Deep Learning for Volumetric Segmentation in Spatio-temporal Data: Application to Segmentation of Prostate in DCE-MRI

Jian Kang[1] Gihan Samarasinghe[1] Upul Senanayake[1]
Sailesh Conjeti[2,3] Arcot Sowmya[1]

[1] School of Computer Science and Engineering,
University of New South Wales,
Sydney, Australia
[2] German Center for Neurodegenerative Diseases,
Bonn, Germany
[3] Computer Aided Medical Procedures,
Technische Universität München,
Munich, Germany

Technical Report
UNSW-CSE-TR-201807
October 2018

UNSW
SYDNEY

School of Computer Science and Engineering
The University of New South Wales
Sydney 2052, Australia

Abstract

Segmentation of the prostate in MR images is an essential step that underpins the success of subsequent analysis methods, such as cancer lesion detection inside the tumour and registration between different modalities. This work focuses on leveraging deep learning for analysis of longitudinal volumetric datasets, particularly for the task of segmentation, and presents proof-of-concept for segmentation of the prostate in 3D+T DCE-MRI sequences. A two-stream processing pipeline is proposed for this task, comprising a spatial stream modelled using a volumetric fully convolutional network and a temporal stream modeled using recurrent neural networks with Long-Short-term Memory (LSTM) units. The predictions of the two streams are fused using deep neural networks. The proposed method has been validated on a public benchmark dataset of 17 patients, each with 40 temporal volumes. When averaged over three experiments, a highly competitive Dice overlap score of 0.8688 and sensitivity of 0.8694 were achieved. As a spatio-temporal segmentation method, it can easily migrate to other datasets.

1 Introduction

Prostate cancer ranks third by incidence (and first in men) amongst 34 types of cancers [1]. Efforts are increasingly directed at early stage diagnosis and assessment of the extent of malignancy. While multiple treatment options are available, their success depends on detection and diagnosis of tumours. Imaging of the prostate is one of the most widely utilised methods for determining clinically useful information that may be important in accurate and well-directed treatment as well as prevention of treatment related morbidities [10]. Amongst various imaging techniques available, including Computed-Tomography (CT) and Transrectal Ultrasound (TRUS), Magnetic Resonance Imaging (MRI) with dedicated sequences for prostate cancer is being increasingly adopted as standard clinical practice, as it offers favourable qualities such as higher spatial resolution, better soft tissue contrast and better safety in the sense of radiation involvement [16]. Novel MRI modalities have been introduced for prostate image acquisition, in addition to traditional morphological T1-weighted and T2-weighted MRI. The most popular and common novel prostate MRI modalities include diffusion weighted MRI (DWI-MRI) and Dynamically Contrast Enhanced MRI (DCE-MRI). DWI-MRI visualises water cell diffusion rates within soft tissue, while DCE-MRI produces a temporal sequence of T1-weighted MR images that represent perfusion of an administered contrast agent between blood vessels and extracellular, extra-vascular regions within soft tissues [18].

Segmentation of the prostate in images is an essential procedure that underpins the success of further analysis for classification or recognition of disease [7]. However, prostate segmentation is a challenging task on any imaging protocol, including different MRI modalities, as the prostate does not have a rigid boundary, and it is surrounded by many other organs that show complex intensity variations in images. Segmentation of the prostate in morphological MRI has been addressed widely in the literature. Recently, segmentation methods based on deep learning algorithms have achieved good success for many different organs, including the prostate. For T2 MRI-based prostate segmentation, V-net [12] was the first attempt at 3D volumetric end-to-end segmentation. Later, Yu et al. [19] improved the performance by adding long and short Res-connections to the networks. Moreover, densely connected layers were used in Auto-DenseSeg [2]. For prostate Diffusion MRI segmentation, Clark et al. [3] applied a modified version of ResNet to process 2D slices.

Surprisingly, segmentation of the prostate in DCE-MRI has been limited. Different from morphological MRI modalities, DCE-MRI is a set of volumes over time, which gives rise to spatio-temporal (3D+T) datasets. Firjani et al.'s work [4] is the state-of-art for DCE-MRI prostate segmentation. In their work, three types of information are taken into account, namely intensity, spatial interaction and shape information. Intensity refers to intensity differences between prostate and non-prostate voxels over time. Spatial interaction indicates rotations and deformations of the prostate view. When it comes to shape information, a shape prior is learned from co-aligned 3D segmented prostate data. Their method was evaluated on 270 DCE-MRI series from 15 independent subjects and achieved DSC of 92% and sensitivity of 85%.

