas suffering from partial volume effects.

An approach to compute a high-quality prostate mesh was proposed by Shah *et al.* [2], where three masks resulting from manual contouring on axial, coronal, and sagittal MR acquisitions were merged by the means of shape-based interpolation. Cheng *et al.* introduced a fully automatic segmentation algorithm incorporating multi-planar MR information [32]. The algorithm includes an ensemble of three 2D neural networks trained separately on axial, coronal, and sagittal MR scans, respectively. The outputs are fused before a high-resolution prostate segmentation is extracted. Furthermore, Lozoya *et al.* [33] assessed the effect of single and dual plane segmentation by training ensembles of 2D CNNs independently on axial and sagittal volumes. The models process three consecutive image slices (downsampled to a 128×128 resolution) to segment the middle one. The results showed an improvement of 4% for the dual plane approach.

While these multi-planar approaches show that the exploitation of multi-planar MR images is beneficial for the segmentation quality, they have some limitations. First, both approaches train independent CNNs per MRI orientation, which prevents the models from learning how to combine the information coming from different orientations. Second, only 2D neural networks are employed which cannot capture the inherent volumetric information of MRI scans. Being able to analyze the 3D image context is important for the prostate segmentation as demonstrated by Ye *et al.* [34], who developed a volumetric ConvNet model that achieved the best performance at the PROMISE12 challenge so far [35]. In this work, we address both limitations by presenting a multi-stream 3D CNN architecture that processes simultaneously anisotropic multi-planar MR images to produce a high-resolution prostate segmentation. This paper builds on our previous work [36], where we demonstrated initial results of an isotropic multi-stream 3D network on a smaller dataset.

Additionally, we evaluate the performance of one, two and three input scan directions on the same dataset. While Lozoya *et al.* [33] only include two scan directions, Cheng *et al.* [32] and our previous work [36] use three planes. All works use different methods and datasets and therefore a thorough investigation of the difference between two and three planes has been impossible so far.

1.2. Contribution of this work

The contribution of this work is two-fold:

1. We propose an anisotropic 3D multi-stream CNN architecture and show that it can process multi-planar MR images to produce a high-resolution prostate segmentation. Contrary to our prior work [36], the proposed network design fuses information from anisotropic images alleviating the need for image resampling to isotropic voxel size. Additionally, the proposed architecture is computationally less expensive, which allows for faster inference and more efficient training.
2. We quantify the influence of information from additional image orientations on segmentation quality by comparing

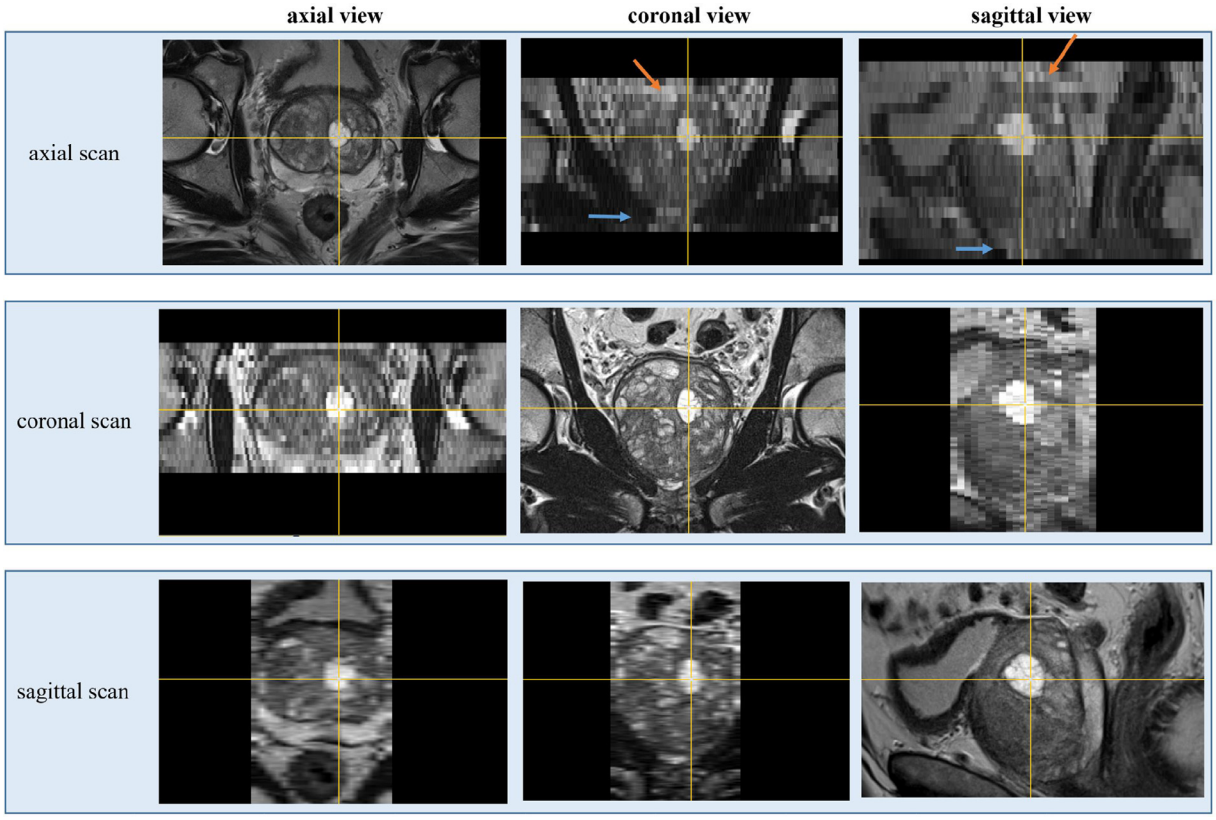


Fig. 1. Visualization of the independent orthogonal scans of one patient illustrating their anisotropic nature. The first row depicts the axial scan that is normally used for segmentation. As can be seen in the sagittal and coronal view of that axial scan, the apical (blue arrow) and base (orange arrow) region lack clear boundaries of the prostate due to partial volume effect. In the sagittal and coronal scans, the prostate tissue in these regions can be distinguished more clearly from non-prostate tissue.

performance of a baseline single-plane model (processing only axial images) with dual-plane (axial + sagittal) and triple-plane (axial + sagittal + coronal) models. To allow a fair comparison of the three approaches, we employ an automatic hyperparameter optimization strategy. We report quantitative results for whole-gland and base, mid and apex regions using image data from two datasets and multiple sites.

Our source code is available on GitHub[2], and we published ground truth segmentations that were created as part of this project for a publicly available challenge dataset [37].

In the following section, we describe the proposed architecture for the multi-plane segmentation of the prostate as well as our hyperparameter optimization method. Furthermore, we will give a description of the datasets used in this work, the training procedure, and the evaluation measures.

2. Materials and Methods

2.1. Multi-Stream Segmentation Network

With respect to the literature, we can basically define two variants of combining multiple planes for CNNs. The first way is to train three networks separately with each network taking one orthogonal scan as input. The output of the three networks is then fused in a postprocessing step. The alternative is to input all planes in one multi-stream network and and process them simultaneously. We compared the two variants to each other and could not find any significant difference in their performance (see results in Section 3.5). Due to its simplicity in deployment, we focus our work on the multi-stream architecture in this project. This has the additional benefit, that we can investigate the influence of additional planes directly, as the ensembling of network outputs has a benefit on performance in general.

