



Deep Learning-Based Boundary Detection for Model-Based Segmentation with Application to MR Prostate Segmentation

Tom Brosch^(✉), Jochen Peters, Alexandra Groth, Thomas Stehle, and Jürgen Weese

Philips GmbH Innovative Technologies, Hamburg, Germany
tom.brosch@philips.com

Abstract. Model-based segmentation (MBS) has been successfully used for the fully automatic segmentation of anatomical structures in medical images with well defined gray values due to its ability to incorporate prior knowledge about the organ shape. However, the robust and accurate detection of boundary points required for the MBS is still a challenge for organs with inhomogeneous appearance such as the prostate and magnetic resonance (MR) images, where the image contrast can vary greatly due to the use of different acquisition protocols and scanners at different clinical sites. In this paper, we propose a novel boundary detection approach and apply it to the segmentation of the whole prostate in MR images. We formulate boundary detection as a regression task, where a convolutional neural network is trained to predict the distances between a surface mesh and the corresponding boundary points. We have evaluated our method on the Prostate MR Image Segmentation 2012 challenge data set with the results showing that the new boundary detection approach can detect boundaries more robustly with respect to contrast and appearance variations and more accurately than previously used features. With an average boundary distance of 1.71 mm and a Dice similarity coefficient of 90.5%, our method was able to segment the prostate more accurately on average than a second human observer and placed first out of 40 entries submitted to the challenge at the writing of this paper.

1 Introduction

Model-based segmentation (MBS) [1] has been successfully used for the automatic segmentation of anatomical structures in medical images (e.g., heart [1]) due to its ability to incorporate prior knowledge about the organ shape into the segmentation method. This allows for robust and accurate segmentation, even when the detection of organ boundaries is incomplete. MBS approaches typically use rather simple features for detecting organ boundaries such as strong

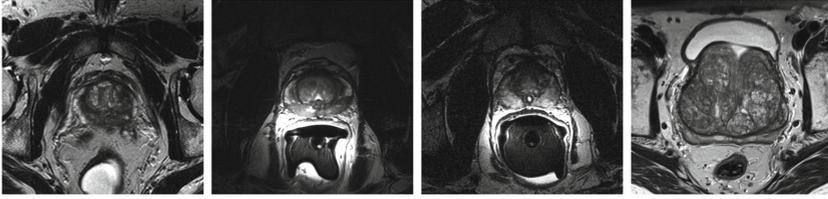


Fig. 1. Example images showing the large variability in image and prostate appearance.

gradients [8] and a set of additional constraints based on intensity value intervals [7] or scale invariant feature transforms [9]. Those features can detect organ boundaries reliably when they operate on well calibrated gray values, as is the case for computed tomography (CT) images. However, defining robust boundary features for the segmentation of organs with heterogeneous texture, such as the prostate, and varying MR protocols and scanners still remains a challenge due to the presence of weak and ambiguous boundaries caused by low signal-to-noise ratio and the inhomogeneity of the prostate, as well as the large variability in image contrast and appearance (see Fig. 1). To increase the robustness of the boundary detection for segmenting the prostate in MR images, Martin et al. [5] have used atlas matching to derive an initial organ probability map and then fine-tuned the segmentation using a deformable model, which was fit to the initial organ probability map and additional image features. Guo et al. [3] have extended this approach by using learned features from sparse stacked autoencoders for multi-atlas matching. Alternatively, Middleton et al. [6] have used a neural network to classify boundary voxels in MR images followed by the adaptation of a deformable model to the boundary voxels for lung segmentation. To speed up the detection of boundary points, Ghesu et al. [2] have used a sparse neural network for classification and restricted the boundary point search to voxels that are close to the mesh and aligned with the triangle normals.

We propose a novel boundary detection approach for fully automatic model-based segmentation of medical images and apply it to the segmentation of the whole prostate in MR images. We formulate boundary detection as a regression task, where a convolutional neural network (CNN) is trained to predict the distances between the mesh and the organ boundary for each mesh triangle, thereby eliminating the need for the time-consuming evaluation of many boundary voxel candidates. Furthermore, we combine the per-triangle boundary detectors into a single network in order to facilitate the calculation of all boundary points in parallel and designed it to be locally adaptive to cope with variations of appearance for different parts of the organ. We have evaluated our method on the Prostate MR Image Segmentation 2012 (PROMISE12) challenge [4] data set with the results showing that the new boundary detection approach can detect boundaries more robustly with respect to contrast and appearance variations and more accurately than previously used features and that the combination of shape-regularized model-based segmentation and deep learning-based boundary detection achieves the highest accuracy on this very challenging task.