To date, there is no reported work on DCE-MRI prostate segmentation based on deep learning algorithms. This motivated the work on segmentation of the prostate DCE-MRI datasets by adopting deep learning algorithms. Video

segmentation [6, 17] is a closely related problem that deals with 2D+T data, and it is usually achieved using optical flow-based methods to track moving objects. Another related work[5] is a longitudinal study where a structural MRI is acquired at baseline and followup points. Since their goal is to track and predict disease progression, the time between two MRI volume acquisitions is of the order of months. In this work, the task is to segment, at acquisition time itself, a volume of relatively stable size but gradually increasing contrast over a relatively short period of a few seconds.

The work in this paper demonstrates the feasibility of deep learning for spatio-temporal segmentation of the prostate in DCE-MRI datasets. A deep learning framework is implemented for prostate segmentation using a publicly available benchmark DCE-MRI dataset [11]. In this work, 3D spatial information is extracted using a modified U-net architecture [13] and temporal information is dealt with by utilising Long Short Term memory (LSTM) networks [8] on voxels. Then the spatial and temporal information are fused together using a Multi-Layer Perceptron (MLP) network. The main contributions include (i) a novel method to combine spatial and temporal information in 3D+T datasets and (ii) utilisation of LSTM for sequential information processing in biomedical images.

2 Method

The problem can be defined as follows : given a sequence of t volumes $\{V_1,V_2,V_3, ..., V_t\}$, an output voxel in patient volume $\mathcal{V}$ should be mapped to a binary mask L $\{0,1\}$ which indicates prostate or background. To achieve this, a step-by-step approach based on a deep learning architecture is proposed.

The proposed approach is presented in Fig. 1. Since DCE-MRI is 4-dimensional, using 4D kernels directly is computationally expensive and will lead to parameter explosion, therefore the spatial and temporal components are dealt with separately. On the 3D spatial data, an extension of U-net is applied directly on each volume. Thus, the volumetric output is a sequence of volumes, which was downsampled before fusion with sequential (temporal) outputs. On the temporal front, the volumes of sequences are resampled into voxel sequences, and each voxel sequence is evaluated with five separate LSTM networks to obtain five probability outputs for each voxel. Then the outputs are fused with a Multi-Layer Perceptron to obtain the final segmentation.

2.1 Spatial Volumetric Segmentation based on U-net

The original U-net [13] deals with 2D images, and in this work it is extended to 3D volumetric images. Spatial dropout layers [15] are added to U-net in order to reduce the effects of over-fitting. The work in this section is similar to that of Milletarì et al. [12], except that they replace the pooling with convolutional layers. Two minor modifications to U-net are performed: acceleration with Batch Normalisation Layer and reduction of overfitting with Spatial Dropout Layer. Batch-Normalisation layers [9] are added to regularise training and reduce internal co-variate shifts. The architecture of the modified U-net is shown in Fig 2.2, while Table 2.1 shows the values of the parameters used in the proposed network.

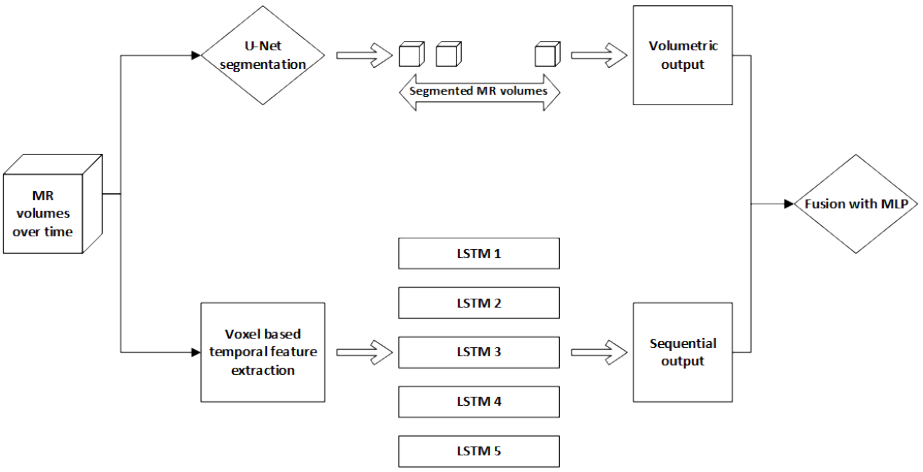

Figure 2.1: Proposed deep fusion architecture where volumetric segmentation is performed with a modified version of U-net, and the temporal segmentation with an ensemble of LTSMs