Our multi-stream model is a 3D U-Net-like architecture following an encoder/decoder design with four resolution levels [38]. The proposed network design is flexible with respect to the number of inputs, which enables information extraction from more than one volume. Fig. 2 illustrates a triple-planar model instance processing axial, coronal, and sagittal acquisitions. Depending on the desired input configuration (single-plane, dual-plane, or triple-plane), the analysis path has corresponding input specific branches on the first two resolution levels. These branches perform downsampling by max pooling operations with anisotropic pool size (e.g., $2 \times 2 \times 1$ for the axial volume) to obtain equally sized outputs, which are concatenated before entering the third resolution level. Features from the analysis path are passed via long skip connections to the synthesis path. The skip connections contain optionally upsampling layers to bring the feature maps to the corresponding spatial size. To prevent information bottleneck, we double the filter size in the convolution layers directly followed by max-pooling. All convolution operations use a $3 \times 3 \times 3$ kernel and ReLU nonlinearity. If batch normalization layers are configured, then they are inserted between the convolution and the nonlinearity. We employ dropout layers in the synthesis path to avoid overfitting. The final layer is a $1 \times 1 \times 1$ convolution with a sigmoid activation to constrain model output to a $[0, 1]$ range.

[2] https://github.com/AnnekeMeyer/AnisotropicMultiStreamCNN

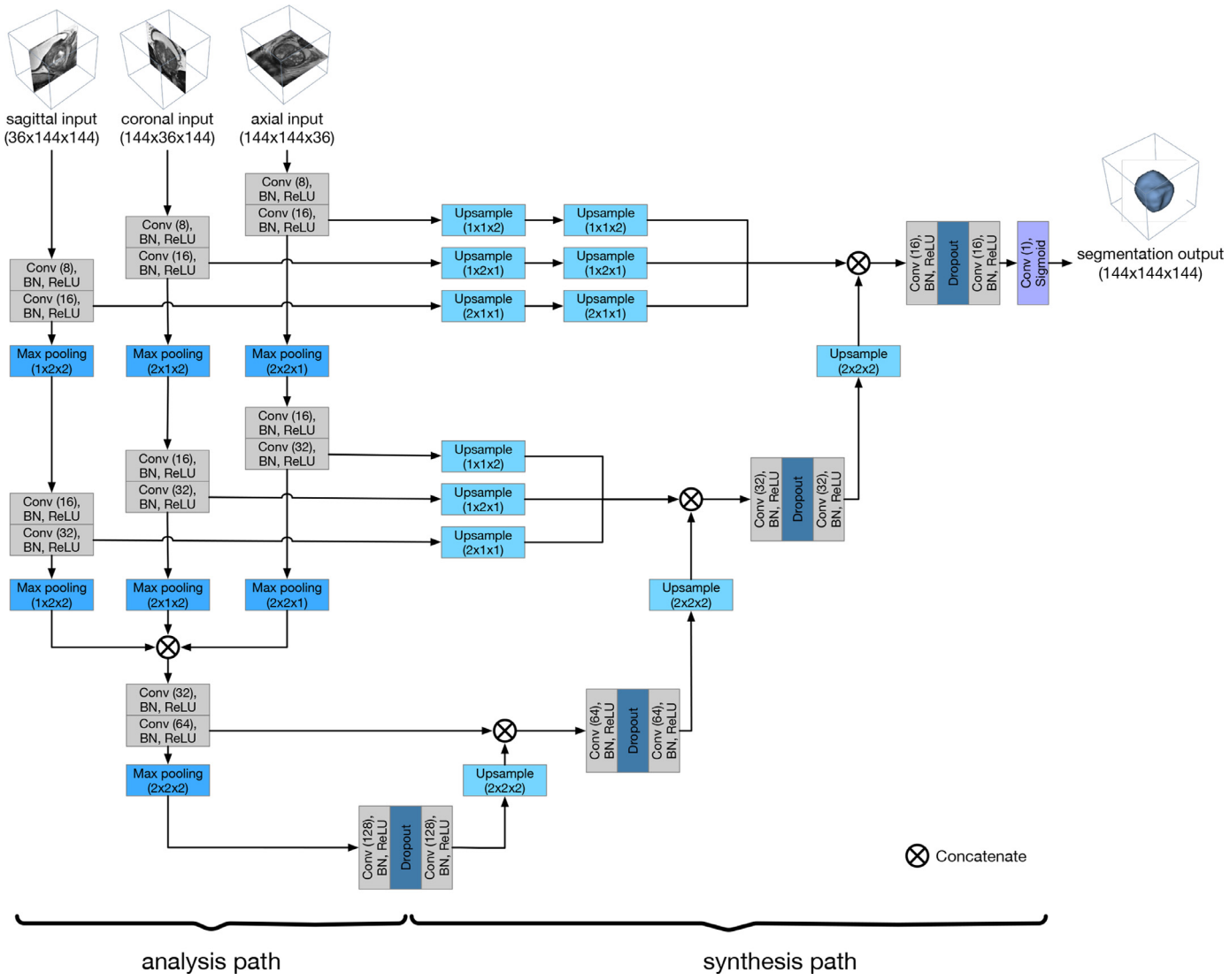


Fig. 2. Triple-planar multi-stream 3D network processing axial, coronal, and sagittal MR volumes. The number in parentheses denotes feature map count (conv layer), pool size (max pooling), and upsampling factor (upsampling). The upsampling is performed either by trilinear upsampling or 3D transposed convolution.

2.2. Hyperparameter Optimization

Careful tuning of neural network hyperparameters, such as learning rate or regularization strength, is important in getting the best possible model performance. Moreover, hyperparameter optimization (HPO) should be performed whenever the architecture or the learned task changes, as a direct transfer of hyperparameter values may lead to a sub-optimal prediction quality. We run HPO to find hyperparameter values yielding the best segmentation performance for all three architectures (single-, dual-, and triple-plane) independently. This strategy minimizes the influence of the chosen hyperparameters, yielding a fair comparison among the investigated models.

We employed the HPO strategy that was proposed by Falkner *et al.* in [39]. The method involves a combination of Hyperband (HB) with Bayesian optimization (BO) to achieve fast convergence to optimal configurations. HB is an HPO method that evaluates n randomly sampled configurations with a small budget (e.g., maximal training epoch count), keeps the best half, and doubles their budget [40]. This process is repeated until only one configuration is left. BO builds a probabilistic model based on the already evaluated configurations [41]. This model is then employed to sample hyperparameter values that should result in better model performance. One iteration of our HPO involves sampling $\frac{n}{2}$ configurations from the Bayesian model and another $\frac{n}{2}$ by random sampling. The sampled configurations are then evaluated using the HB method.

2.3. Dataset Description

We used two datasets for the evaluation of the proposed approaches. The first dataset is an in-house dataset containing 89 axial, sagittal and coronal T2-weighted MR scans acquired on a Philips Achieva 3T imager. In the clinical routine, gland segmentations have been obtained with commercial software (DynaCAD, Philips Invivo) in a semi-supervised manner. As the software only considers the axial T2 volumes, we resampled the segmentations to an isotropic resolution via shape-based interpolation as in Herman *et al.* [42]. Subsequently, an expert urologist reviewed and corrected the isotropic segmentations with 3D Slicer [43] by simultaneously considering all three orthogonal scans.

The second dataset ProstateX is publicly available through the SPIE-AAPM-NCI Prostate MR Classification Challenge [44–46], which was designed for predicting the clinical significance of prostate lesions. The dataset comprises multiparametric MRI ac-

Table 1
Resolution details for Prostate MRI datasets.