2 Method

In this section, we will give a brief introduction to the model-based segmentation framework followed by a description of two network architectures for boundary detection: a global neural network-based boundary detector that uses the same parameters for all triangles, and a triangle-specific boundary detector that uses locally adaptive neural networks to search for the right boundary depending on the triangle index. A comprehensive introduction the model-based segmentation and previously designed boundary detection functions can be found in the papers by Ecabert et al. [1] and Peters et al. [7].

Model-Based Segmentation. The prostate surface is modeled as a triangulated mesh with fixed number of vertices V and triangles T . Given an input image I , the mesh is first initialized based on a rough localization of the prostate using a 3D version of the generalized Hough transformation (GHT) [1], followed by a parametric and a deformable adaptation. Both adaptation steps are governed by the external energy that attracts the mesh surface to detected boundary points. The external energy, E_{ext} , given a current mesh configuration and an image I is defined as

$$E_{\text{ext}} = \sum_{i=1}^T \left(\frac{\nabla I(\mathbf{x}_i^{\text{boundary}})}{\|\nabla I(\mathbf{x}_i^{\text{boundary}})\|} (\mathbf{c}_i - \mathbf{x}_i^{\text{boundary}}) \right)^2, \quad (1)$$

where \mathbf{c}_i denotes the center of triangle i , $\mathbf{x}_i^{\text{boundary}}$ denotes the boundary point for triangle i , and $\nabla I(\mathbf{x}_i^{\text{boundary}})$ is the image gradient at the boundary point $\mathbf{x}_i^{\text{boundary}}$. The boundary point difference $(\mathbf{c}_i - \mathbf{x}_i^{\text{boundary}})$ is projected onto the image gradient to allow cost-free lateral sliding of the triangles on the organ boundary. For the parametric adaptation, the external energy is minimized subject to the constraint that only affine transformations are applied to the mesh vertices. For the deformable adaptation, the vertices are allowed to float freely, but an internal energy term is added to the energy function, which penalizes deviations from a reference shape model of the prostate.

Neural Network-Based Boundary Detection. For each triangle, the corresponding boundary point is searched for on a line that is aligned with the triangle normal and passes through the triangle center. In previous work (e.g., [7]), candidate points on the search line were evaluated using predefined feature functions and the candidate point with the strongest feature response was selected as a boundary point. In contrast, we directly predict the signed distances d_i , $i \in [1, T]$, of the triangle centers to the organ boundary using neural networks, $f_i^{\text{CNN}} : \mathbb{R}^{D \times H \times W} \mapsto \mathbb{R}$, that process small subvolumes of I with depth D , height H , and width W such that

$$\mathbf{x}_i^{\text{boundary}} = \mathbf{c}_i + d_i \frac{\mathbf{n}_i}{\|\mathbf{n}_i\|} \quad (2)$$

with

$$d_i = f_i^{\text{CNN}}(S(I; \mathbf{c}_i, \mathbf{n}_i)), \quad (3)$$

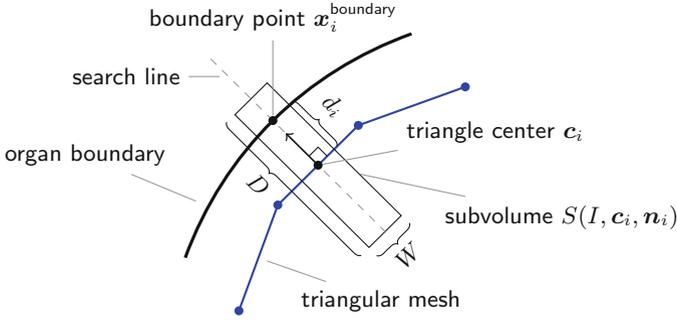


Fig. 2. Illustration of the boundary point search. For simplicity, the boundary point search is illustrated in 2D. The subvolume $S(I, \mathbf{c}_i, \mathbf{n}_i)$ is extracted from the image I and used by a neural network as input to predict the signed distance d_i of triangle i to its boundary point $\mathbf{x}_i^{\text{boundary}}$.