Layer Type	Kernel Size
Convolution	5*5*5
Max Pooling	2*2*2
Up Sampling	2*2*2
Convolution (mid)	3*3*3
Binarisation	1*1*1

Block	Number of Filters
Down1	32
Down2	64
Down3	128
Down4	256
Down5	512

Block	Number of Filters
Up1	32
Up2	64
Up3	128
Up4	256

Table 2.1: Kernel Size of the Layers, and Number of Filters for Each Block

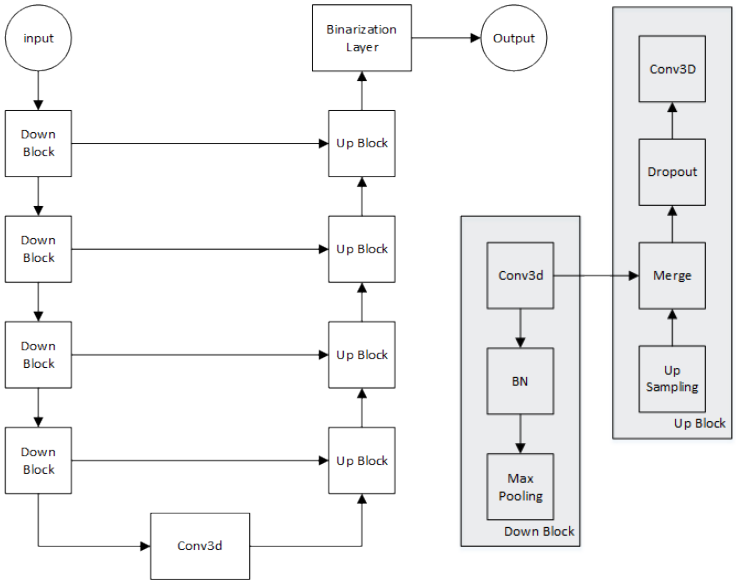

Figure 2.2: Modified U-net Architecture after adding batch normalization layers in Down block and Dropout layers in Up block

2.2 Temporal segmentation based on LSTM

In DCE-MRI, image contrast is enhanced by injection of a gadolinium (Gd)-based contrast agent that decreases the relaxation time of water protons. Therefore, the Region of Interest (RoI) in the image acquired after the injection should show higher intensity values. In Fig. 3, intensity changes over time of a pair of randomly selected prostate (red) and non-prostate(blue) voxels are shown. Based on the difference in contrast over time between the pairs of voxels, common sequential analysis methods may be used to address the segmentation problem. A widely used deep learning architecture for sequential data is the Long Short Term Memory(LSTM) [8], which is now utilised for this purpose.

The temporal segmentation based on LSTM is based only on voxel intensity changes, and the data is managed with manual pre-processing. In order to apply LSTM to the data, two pre-processing steps are required, as shown in Fig 2.4. First, the data intensity is normalised for all patients to the range 0-255, to avoid gaps in contrast for different patients. Secondly, volumetric sequential samples (3D+T) are broken down into voxel sequential samples (1D+T). All prostate voxels are selected as positive samples. All the non-prostate voxels cannot be entirely used as negative samples, as the number of prostate voxels are very few compared to the non-prostate ones, leading to very high data imbalance. As this may lead to poor model learning, a similar number of negative samples to the positive ones are selected instead, by randomly selecting negative voxels near the prostate voxels inside a small 50*50*40 cube. After that, an LSTM network is utilised to perform temporal segmentation over voxel intensity changes. The input is an intensity sequence of length 32 and the intensities range over 0-64, with the original intensity range sequence length decreased to reduce the training load from a limited number of samples. Finally, post-processing is used to improve the performance of the LSTMs, including edge smoothing, removal of noisy voxels by finding the biggest component, and removal of the bladder if included in the output.