Dataset	Scan	Resolution [mm]
ProstateX	axial	[0.5-0.6] x [0.5-0.6] x [3-5]
	sagittal	0.56 x 0.56 x [3-4]
	coronal	[0.56-0.6] x [0.56-0.6] x [3-4.5]
In-House	axial	0.5 x 0.5 x 2.75
	sagittal	0.5 x 0.5 x 3.25
	coronal	0.5 x 0.5 x 2.76
PROMISE12	axial	[0.27-0.63] x [0.27-0.63] x [2.2-3.6]
	sagittal	not available
	coronal	not available

quired on two different types of Siemens 3T MR imagers; the MAGNETOM Trio and Skyra. As no reference segmentation of the glands is available in the challenge dataset, we created 66 segmentations for randomly chosen T2-weighted volumes. The segmentations were obtained manually for each scan direction by a medical student, followed by a review and corrections of an expert urologist with 3D Slicer under consideration of all three orthogonal scans. The final isotropic high-resolution prostate mask is extracted by taking the average of linearly resampled distance transformations of the individual segmentations and thresholding the result at zero (similar to the approach employed by Herman *et al.* [42]). The final masks were reviewed by an expert and corrected if necessary using 3D Slicer. These segmentations were published as part of the study to support open research [37,46].

The scans of both datasets were acquired without an endorectal coil. Details on the resolution of the orthogonal scans can be found in Table 1. The scans represent prostates with clinical variability such as tumors, cysts, benign prostatic hyperplasia, and scars from previous minimally invasive surgeries. The alignment of the orthogonal scans was checked visually using 3D Slicer. In about 10% of the cases, the scans were misaligned due to, for example, patient or bowel motion. For these cases, we performed a manual rigid registration of affected images. Volumes in the ProstateX dataset that did not contain the whole prostate were excluded from this study to have a fair comparison between the approaches. For the in-house dataset, no such cases were found.

Methods regarding the segmentation of the prostate glands are often compared to each other in the PROMISE12 challenge [35]. As this challenge dataset only consists of axial T2-weighted MR images (see Table 1), we were not able to make this comparison in this project. Instead, we focus on the comparison of different network architectures that are based on the multi-planar input volumes.

2.4. Preprocessing

For network training and prediction, the three scans are preprocessed by resampling (linear interpolation) them into a common coordinate system. The resulting resolution is $0.5\times0.5\times2.0$mm for axial scans, $0.5\times2.0\times0.5$mm for coronal scans, and $2.0\times0.5\times0.5$mm for the sagittal scans corresponding to their anisotropic acquisition. Next, the images are cropped, such that the resulting volume is the intersection of the three scans. They are further cropped or resized to an in-plane size of 184×184 and an out-of-plane size of 46. As intensity normalization, the gray values are cropped to the 1st and 99th percentiles and afterwards normalized to a range of [0,1].

2.5. CNN Training

We set aside randomly chosen 19 test cases for each dataset that were not considered for training. The remaining images were split into four folds for cross-validation. Hence, the folds of the in-house dataset consist of 52 training and 18 validation images each, while the ProstateX fold contains 35 training and 12 validation images. To augment the training set, random operations such as axial flips, elastic deformations, translations and rotations were used. Unnatural transformations such as top-bottom and front-back flips were not considered. The input images were cropped to a size of $144\times144\times144$ voxels, before being fed to the network.

The objective function of our networks is the negative soft Dice similarity coefficient (DSC)

$$loss = -\frac{2\sum_i^N p_i g_i + \epsilon}{\sum_i^N p_i^2 + \sum_i^N g_i^2 + \epsilon},$$

with N being the total number of voxels, p_i and g_i the predicted and reference voxels, respectively, and ϵ a small constant to ensure numerical stability. We ran the training with the Adam optimizer [47] for a maximum of 270 epochs, with an early stop criterium if the validation loss does not improve by at least $\delta = 0.001$ for 100 iterations. The mini-batch size was set to one due to GPU memory capacity (NVIDIA GeForce GTX 1080 Ti).

The prediction was post-processed with a connected components analysis, removing every component except for the largest. We ran the HPO on the concatenation of the first folds from both datasets. For each approach (single, dual and triple-plane), a separate HPO was performed. We optimized the hyperparameters which were empirically found to have substantial influence on model performance: learning rate (range $[10^{-6}, 10^{-3}]$), dropout rate ($0.0, 0.2, 0.4, 0.6$ or 0.8), upsampling mode (tri-linear or transposed convolution), and batch normalization (yes or no). The best performing hyperparameters for each approach, selected based on the validation loss, are summarized in Table 2. The total numbers of trainable parameters for the single-, dual, and triple-plane of the proposed network architectures are 1.4, 1.6, and 1.7 million, respectively. Thus, the proposed strategies are using similar network capacity.

2.6. Training Scenarios

We implemented two training scenarios:

- *Scenario I* - train one model on merged datasets
- *Scenario II* - train separate models for each dataset

By comparing models resulting from both scenarios, we can verify whether segmentation quality for a target dataset can benefit from training on multi-site data. For each scenario, four-fold cross-validation was performed.

2.7. Evaluation Measures

We evaluated the investigated models with the following measures that were also used in the PROMISE12 challenge [35]: Dice similarity coefficient (DSC) as well as the average boundary distances (ABD) and the 95th percentile Hausdorff-Distance (95-HD) between surface points of both volumes.

The Dice similarity coefficient is defined as

$$\mathrm{DSC}(X, Y) = \frac{2\,|X \cap Y|}{|X| + |Y|} \tag{1}$$

with X being the predicted and Y being the ground truth voxels. The average boundary distance is defined as:

$$\begin{aligned}\mathrm{ABD}(X_S, Y_S) = \tfrac{1}{|X_S|+|Y_S|}(&\textstyle\sum_{x\in X_S} \min_{y\in Y_S} \mathrm{ED}(x, y)\\ +\ &\textstyle\sum_{y\in Y_S} \min_{x\in X_S} \mathrm{ED}(y, x))\end{aligned} \tag{2}$$

where X_S and Y_S are the sets of surface points of the predicted and ground truth segmentation. ED is the Euclidean distance operator.

Table 2
Best performing hyperparameter for each of the investigated network architectures.

	Single-Plane	Dual-Plane	Triple-Plane
learning rate	1.28×10^{-4}	1.31×10^{-4}	2.99×10^{-4}
dropout rate	0.6	0.2	0.2
batch normalization	no	no	yes
upsampling mode	tri-linear	transposed convolution	transposed convolution

Table 3
Evaluation measures for scenario I (training on merged datasets) averaged across all folds. Asterisks mark significantly better results when compared to the single-plane model.

		Merged Datasets			ProstateX			In-House		
		Single	Dual	Triple	Single	Dual	Triple	Single	Dual	Triple
DSC	Whole	0.927	**0.933****	0.931*	0.917	**0.925***	0.922	0.936	**0.941***	0.939
	Apex	0.888	**0.901***	0.896	0.854	**0.880*****	0.872	**0.922**	0.921	0.920
	Mid	0.956	**0.958***	0.954	**0.957**	0.956	0.950	0.955	**0.960***	0.959
	Base	0.898	0.904**	**0.906***	0.884	0.890*	**0.893**	0.912	**0.919**	0.918
ABD[mm]	Whole	0.901	**0.841****	0.877	1.088	**1.019***	1.048	0.714	**0.664***	0.705
	Apex	0.990	**0.863***	0.916	1.343	**1.084*****	1.160	**0.637**	0.643	0.672
	Mid	**0.762**	0.779*	0.827	**0.797**	0.918	0.971	0.727	**0.640****	0.684
	Base	1.007	**0.946***	0.947*	1.230	1.176	**1.143**	0.783	**0.715**	0.751
95-HD[mm]	Whole	3.101	**3.005***	3.072	3.916	3.927	**3.721**	2.286	**2.083***	2.422
	Apex	2.992	**2.651***	2.810	4.017	**3.288*****	3.520*	**1.967**	2.015	2.100
	Mid	**2.483**	2.687**	2.740	**2.754**	3.439	3.212	2.213	**1.935***	2.269
	Base	3.097	2.932*	**2.899**	3.670	3.706	**3.478**	2.524	**2.158***	2.321

Best results are marked bold. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table 4
Evaluation measures for scenario II (models are trained and evaluated on each dataset individually) averaged across all folds. Asterisks mark significantly better results when compared to the single-plane model.