where \mathbf{n}_i are the normals of triangles i . The subvolumes $S(I, \mathbf{c}_i, \mathbf{n}_i)$ are sampled on a $D \times H \times W$ grid that is centered at \mathbf{c}_i and aligned with \mathbf{n}_i (see Fig. 2). The depth of the subvolumes is chosen such that they overlap with the organ boundary for the expected range of boundary distances called the capture range. The physical dimension of the subvolume is influenced by the number of voxels in each dimension of the subvolume and the spacing of the sampling grid. To keep the number of sampling points constant and thereby to allow the same network architecture to be used for different capture ranges, we change the voxel spacing in normal direction to account for different expected maximum distances of a triangle from the organ boundary. The parametric adaptation uses boundary detectors that were trained for an expected capture range of ± 20 mm and a sampling grid spacing of $2 \times 1 \times 1$ mm. We padded the size of the subvolume to account for the reduction of volume size caused by the first few convolutional layers, resulting in a subvolume size of $40 \times 5 \times 5$ voxels or $80 \times 5 \times 5$ mm. After the parametric adaptation, the prostate mesh is already quite well adapted to the organ boundary so we trained a second set of boundary detectors for a capture range of ± 5 mm and a sampling grid spacing of $0.5 \times 1 \times 1$ mm to facilitate the fine adaptation of the surface mesh during the deformable adaptation.

We propose and evaluate two different architectures for the boundary detection networks: a global boundary detector network that uses the same parameters for all triangles, and a locally adaptive network that adds a triangle-specific channel weighting layer to the global network and thereby facilitates the search for different boundary features depending on the triangle index. For both architecture, we combine the per-triangle networks, f_i^{CNN} , into a single network f^{CNN} that predicts all distances in one feedforward pass in order to speed up the prediction of all triangle distances and to allow for the sharing of parameters between the networks f_i^{CNN} :

$$(d_1, d_2, \dots, d_T) = f^{\text{CNN}}(S(I, \mathbf{c}_1, \mathbf{n}_1), S(I, \mathbf{c}_2, \mathbf{n}_2), \dots, S(I, \mathbf{c}_T, \mathbf{n}_T)). \quad (4)$$

Table 1. Network architecture with optional feature selection layer and corresponding dimensions used for predicting boundary point distances for each triangle for a subvolume size of $40 \times 5 \times 5$ voxels.

Layer type	Input dimension	Kernel size	# kernels	Output dimension
Conv + BN + ReLU	$T \times 40 \times 5^2$	$1 \times 7 \times 25$	32	$T \times 34 \times 32$
Conv + BN + ReLU	$T \times 34 \times 32$	$1 \times 7 \times 32$	32	$T \times 28 \times 32$
Conv + BN + ReLU	$T \times 28 \times 32$	$1 \times 7 \times 32$	32	$T \times 22 \times 32$
Conv + BN + ReLU	$T \times 22 \times 32$	$1 \times 22 \times 32$	32	$T \times 1 \times 32$
Conv + BN + ReLU	$T \times 1 \times 32$	$1 \times 1 \times 32$	32	$T \times 1 \times 32$
(Per-triangle weighting)	$T \times 1 \times 32$	—	—	$T \times 1 \times 32$
Conv	$T \times 1 \times 32$	$1 \times 1 \times 32$	1	$T \times 1 \times 1$

To simplify the network architecture, we assume that the width of all subvolumes is equal to their height and additionally reshape all subvolumes from size $D \times W \times W$ to $D \times W^2$. Consequently, the neural network for predicting the boundary distances is a function of the form $f^{\text{CNN}} : \mathbb{R}^{T \times D \times W^2} \mapsto \mathbb{R}^T$. The network input is processed using several blocks of convolutional (Conv), batch normalization (BN), and rectified linear unit (ReLU) layers called CBR blocks as summarized in Table 1, where each $1 \times A \times B$ kernel only operates on the input values and hidden units corresponding to a single triangle. Through the repeated application of valid convolutions, the network input of size $T \times D \times W^2$ is reduced to $T \times 1 \times 1$, where each element of the output vector represents the boundary distance of a particular triangle. Because the kernels are shared between all triangles, the network essentially calculates the same function for each triangle. However, the appearance of the interior and exterior of the organ might vary over the organ boundary and hence a triangle-specific distance function is often required. To allow for the learning of triangle-specific distance estimators, we extend the global network to a locally adaptive network by introducing a new layer that is applied before the last convolutional layer and defined as

$$\mathbf{x}_{L-1} = \mathbf{F} \odot \mathbf{x}_{L-2}, \quad (5)$$

where L is the number of layers of the network, \mathbf{x}_l is the output of layer l , \odot denotes element-wise multiplication, and $\mathbf{F} \in \mathbb{R}^{T \times 1 \times 32}$ is a trainable parameter matrix with one column per triangle and one row per channel of the output of the last CBR block. The locally adaptive network learns a pool of distance estimators, which are encoded in the convolutional kernels and shared between all triangles, along with triangle-specific weighting vectors encoded in the matrix \mathbf{F} that allow the distance estimation to be adapted for different parts of the surface mesh.