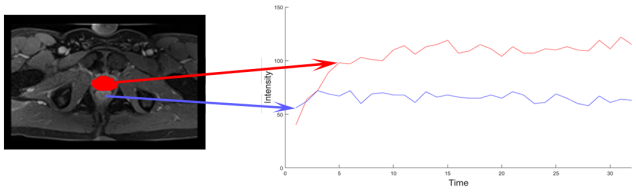

Figure 2.3: Different Behaviours of Intensity Changes Exhibited by Prostate and Non-prostate Regions. Prostate voxel is red, non-prostate voxel is blue

2.3 Fusing Outputs with Multi-layer Perceptron

The outputs from the 3D U-net and the LSTM are fused by an Artificial Neural Network (Multi-Layer Perceptron). In order to achieve this, the difference in output sizes is a large hurdle to overcome. The outputs are volumes in the 3D spatial domain, while they are voxel labels in the 1D temporal domain. In order to fuse them together, the spatial volume outputs are transferred into voxel outputs as shown in Fig 2.1.To make full use of the transitional values extracted,

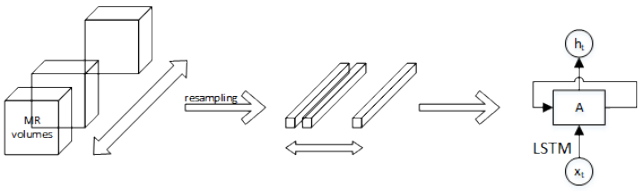

Figure 2.4: Flowchart of the Temporal Segmentation Process

outputs from the volumetric and temporal segmentations are treated as new features to determine the final mixed outputs. The input to the Multi-Layer Perceptron is a feature vector of length of 38, of which 32 are volumetric outputs for each voxel, 5 are sequential outputs and the last feature is the distance of the voxel to the prostate center, extracted from the volumetric output directly. The new feature vector is trained on a Multi-Layer Perceptron of 2 Dense layers with 200 neurons each. The dropouts for the Dense layers are 0.6 and ReLu is used as the activation function. Then a binarisation layer with sigmoid function is added. The network is evaluated with binary cross-entropy and optimised with the Adam optimiser.

3 Experimental Results

3.1 Data

A publicly available dataset from the Initiative for Collaborative Computer Vision Benchmarking [11] is used in this work[1]. This dataset contains T2, DCE and DWI prostate MRI images for 17 patients. The DCE-MRI data includes 40 volumes over time for each patient. Of the 40 time sequences, 32 volumes were selected and the original images were also re-scaled to 192*128*64 for performance speedup reasons. The datasets were split randomly into 15 patients for training and 2 for testing, and the experiments were repeated three times for three different data splits. Results of all the experiments were evaluated using three metrics, namely Dice Similarity Coefficient DSC, sensitivity SEN and Volume Difference VD [20].

3.2 Spatial Volumetric output from U-net

The results are shown in Table 3 in the column labeled as Volumetric Output. As 32 volumes were used for each patient, the total number of volumes used for training and testing are 480 and 64 respectively. The average over 32 volumes per patient for the 2 patients tested is reported as the volumetric output.

3.3 Sequential Output from LSTM networks

Five LSTM models were trained to evaluate the method for temporal segmentation. Taking the variety of negative samples into consideration, all negative samples for each model were re-selected each time, instead of using cross-validation

[1] http://i2cvb.github.io/

	Experiment 1(%)		
	Volumetric Output	LSTM Output	Fused Output
DSC	83.38 ± 0.40	71.02± 3.43	84.97 ± 4.89
SEN	84.81 ± 7.01	73.43 ± 4.69	83.84 ± 7.72
VD	15.42 ± 2.96	5.54 ± 3.66	4.98 ± 4.28
	Experiment 2(%)		
DSC	85.33 ±3.50	77.89 ± 3.20	89.07 ± 0.84
SEN	82.95 ± 8.81	90.15 ± 2.55	88.69 ± 1.17
VD	9.52 ± 9.14	24.39 ± 10.36	0.88 ± 0.78
	Experiment 3(%)		
DSC	77.73 ± 6.74	71.93 ± 2.22	86.58 ± 1.40
SEN	87.93 ± 4.8	80.00 ± 5.60	88.30 ± 3.94
VD	23.35 ± 5.53	17.61 ± 11.35	8.46 ± 5.13
	Average(%)		
DSC	82.15±4.91	73.61±4.24	86.88± 2.96
SEN	85.23±6.71	81.19±8.22	86.94±4.60
VD	16.10±7.94	18.02±13.11	4.77±4.54

Table 3.1: Summary Results of Three Different Experiments Conducted using the Proposed Fusion Pipeline.