		ProstateX			In-House		
		Single	Dual	Triple	Single	Dual	Triple
DSC	Whole	0.919	0.923*	**0.926**	0.927	**0.939**	**0.939***
	Apex	0.865	0.873	**0.875**	0.917	0.919	**0.920**
	Mid	**0.956**	0.952	0.953	0.946	**0.960**	0.959
	Base	0.886	0.896	**0.900***	0.897	0.914	**0.915**
ABD[mm]	Whole	1.056	1.014*	**0.994**	0.793	0.704	**0.680**
	Apex	1.228	1.144	**1.118**	0.673	0.677	**0.662**
	Mid	**0.808**	0.906	0.910	0.810	**0.652**	0.658
	Base	1.207	1.094*	**1.065***	0.904	0.785	**0.729**
95-HD[mm]	Whole	3.731	**3.600**	3.666	2.604	2.405	**2.155**
	Apex	3.573	**3.393**	3.413	2.076	2.052	**1.999**
	Mid	**2.726**	3.054	3.375	2.482	**2.004**	2.016
	Base	3.718	**3.347**	3.456	2.920	2.676	**2.096****

Best results are marked bold. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

The Hausdorff distance is defined as

$$\begin{aligned} \mathrm{HD}(X_S, Y_S) &= \max\left(\mathrm{HD}'(X_S, Y_S), \mathrm{HD}'(Y_S, X_S)\right) \\ \text{with } \mathrm{HD}'(X_S, Y_S) &= \max_{x \in X_S}(\min_{y \in Y_S} \mathrm{ED}(x, y)). \end{aligned} \tag{3}$$

As done in [35], we used the 95th percentile for implementation of HD (the so-called 95-HD), as this measure is more often used, leveraging comparability with previous works.

All evaluation measures are computed in 3D each for the whole gland, apex, base, and mid-gland regions. Each region corresponds to ca. one-third of the prostate and was partitioned in a slice-based manner with regards to the manual reference segmentation.

3. Results and Discussion

We report quantitative results (averaged across folds) of both scenarios in Table 3 and Table 4, respectively. Each approach was subject to four-fold cross-validation, and the performance of the resulting models was evaluated on left-out test cases. We applied the Wilcoxon signed-rank test to obtain the statistical significance of quantitative differences between single and dual or triple plane approaches. The rationale against a standard Student t-Test is that we cannot assume Gaussianity for the distribution of the result quality.

In general, the additional scans used by the dual- and triple-plane models improved the segmentation quality when compared with the single-plane model. In the following, we present a more detailed result analysis for both considered scenarios as well as comparison with the inter-rater variability.

3.1. Scenario I

In training scenario I (training on merged datasets), the dual-plane approach that incorporates axial and sagittal volumes, works significantly better ($p < 0.05$) than the single-plane approach on both datasets and every region of the prostate with regards to every evaluation measure. The dual-plane method achieved an average DSC of 0.933 for the whole gland (vs. 0.927 for single-plane), 0.901 (vs. 0.888) in the apex and 0.958 (vs. 0.956) and 0.904 (vs. 0.898) for mid-gland and base, respectively. It has to be noted that the ABD and 95-HD for the mid-region are worse for dual-plane

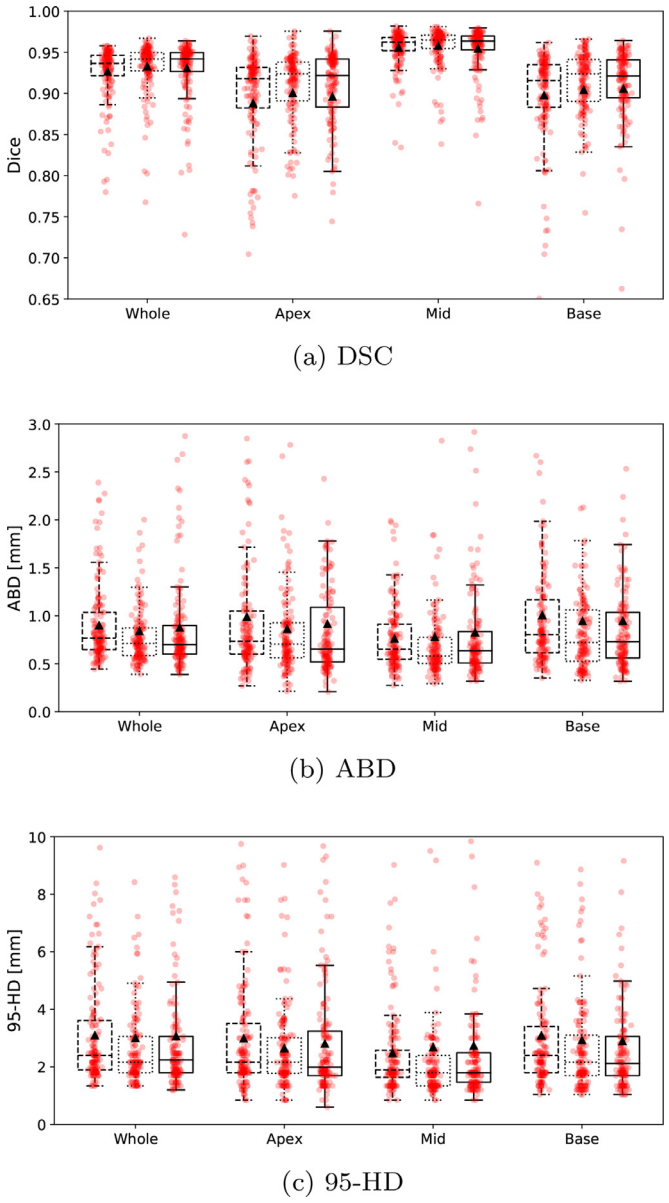


Fig. 3. Boxplots showing (a) DSC, (b) ABD, and (c) 95-HD for the whole gland and its subregions for single- (dashed), dual- (dotted), and triple-plane (solid) models. Models were trained on merged datasets (scenario I).

than single-plane, but the boxplots in Fig. 3 indicate that the dual-plane model performs better when the median is considered. The triple-plane model performed significantly better ($p < 0.05$) than the single-plane model regarding the DSC of the whole prostate as well as of the base region. Regarding distance-based measures, only the ABD of the base region was significant ($p < 0.05$).

Overall, the difference in performance between dual and triple-plane is less than between single-plane and triple- or dual-plane for training scenario I. Examples for the above-described segmentation quality differences are depicted in Fig. 4.

3.2. Scenario II

For scenario II, we can find less significant differences between the different approaches (see Tab. 4). This may be caused by the fact that less training data was available for each experiment. Opposed to scenario I (Table 3), where the dual-plane approach achieved the best performance for the evaluation measures in general, the triple-plane approach generally performs better in scenario II than dual-plane for each region and evaluation measure.