Training. Training of the boundary detectors requires a set of subvolumes that are extracted around each triangle and corresponding boundary distances, which can be generated from a set of training images and corresponding reference

meshes. To that end, we adapt a method previously used for selecting optimal boundary detectors from a large set of candidates called Simulated Search [7]. At each training iteration, mesh triangles are transformed randomly and independently of each other using three types of basic transformations: (a) random translations along the triangle normal, (b) small translations orthogonal to the triangle normal, and (c) small random rotations. Then, subvolumes are extracted for each transformed triangle and the distance of the triangle to the reference mesh is calculated. The network parameters are optimized using stochastic gradient descent by minimizing the root mean square error between the predicted and simulated distances. The coarse and fine boundary detectors have been trained with a translation range along the triangle normal of ± 20 mm and ± 5 mm, which matches the capture range of the respective networks.

3 Results

We have evaluated our method on the training and test set from the Prostate MR Image Segmentation 2012 (PROMISE12) challenge¹ [4]. The training set consists of 50 T2-weighted MR images showing a large variability in organ size and shape. The training set contains acquisitions with and without endorectal coils and was acquired from multiple clinical centers using scanners from different vendors, thereby further adding to the variability in appearance and contrasts of the training images. Training of the boundary detection networks took about 6 h on an NVIDIA GeForce 1080 GTX graphics card. Segmentation of the prostate took about 37 s on the GPU and 98 s on the CPU using 8 cores. A comparison of the global and locally adaptive boundary detection networks with previously proposed boundary detection functions [7] was performed on the training set using 5-fold cross-validation. For a direct comparison to state-of-the-art methods, we submitted the segmentation results produced by the locally adaptive method on the test set for evaluation to the challenge.

For the comparison of different boundary detectors, we measured the segmentation accuracy in terms of the average boundary distance (ABD) between the produced and the reference segmentation. We were not able to achieve good segmentation results (ABD = 6.09 mm) using designed boundary detection functions with trained parameters as described in [7], which shows the difficulty of detecting the right boundaries for this data set. Using the global boundary detection network, we were able to achieve satisfying segmentation results with a mean ABD of 2.08 mm. The ABD could be further reduced to 1.48 mm using the locally adaptive network, which produced similar results compared to the global network, except for a few cases where the global network was not able to detect the correct boundary (see Fig. 3(a)) due to the inhomogeneous appearance of the prostate. In those cases, the global network only detected the boundary of the central gland, which produces the correct result for the anterior part of the prostate, but causes errors where the prostate boundary is defined by the peripheral zone. In contrast, locally adaptive networks (see Fig. 3(b)) are able

¹ <https://promise12.grand-challenge.org/>.

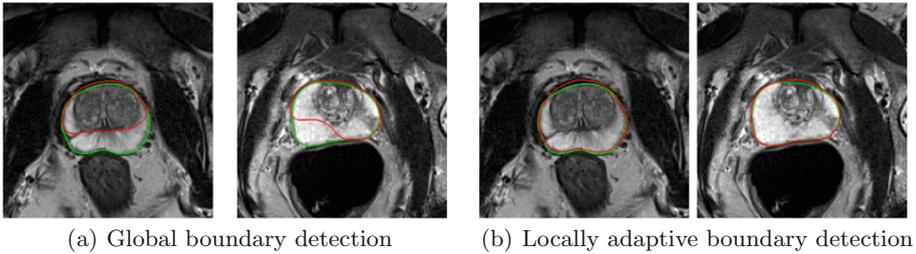


Fig. 3. Comparison of segmentation results (red) and reference meshes (green) using two network architectures. The locally adaptive network correctly detects the prostate hull for the central gland and the peripheral zone, despite the large appearance differences of the two structures.

to switch between the detection of the central gland and the peripheral zone depending on the triangle index, consequently detecting the true boundary in all cases.