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

Figure 3.1: Illustration of post-processing carried out for patient 1 using the axial slice 31: (a) output of LSTM, (b) largest component, (c) removal of the bladder, (d) ground truth, and (e) overlap between the prediction and ground truth

over a single set. The results are shown in Table 3 in the column labelled as LSTM Output. Some illustrative sample outputs are in Figure 5. In some slices, the bladder is also picked as part of the prostate. In these cases, even expert manual observation can barely distinguish any intensity changes between the prostate and bladder. Therefore the bladder is sliced off in the post processing stage, by identifying 20% of the whole length of the identified white regions along the y axis, shown in Figure 5(b).

3.4 Fused output

The results after fusion are also shown in Table 2 in the Fused Output column. From the results, clearly the spatial volumetric output is improved by addition of the temporal segmentation output, with a DSC improvement of 4.73% on average.The LSTM outputs themselves are not satisfactory. This may be due to intra-patient variations, the amount of agent injected and other factors. Nevertheless, the LSTM outputs do improve volumetric only outputs. An example of the results for slices 22, 32 and 42 of one patient is shown in Figure 6. The yellow region (TP) is overlapping parts of the predicted zone and the ground truth zone. The red region represents False Positives (FP), while the green region represents False Negatives (FN).

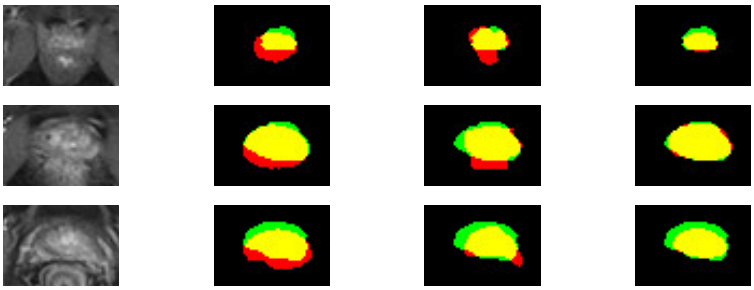

Figure 3.2: Overlap between the Ground Truth and the Prediction for 2D patient scans for a selected patient in Experiment 3, for three different MRI slices. First column shows 22nd, 23rd and 42th prostate slices (top to bottom), and the second column shows the respective volumetric outputs, the third column shows he respective LSTM outputs and the fourth column shows the respective Fused outputs.

4 Discussion and Conclusion

A novel deep fusion pipeline is proposed that utilises both the volumetric and temporal information captured from DCE-MRI sequences. LSTMs are used to handle DCE-MRI based temporal features, while convolutional neural networks are used for volumetric processing. A sequence of ablative testing has been performed to test the efficacy of the proposed method. As can be seen, temporal segmentation is often inferior to spatial segmentation, however with the novel method proposed to combine spatial and temporal information, equal or better results are achieved consistently, compared to spatial segmentation alone.

As an extension of this work, a deep fusion pipeline that can handle both spatial and temporal sequences end-to-end without post-processing is being built. This will ultimately make pre-trained models available to the wider medical imaging community. The proposed methodology for spatio-temporal segmentation can easily be migrated to other biomedical applications where multi-dimensional datasets arise [14].

Bibliography

[1] American Cancer Society cancer ranks. `https://www.cancer.org/cancer/prostate-cancer.html`.

[2] Toan Duc Bui, Jitae Shin, and Taesup Moon. 3d densely convolution networks for volumetric segmentation. *arXiv preprint arXiv:1709.03199*, 2017.

[3] Tyler Clark, Junjie Zhang, Sameer Baig, Alexander Wong, Masoom A Haider, and Farzad Khalvati. Fully automated segmentation of prostate whole gland and transition zone in diffusion-weighted mri using convolutional neural networks. *Journal of Medical Imaging*, 4(4):041307, 2017.