3.3. Scenario I vs. II

In general, the differences in performance between scenario II (training on individual datasets) and scenario I (training on merged datasets) were not substantial. However, we can see a slight improvement in the boxplots in Fig. 5 for the whole gland and most regions when models are trained on merged datasets.

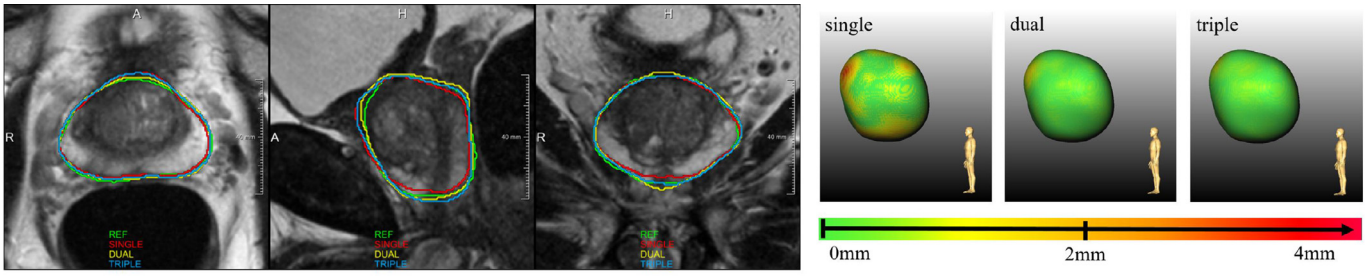
We observed that the quantitative evaluation measures in both scenarios are considerably better for the in-house datasets than for the ProstateX data. We assume that the reason for these results is two-fold: Firstly, the number of cases in the datasets are not balanced: the in-house dataset had almost 50% more cases available for training (n=70) than the ProstateX dataset (n=47). Secondly, the reference annotations were created with different methods: while the annotations for the ProstateX dataset were created entirely manually, the in-house dataset was segmented semi-automatically in the first stage and later refined manually. Even when experts review and correct the semi-automatically generated segmentations, there may still be a potential bias towards the semi-automatic segmentations, which could result in more consistent segmentations than with manual delineation. One might also argue that the image quality is another factor for performance quality. However, we could not confirm this visually.

Another observation we made is that, on the one hand, the triple-plane approach performs better than dual-plane if models are trained on separate datasets (scenario II). On the other hand, the dual-plane approach is more often significantly better than single-plane as the triple-plane approach is when trained on merged datasets (scenario I). Thus, dual-plane seems to be more robust to variations in the training data if multiple data sources are used. However, the quantitative differences between dual- and triple-plane in both training scenarios are not statistically significant.

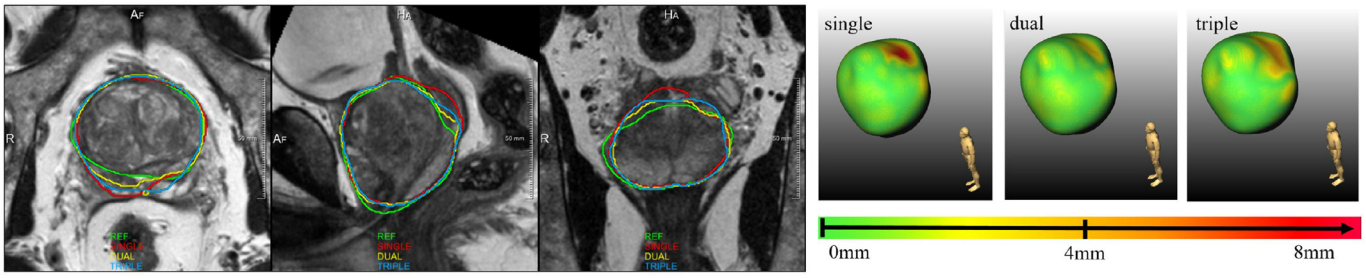
3.4. Inter-Rater Variability

To put our automatic segmentation results into perspective, we were interested to see in what range the inter-observer variability of prostate segmentation is (see Table 5). In the literature, second observer segmentation evaluation has been investigated within the scope of the PROMISE12 challenge [35]. The authors report a mean DSC of 0.90 between two expert segmentations for the whole gland and 0.80 and 0.86 for the apex and base, respectively. For the whole gland, they report an inter-rater variability of 5.64 mm for 95-HD.

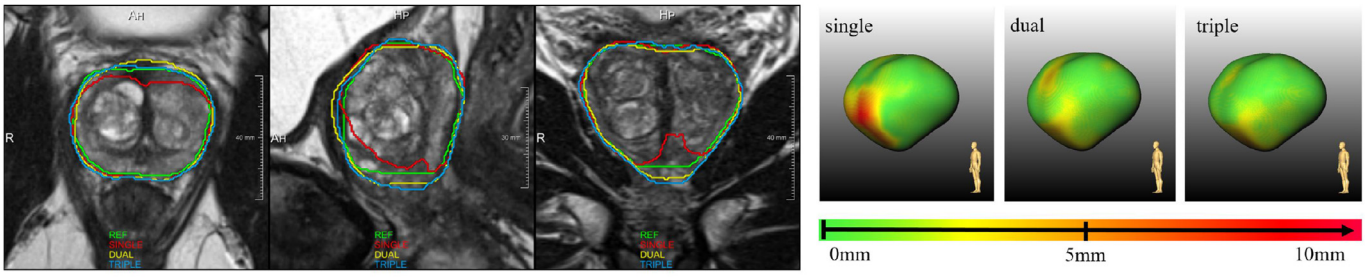
We carried out a similar study as part of another project where we asked two urologists to outline the glandular structures in the axial scans of 20 cases from the ProstateX challenge [48]. It has to be noted that those cases do not cover the test cases of this work. Nevertheless, we can still get a notion of how much two expert segmentations can vary. The inter-rater DSC for the whole gland, apex and base for these 20 cases were 0.93, 0.90 and 0.89, respectively. The 95-HD was 3.15 mm for the whole gland, which corresponds approximately to the thickness of one slice. Comparing these results to the overall DSC of 0.93 for the dual- and triple-planar model, we are clearly in the range of inter-rater variability. However, individual cases, as shown in Fig. 4d, still indicate that automatic segmentations need to be further improved in the future.



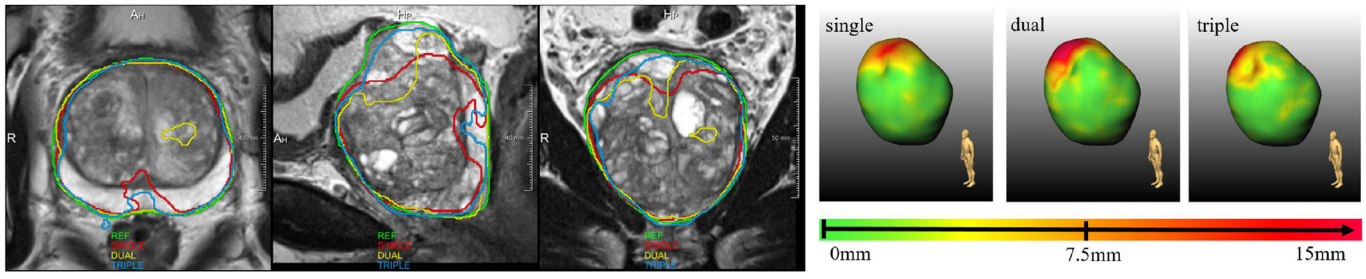
(a) Simple case where all approaches perform about equally well.