A comparison of our method to the best performing methods on the PROMISE12 challenge in terms of the Dice similarity coefficient (DSC), the average boundary distance (ABD), the absolute volume difference (VD), and the 95 percentile Hausdorff distance (HD95) calculated over the whole prostate is summarized in Table 2. The “score” relates a metric to a second observer, where a score of 85 is assigned if a method performs as well as a second observer and a score of 100 corresponds to a perfect agreement with the reference segmentation. At the writing of this paper, our method placed first in the challenge out of 40 entries, although the scores of the top three methods are very close. With

Table 2. Comparison of our method to state-of-the-art methods on the PROMISE12 challenge in terms of the Dice similarity coefficient (DSC), the average boundary distance (ABD), absolute volume difference (VD), and the 95 percentile Hausdorff distance (HD95) calculated over the whole prostate. Our method ranks first in all metrics except for HD95 and performs better on average than a second observer (score > 85).

Rank	Method (Year)	DSC	ABD	VD	HD95	Score
1	Our method (2018)	90.5	1.71	6.6	4.94	87.21
2	AutoDenseSeg (2018)	90.1	1.83	7.6	5.36	87.19
3	CUMED (2016)	89.4	1.95	7.0	5.54	86.65
4	RUCIMS (2018)	88.8	2.05	8.5	5.59	85.78
5	CREATIS (2017)	89.3	1.93	9.2	5.59	85.74
6	methinks (2017)	87.9	2.06	8.7	5.53	85.41
7	MedicalVision (2017)	89.8	1.79	8.2	5.35	85.33
8	BDSlab (2017)	87.8	2.35	9.1	7.59	85.16
9	IAU (2018)	89.3	1.86	7.7	5.34	84.84
10	UBCRCL (2017)	88.8	1.91	10.6	4.90	84.48

a DSC of 90.5%, an average boundary distance of 1.71 mm, and a mean absolute volume difference of 6.6% calculated over the whole prostate, our method achieved the best scores in these three metrics. Our method is second in only one metric, the 95 percentile Hausdorff distance, where our method achieved the second best value (4.94 mm) and is only slightly worse than the fully convolutional neural network approach by UBCRCL, which achieved a distance of 4.90 mm. Overall, our method demonstrated very good segmentation results and performed better on average (score > 85) than a second human observer.

4 Conclusion

We presented a novel deep learning-based method for detecting boundary points for the model-based segmentation of the prostate in MR images. We showed that using neural networks to directly predict the distances to the organ boundary instead of evaluating several boundary candidates using hand-crafted boundary features significantly improves the accuracy and robustness to large contrast variations. The accuracy could be further improved by making the network locally adaptive, which facilitates the learning of boundary detectors that are tuned for specific parts of the boundary. With an average boundary distance of 1.71 mm and a Dice similarity coefficient of 90.5%, our method was able to segment the prostate more accurately on average than a second human observer and placed first out of 40 submitted entries on this very challenging data set.

References

1. Ecabert, O., et al.: Automatic model-based segmentation of the heart in CT images. *IEEE Trans. Med. Imaging* **27**(9), 1189–1201 (2008)
2. Ghesu, F.C., et al.: Marginal space deep learning: efficient architecture for volumetric image parsing. *IEEE Trans. Med. Imaging* **35**(5), 1217–1228 (2016)
3. Guo, Y., Gao, Y., Shen, D.: Deformable MR prostate segmentation via deep feature learning and sparse patch matching. In: *Deep Learning for Medical Image Analysis*, pp. 197–222. Elsevier (2017)
4. Litjens, G., et al.: Evaluation of prostate segmentation algorithms for MRI: the PROMISE12 challenge. *Med. Image Anal.* **18**(2), 359–373 (2014)
5. Martin, S., Troccaz, J., Daanen, V.: Automated segmentation of the prostate in 3D MR images using a probabilistic atlas and a spatially constrained deformable model. *Med. Phys.* **37**(4), 1579–1590 (2010)
6. Middleton, I., Damper, R.I.: Segmentation of magnetic resonance images using a combination of neural networks and active contour models. *Med. Eng. Phys.* **26**(1), 71–86 (2004)
7. Peters, J., Ecabert, O., Meyer, C., Kneser, R., Weese, J.: Optimizing boundary detection via simulated search with applications to multi-modal heart segmentation. *Med. Image Anal.* **14**(1), 70–84 (2010)
8. Vincent, G., Guillard, G., Bowes, M.: Fully automatic segmentation of the prostate using active appearance models. In: *2012 MICCAI Grand Challenge: Prostate MR Image Segmentation* (2012)
9. Yang, M., Yuan, Y., Li, X., Yan, P.: Medical image segmentation using descriptive image features. In: *BMVC*, pp. 1–11 (2011)