[4] Ahmad Firjani, Ahmed Elnakib, Fahmi Khalifa, G Gimel'Farb, M Abo El-Ghar, J Suri, Adel Elmaghraby, and Ayman El-Baz. A new 3d automatic

segmentation framework for accurate segmentation of prostate from dce-mri. In *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on*, pages 1476–1479. IEEE, 2011.

[5] Yang Gao, Jeff M. Phillips, Yan Zheng, Martin Renqiang Min, P. Thomas Fletcher, and Guido Gerig. Fully convolutional structured lstm networks for joint 4d medical image segmentation. *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 1104–1108, 2018.

[6] Alberto Garcia-Garcia, Sergio Orts-Escolano, Sergiu Oprea, Victor Villena-Martinez, and Jose Garcia-Rodriguez. A review on deep learning techniques applied to semantic segmentation. *arXiv preprint arXiv:1704.06857*, 2017.

[7] Soumya Ghose, Arnau Oliver, Robert Martí, Xavier Lladó, Joan C Vilanova, Jordi Freixenet, Jhimli Mitra, Désiré Sidibé, and Fabrice Meriaudeau. A survey of prostate segmentation methodologies in ultrasound, magnetic resonance and computed tomography images. *Computer methods and programs in biomedicine*, 108(1):262–87, October 2012.

[8] Alex Graves, Abdel-rahman Mohamed, and Geoffrey Hinton. Speech recognition with deep recurrent neural networks. In *Acoustics, speech and signal processing (icassp), 2013 ieee international conference on*, pages 6645–6649. IEEE, 2013.

[9] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.

[10] K Yu Kyle and Hedvig Hricak. Imaging prostate cancer. *Radiologic Clinics of North America*, 38(1):59–85, 2000.

[11] Guillaume Lemaître, Robert Martí, Jordi Freixenet, Joan C. Vilanova, Paul M. Walker, and Fabrice Meriaudeau. Computer-aided detection and diagnosis for prostate cancer based on mono and multi-parametric mri: A review. *Computers in Biology and Medicine*, 60:8 – 31, 2015.

[12] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *3D Vision (3DV), 2016 Fourth International Conference on*, pages 565–571. IEEE, 2016.

[13] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[14] Roland F Schwarz, Charlotte KY Ng, Susanna L Cooke, Scott Newman, Jillian Temple, Anna M Piskorz, Davina Gale, Karen Sayal, Muhammed Murtaza, Peter J Baldwin, et al. Spatial and temporal heterogeneity in high-grade serous ovarian cancer: a phylogenetic analysis. *PLoS medicine*, 12(2):e1001789, 2015.

[15] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1):1929–1958, 2014.

[16] Baris Turkbey, Peter A Pinto, and Peter L Choyke. Imaging techniques for prostate cancer: implications for focal therapy. *Nature Reviews Urology*, 6(4):191–203, 2009.

[17] Sepehr Valipour, Mennatullah Siam, Martin Jagersand, and Nilanjan Ray. Recurrent fully convolutional networks for video segmentation. In *Applications of Computer Vision (WACV), 2017 IEEE Winter Conference on*, pages 29–36. IEEE, 2017.

[18] S. Verma, B. Turkbey, N. Muradyan, A. Rajesh, F. Cornud, M.A. Haider, P.L. Choyke, and M. Harisinghani. Overview of dynamic contrast-enhanced mri in prostate cancer diagnosis and management. *American Journal of Roentgenology*, 198(6):1277–1288, 2012.

[19] Lequan Yu, Xin Yang, Hao Chen, Jing Qin, and Pheng-Ann Heng. Volumetric convnets with mixed residual connections for automated prostate segmentation from 3d mr images. In *AAAI*, pages 66–72, 2017.

[20] Kelly H Zou, Simon K Warfield, Aditya Bharatha, Clare MC Tempany, Michael R Kaus, Steven J Haker, William M Wells III, Ferenc A Jolesz, and Ron Kikinis. Statistical validation of image segmentation quality based on a spatial overlap index1: scientific reports. *Academic radiology*, 11(2):178–189, 2004.