(b) Challenging case where dual/triple plane approaches are necessary. When considering only the axial plane, we yield overestimation in the base region.



(c) Challenging case where dual/triple plane approaches are necessary. Segmentation in apical region of the prostate is improved.



(d) Challenging case, where all approaches fail, presumably due to strong heterogeneity in the prostate gland.

Fig. 4. Four examples with different characteristics. On the left, segmentations in the image plane are depicted. Left column is the axial view, central column is sagittal view, and right column depicts the coronal view. On the right the surface distance between ground truth and prediction are shown for each approach.

3.5. Multi-Stream vs. Ensemble

We compared our triple-plane architecture processing all orthogonal images simultaneously, which directly outputs a prostate segmentation, with an ensemble approach from the literature [32,33]. In the ensemble approach three independent 3D models are trained for each image orientation, which outputs are combined in a post-processing step. We trained three single-plane models and used majority vote to compute the final segmentation. The experiment was performed on the ProstateX dataset and results were averaged across 4 folds. The results are listed in Table 6. No significant differences were found between the two meth-

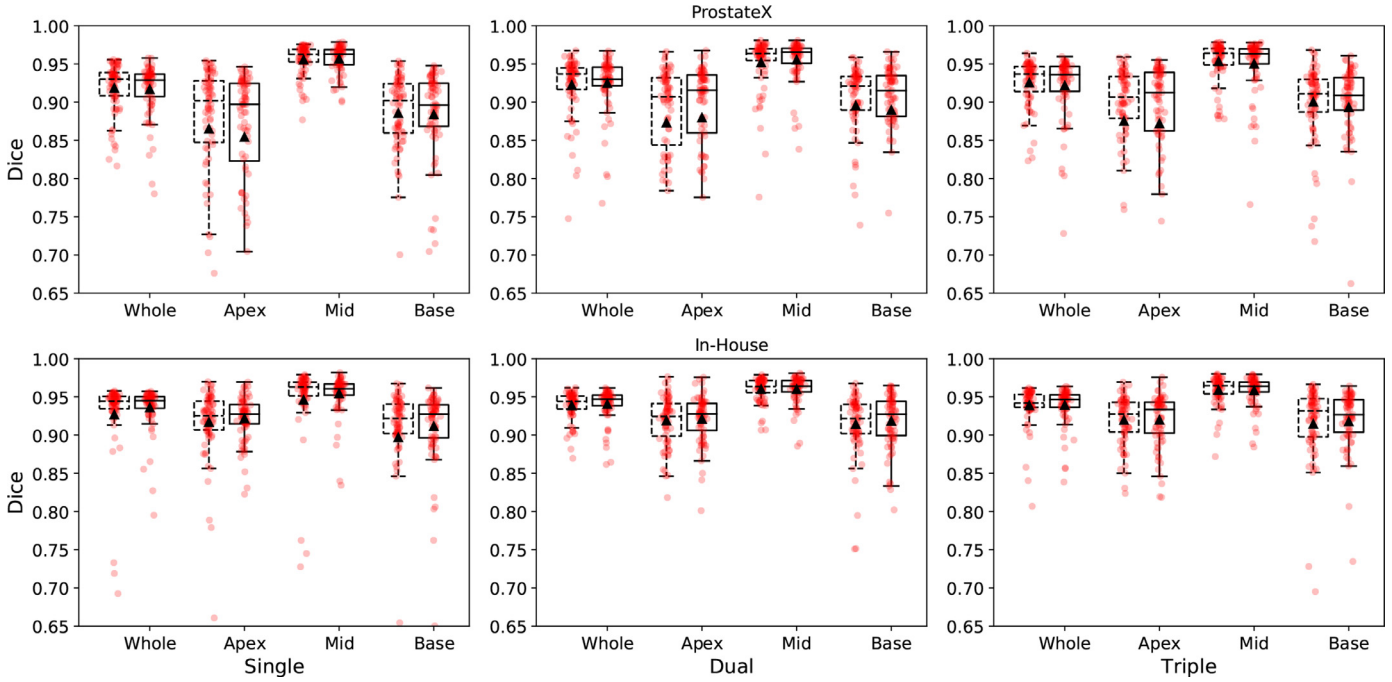


Fig. 5. Boxplots comparing Dice similarity coefficients for the whole gland and its subregions for models trained on only one dataset (scenario II, dashed) and merged datasets (scenario I, solid). Results are accumulated from all folds. The quality differences between scenario I and II are not substantial, yet, a slight improvement for the whole gland and most regions for models resulting from scenario I can be observed.

Table 5
Evaluation measures for inter-rater variability

		Inter-Observer	
		PROMISE12 (n=30)	ProstateX (n=20)
DSC	Whole	0.90	0.93
	Apex	0.80	0.90
	Mid	n.a.	0.96
	Base	0.86	0.89
ABD	Whole	1.82	0.66
	Apex	2.55	0.63
	Mid	n.a	0.49
	Base	2.21	0.86
95-HD	Whole	5.64	3.15
	Apex	6.36	2.84
	Mid	n.a	2.02
	Base	6.28	3.56

ods for any region and evaluation measure (Wilcoxon signed-rank test). The results are also in line with the outcome of our study that the input of multiple planes improves over a single-plane input.

Although no differences were found, we think that the multi-stream approach is superior to the ensemble because it requires less parameters (factor of 2.7) and therefore is easier to deploy in production. Moreover, using common decoder for all image orientations (as in multi-stream architecture) can be seen as a regularizer, which can help in minimising the generalization error on other datasets/tasks. For the ensemble we also evaluated output combination using shape-based interpolation, but it worked worse than majority vote.

4. Conclusion and Future Work

We proposed an anisotropic 3D multi-stream segmentation CNN that allows incorporating of different numbers of orthogonal input volumes. The objective of our work was to determine whether segmentation performance could be increased by the incorporation of sagittal and coronal volumes. To allow for a fair comparison between single-, dual- and triple-plane approaches, we included an automatic hyperparameter optimization strategy.

The most important finding of this work is that the use of multi-planar strategies significantly improves segmentation performance compared to using only axial volumes in almost all

Table 6
Comparison of two methods (ensemble and triple-plane) for generating segmentations from tri-planar input. No significant differences were found.

		ensemble	triple-plane
DSC	Whole	0.926 ± 0.03	0.926 ± 0.03
	Apex	0.871 ± 0.12	0.875 ± 0.12
	Mid	0.955 ± 0.02	0.953 ± 0.03
	Base	0.901 ± 0.04	0.900 ± 0.04
ABD[mm]	Whole	0.947 ± 0.36	0.994 ± 0.46
	Apex	0.871 ± 0.12	0.875 ± 0.12
	Mid	0.789 ± 0.30	0.910 ± 0.63
	Base	1.028 ± 0.56	1.065 ± 0.60
95-HD[mm]	Whole	3.10 ± 1.38	3.666 ± 2.23
	Apex	3.13 ± 1.72	3.413 ± 1.78
	Mid	2.29 ± 0.99	3.375 ± 3.26
	Base	3.01 ± 1.70	3.456 ± 2.17

cases. The quantitative differences between the three proposed approaches may not be large, but depending on the clinical application, the improved accuracy can be critical for the conservation of structures like external sphincter, bladder, or seminal vesicles. The clinical utility of the multi-planar approaches would be addressed in future work. Whether to prefer using the dual- or triple-plane variant could not be answered unequivocally. However, the dual-plane approach seems to be a good trade-off between computational cost and segmentation quality.

Future work will include an automatic registration among the orthogonal scans to compensate for potential transformations between them. This may lead to an increased performance of the multi-planar approaches, as the manual registration may not compensate for all motion artifacts and may be less precise than an automatic method. Another field of future research will be the detailed investigation of the multi-stream network architecture. For example, the location where the encoders are merged could be further examined.

Our results quality is comparable to the inter-rater variability. However, as mentioned above, some negative outliers would have never been produced by any medical experts. Hence, future work should also investigate how those outliers could be automatically detected and how much correction time would be required to achieve clinically acceptable segmentations. Furthermore, it would be interesting to apply our multi-stream architecture to other clinical use cases, where multi-planar imaging is acquired (e.g., cardiac MRI).

Conflict of Interest

This work has been funded by the EU and the federal state of Saxony-Anhalt, Germany under grant number ZS/2016/08/80388. Co-Funding was provided by Fraunhofer-Society. The Titan Xp used for this research was donated by the NVIDIA Corporation. Data used in this research were obtained from The Cancer Imaging Archive (TCIA) sponsored by the SPIE, NCI/NIH, AAPM, and Radboud University.

References

[1] R.L. Siegel, K.D. Miller, A. Jemal, Cancer statistics, 2018, CA Cancer J Clin 68 (1) (2018) 7–30, doi:10.3322/caac.21442.
[2] V. Shah, T. Pohida, B. Turkbey, H. Mani, M. Merino, P.A. Pinto, P. Choyke, M. Bernardo, A method for correlating in vivo prostate magnetic resonance imaging and histopathology using individualized magnetic resonance-based molds, Rev Sci Instrum 80 (10) (2009) 104301, doi:10.1063/1.3242697.
[3] M.A. Schmidt, G.S. Payne, Radiotherapy planning using MRI, Phys Med Biol 60 (22) (2015) R323–61, doi:10.1088/0031-9155/60/22/R323.
[4] A. Fedorov, S. Khallaghi, A.C. Sánchez, A. Lasso, S. Fels, K. Tuncali, E.S. Neubauer, T. Kapur, C. Zhang, W. Wells, P.L. Nguyen, P. Abolmaesumi, C. Tempany, Open-source image registration for MRI-TRUS fusion-guided prostate interventions, Int J Comput Assist Radiol Surg 10 (6) (2015) 925–934, doi:10.1007/s11548-015-1180-7.
[5] C.J. Das, A. Razik, A. Netaji, S. Verma, Prostate MRI–TRUS fusion biopsy: a review of the state of the art procedure, Abdominal Radiology (2020), doi:10.1007/s00261-019-02391-8.
[6] S. Ghose, A. Oliver, R. Martí, X. Lladó, J.C. Vilanova, J. Freixenet, J. Mitra, D. Sidibé, F. Meriaudeau, A survey of prostate segmentation methodologies in ultrasound, magnetic resonance and computed tomography images, Comput Methods Programs Biomed 108 (1) (2012) 262–287, doi:10.1016/j.cmpb.2012.04.006.
[7] S. Liao, Y. Gao, A. Oto, D. Shen, Representation learning: a unified deep learning framework for automatic prostate MR segmentation, Med Image Comput Comput Assist Interv 16 (2) (2013) 254–261, doi:10.1007/978-3-642-40763-5_32.
[8] Y. Guo, Y. Gao, D. Shen, Deformable MR prostate segmentation via deep feature learning and sparse patch matching, IEEE Trans Med Imaging 35 (4) (2016) 1077–1089, doi:10.1109/TMI.2015.2508280.
[9] H. Jia, Y. Xia, W. Cai, M. Fulham, D.D. Feng, Prostate segmentation in MR images using ensemble deep convolutional neural networks, in: Proc IEEE 14th Int Symp Biomed Imaging (ISBI), IEEE, 2017, pp. 762–765, doi:10.1109/ISBI.2017.7950630.
[10] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit, 2015, pp. 3431–3440, doi:10.1109/CVPR.2015.7298965.
[11] O. Ronneberger, P. Fischer, T. Brox, U-Net: convolutional networks for biomedical image segmentation, in: Med Image Comput Comput Assist Interv, 2015, pp. 234–241, doi:10.1007/978-3-319-24574-4_28.
[12] Z. Tian, L. Liu, B. Fei, Deep convolutional neural network for prostate MR segmentation, in: Proc SPIE Int Soc Opt Eng, 10135, 2017, p. 101351L, doi:10.1007/s11548-018-1841-4.
[13] K. Yan, X. Wang, J. Kim, M. Khadra, M. Fulham, D. Feng, A propagation-dnn: Deep combination learning of multi-level features for mr prostate segmentation, Comput Methods and Programs Biomed 170 (2019) 11–21, doi:10.1016/j.cmpb.2018.12.031.
[14] Q. Zhu, B. Du, B. Turkbey, P.L. Choyke, P. Yan, Deeply-supervised CNN for prostate segmentation, in: Proc Int Jt Conf Neural Netw, 2017, pp. 178–184. IJCNN.2017.7965852
[15] R. Cheng, H.R. Roth, N. Lay, L. Lu, B.I. Turkbey, W. Gandler, E.S. McCreedy, P. Choyke, R.M. Summers, M.J. McAuliffe, Automatic MR prostate segmentation by deep learning with holistically-nested networks, Proc SPIE Int Soc Opt Eng, 2017, doi:10.1117/12.2254558. pp. 101332H–101332H
[16] B. Wang, Y. Lei, S. Tian, T. Wang, Y. Liu, P. Patel, A.B. Jani, H. Mao, W.J. Curran, T. Liu, X. Yang, Deeply supervised 3D fully convolutional networks with group dilated convolution for automatic MRI prostate segmentation, Med Phys 46 (4) (2019) 1707–1718, doi:10.1002/mp.13416.
[17] B. Wang, Y. Lei, J.J. Jeong, T. Wang, Y. Liu, S. Tian, P. Patel, X. Jiang, A.B. Jani, H. Mao, W.J. Curran, T. Liu, X. Yang, Automatic MRI prostate segmentation using 3D deeply supervised FCN with concatenated atrous convolution, in: Proc SPIE Int Soc Opt Eng, 2019, p. 141, doi:10.1117/12.2512551.
[18] L. Yu, X. Yang, H. Chen, J. Qin, P.-A. Heng, Volumetric ConvNets with mixed residual connections for automated prostate segmentation from 3D MR images., in: Proc Conf AAAI Artif Intell, 2017, pp. 66–72.
[19] M.S. Hossain, A.P. Paplinski, J.M. Betts, Residual Semantic Segmentation of the Prostate from Magnetic Resonance Images, in: L. Cheng, A.C.S. Leung, S. Ozawa (Eds.), Neural Information Processing, Lecture Notes in Computer Science, 11307, Springer International Publishing, Cham, 2018, pp. 510–521, doi:10.1007/978-3-030-04239-4_46.
[20] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, arXiv preprint arXiv:1409.1556 (2014).
[21] K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, in: Proc 29th IEEE Comput Soc Conf Comput Vis Pattern Recognit, IEEE, 2016, pp. 770–778, doi:10.1109/CVPR.2016.90.
[22] H. Jia, Y. Song, D. Zhang, H. Huang, D. Feng, M. Fulham, Y. Xia, W. Cai, 3D Global Convolutional Adversarial Network for Prostate MR Volume Segmentation, arXiv preprint arXiv:1807.06742 (2018).
[23] H. Jia, Y. Xia, Y. Song, D. Zhang, H. Huang, Y. Zhang, W. Cai, 3D APA-Net: 3D Adversarial Pyramid Anisotropic Convolutional Network for Prostate Segmentation in MR Images, IEEE Trans Med Imaging (2019), doi:10.1109/TMI.2019.2928056.
[24] T. Hassanzadeh, L.G.C. Hamey, K. Ho-Shon, Convolutional Neural Networks for Prostate Magnetic Resonance Image Segmentation, IEEE Access 7 (2019) 36748–36760, doi:10.1109/ACCESS.2019.2903284.
[25] Y. Yuan, W. Qin, X. Guo, M. Buyyounouski, S. Hancock, B. Han, L. Xing, Prostate Segmentation with Encoder-Decoder Densely Connected Convolutional Network (Ed-Densenet), in: Proc IEEE 16th Int Symp Biomed Imaging (ISBI), 2019, pp. 434–437, doi:10.1109/ISBI.2019.8759498.
[26] Q. Zhu, Du Bo, P. Yan, Boundary-weighted domain adaptive neural network for prostate MR image segmentation, arXiv preprint arXiv:1902.08128 (2019).
[27] Q. Zhu, B. Du, J. Wu, P. Yan, A deep learning health data analysis approach: Automatic 3D prostate MR segmentation with densely-connected volumetric ConvNets, in: Proc Int Jt Conf Neural Netw, IEEE, 2018, pp. 1–6, doi:10.1109/IJCNN.2018.8489136.
[28] M.N.N. To, D.Q. Vu, B. Turkbey, P.L. Choyke, J.T. Kwak, Deep dense multi-path neural network for prostate segmentation in magnetic resonance imaging, Int J Comput Assist Radiol Surg 13 (11) (2018) 1687–1696, doi:10.1007/s11548-018-1841-4.
[29] Q. Liu, M. Fu, X. Gong, H. Jiang, Densely Dilated Spatial Pooling Convolutional Network using benign loss functions for imbalanced volumetric prostate segmentation, arXiv preprint arXiv:1801.10517 (2018).
[30] T. Brosch, J. Peters, A. Groth, T. Stehle, J. Weese, Deep learning-based boundary detection for model-based segmentation with application to MR prostate segmentation, Med Image Comput Comput Assist Interv (2018) 515–522, doi:10.1007/978-3-030-00937-3_59.
[31] J.C. Weinreb, J.O. Barentsz, P.L. Choyke, F. Cornud, M.A. Haider, K.J. Macura, D. Margolis, M.D. Schnall, F. Shtern, C.M. Tempany, H.C. Thoeny, B. Turkbey, A. Rosenkrantz, G. Villeirs, S. Verma, PI-RADS Prostate Imaging - Reporting and Data System: 2019, Version 2.1, Eur. Urol. 69 (1) (2016) 16–40.
[32] R. Cheng, N. Lay, F. Mertan, B. Turkbey, H.R. Roth, L. Lu, W. Gandler, E.S. McCreedy, T. Pohida, P. Choyke, M.J. McAuliffe, R.M. Summers, Deep learning with orthogonal volumetric HED segmentation and 3D surface reconstruction model of prostate MRI, in: Proc IEEE 14th Int Symp Biomed Imaging (ISBI), 2017, pp. 749–753, doi:10.1109/ISBI.2017.7950627.
[33] R. Cabrera Lozoya, A. Iannessi, J. Brag, S. Patriti, E. Oubel, Assessing the relevance of multi-planar MRI acquisitions for prostate segmentation using deep learning techniques, in: Proc SPIE Int Soc Opt Eng, 2018, p. 45, doi:10.1117/12.2293514.
[34] L. Yu, X. Yang, H. Chen, J. Qin, P.-A. Heng, Volumetric convnets with mixed residual connections for automated prostate segmentation from 3d MR images, in: Proc Conf AAAI Artif Intell, 2017, pp. 66–72.

[35] G. Litjens, R. Toth, W. van de Ven, C. Hoeks, S. Kerkstra, B. van Ginneken, G. Vincent, G. Guillard, N. Birbeck, J. Zhang, et al., Evaluation of prostate segmentation algorithms for MRI: the PROMISE12 challenge, Med Image Anal 18 (2) (2014) 359–373, doi:10.1016/j.media.2013.12.0029.
[36] A. Meyer, A. Mehrtash, M. Rak, D. Schindele, M. Schostak, C. Tempany, T. Kapur, P. Abolmaesumi, A. Fedorov, C. Hansen, Automatic high resolution segmentation of the prostate from multi-planar MRI, in: Proc IEEE 15th Int Symp Biomed Imaging (ISBI), 2018, pp. 177–181, doi:10.1109/ISBI.2018.8363549.
[37] D. Schindele, A. Meyer, D.F. von Reibnitz, V. Kiesswetter, M. Schostak, M. Rak, C. Hansen, High resolution prostate segmentations for the ProstateX-Challenge [Data set], 2020, (The Cancer Imaging Archive). 10.7937/TCIA.2019.DEG7ZG1U
[38] Ö. Çiçek, A. Abdulkadir, S.S. Lienkamp, T. Brox, O. Ronneberger, 3D U-Net: learning dense volumetric segmentation from sparse annotation, in: Med Image Comput Comput Assist Interv, 2016, pp. 424–432, doi:10.1007/978-3-319-46723-8_19.
[39] S. Falkner, A. Klein, F. Hutter, BOHB: Robust and efficient hyperparameter optimization at scale, arXiv preprint arXiv:1807.01774 (2018).
[40] L. Li, K. Jamieson, G. DeSalvo, A. Rostamizadeh, A. Talwalkar, Hyperband: A novel bandit-based approach to hyperparameter optimization, arXiv preprint arXiv:1603.06560 (2016).
[41] B. Shahriari, K. Swersky, Z. Wang, R.P. Adams, N. De Freitas, Taking the human out of the loop: A review of bayesian optimization, Proc IEEE 104 (1) (2015) 148–175, doi:10.1109/JPROC.2015.2494218.
[42] G.T. Herman, J. Zheng, C.A. Bucholtz, Shape-based interpolation, IEEE Comput Graph Appl 12 (3) (1992) 69–79, doi:10.1109/38.135915.
[43] A. Fedorov, R. Beichel, J. Kalpathy-Cramer, J. Finet, J.-C. Fillion-Robin, S. Pujol, C. Bauer, D. Jennings, F. Fennessy, M. Sonka, et al., 3D Slicer as an image computing platform for the quantitative imaging network, Magn Reson Imaging 30 (9) (2012) 1323–1341, doi:10.1016/j.mri.2012.05.001.
[44] G. Litjens, O. Debats, J. Barentsz, N. Karssemeijer, H. Huisman, ProstateX Challenge data, 2017, (The Cancer Imaging Archive). 10.7937/K9TCIA.2017.MURS5CL
[45] G. Litjens, O. Debats, J. Barentsz, N. Karssemeijer, H. Huisman, Computer-aided detection of prostate cancer in mri, IEEE Trans Med Imaging 33 (5) (2014) 1083–1092, doi:10.1109/TMI.2014.2303821.
[46] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, L. Tarbox, F. Prior, The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository, Journal of digital imaging 26 (6) (2013) 1045–1057, doi:10.1007/s10278-013-9622-7.
[47] D. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980 (2014).
[48] A. Meyer, M. Rak, D. Schindele, S. Blaschke, M. Schostak, A. Fedorov, C. Hansen, Towards patient-individual PI-Rads v2 sector map: CNN for automatic segmentation of prostatic zones from T2-weighted MRI, in: Proc IEEE 16th Int Symp Biomed Imaging (ISBI), 2019, pp. 696–700, doi:10.1109/ISBI.2019.8759572.