PROCEEDINGS OF SPIE

Automatic slice segmentation of intraoperative transrectal ultrasound images using convolutional neural networks

Nooshin Ghavami, Yipeng Hu, Ester Bonmati, Rachael Rodell, Eli Gibson, et al.

SPIE.

Automatic Slice Segmentation of Intraoperative Transrectal Ultrasound Images using Convolutional Neural Networks

Nooshin Ghavami[1], Yipeng Hu[1], Ester Bonmati[1], Rachael Rodell[1], Eli Gibson[1], Caroline Moore[2, 3], Dean Barratt[1]

[1]UCL Centre for Medical Image Computing, Department of Medical Physics and Biomedical Engineering, University College London, London, UK
nooshin.ghavami.15@ucl.ac.uk
[2]Division of Surgery & Interventional Science, University College London, UK
[3]Department of Urology, University College London Hospitals NHS Foundation Trust, London, UK

ABSTRACT

Clinically important targets for ultrasound-guided prostate biopsy and prostate cancer focal therapy can be defined on MRI. However, localizing these targets on transrectal ultrasound (TRUS) remains challenging. Automatic segmentation of the prostate on intraoperative TRUS images is an important step towards automating most MRI-TRUS image registration workflows so that they become more acceptable in clinical practice. In this paper, we propose a deep learning method using convolutional neural networks (CNNs) for automatic prostate segmentation in 2D TRUS slices and 3D TRUS volumes. The method was evaluated on a clinical cohort of 110 patients who underwent TRUS-guided targeted biopsy. Segmentation accuracy was measured by comparison to manual prostate segmentation in 2D on 4055 TRUS images and in 3D on the corresponding 110 volumes, in a 10-fold patient-level cross validation. The proposed method achieved a mean 2D Dice score coefficient (DSC) of 0.91±0.12 and a mean absolute boundary segmentation error of 1.23±1.46mm. Dice scores (0.91±0.04) were also calculated for 3D volumes on the patient level. These suggest a promising approach to aid a wide range of TRUS-guided prostate cancer procedures needing multimodality data fusion.

Keywords: Convolutional Neural Networks, Prostate Cancer, Segmentation, Transrectal Ultrasound, Registration

1   INTRODUCTION

Prostate cancer is the most commonly diagnosed non-cutaneous cancer in men in many parts of the Western World, and is a major cause of cancer-related death internationally. It is now widely acknowledged that the widespread use of prostate specific antigen (PSA) screening has led to an over-diagnosis of low-risk prostate cancer, and over-treatment of patients with low-to-intermediate risk with conventional radical radiation therapy or prostatectomy that carry a significant risk of side-effects is a source of concern [1].

Consequently, accurate patient stratification and less invasive treatments are both critical to improving prostate cancer care. These goals are supported by transrectal-ultrasound (TRUS)-guided biopsies and therapies, respectively. In addition to real-time visualisation of instruments such as needles as well as anatomical structures, TRUS images also enable visualization of real-time deformations of the prostate during procedures at low cost. However clinically important tumours are frequently not visible on ultrasound images. Fusion of pre-procedural imaging modalities, such as multi-parametric MRI (mpMRI), with intraoperative TRUS is therefore necessary to localize tumours [2]. Sankineni et al. [2] showed that in 26% of patients with prostate cancer, TRUS-MRI fusion-guided biopsy detected the cancer, whereas the systematic 12-core sextant biopsy did not.

Automatic registering of mpMRI visible targets with intraoperative TRUS images remains challenging due to differences in underlying imaging generation processes between different modalities, substantial tissue deformation and patient

motion. A common technique employed in most commercial guidance systems, requires segmentations of the prostate gland on both modalities to define their relative alignment. However, time constraints on the clinical workflow mean that accurate manual segmentation of the prostate boundary on intraoperative imaging can be very difficult. In addition, manually updating segmentations to account for real-time soft-tissue deformation during procedures places a significant additional time burden and therefore is rarely done in practice. Furthermore, intra- and inter-observer variability in manual delineation of the prostate can reduce the accuracy and reproducibility of the alignment. Having an automated method for localising and segmenting the prostate allows this variability to be reduced [2].

Deep convolutional neural networks are a promising technique for automated segmentations of TRUS images of the prostate. They have yielded high accuracies for automatic prostate segmentations and cancer detection from MR images [3-5], using a variety of network architectures. Liu et al. [3], used an architecture called XmasNet for cancer classification and achieved high area under the curve (AUC) values of 0.84 on the testing data and outperformed 69 different methods. Tsehay et al. [4] used a deep learning CNN adopted from a state-of-the-art edge detector which outputs image probability maps. The efficacy of their network was compared with an existing prostate computer aided design (CAD) system based on hand-crafted features and their method produced a detection rate of 86% while the existing system had a detection rate of 80%. Finally Liao et al.[5] used an independent subspace analysis (ISA) network using representation learning to learn the most effective features in a hierarchical and unsupervised manner. Their method achieved an average dice score of 86.7±2.2 which was higher than widely-used hand crafted features and state-of-the-art segmentation methods.

However, for TRUS prostate images, even though there has been work for automatic prostate segmentation, most studies in this field have used either feature based methods with support vector machines (SVM) [6-7] or statistical shape models [8]. However, since CNN approaches have shown high accuracy when applied to MR images of the prostate, the aim of this work is to test and evaluate this methodology on TRUS images. To the best of our knowledge only one paper has evaluated a CNN for whole prostate segmentation, on a limited cohort of ultrasound images from 17 patients cropped to prostate volume [9].

In this paper, we evaluate the accuracy of a CNN-based method for automatic prostate segmentation on clinically acquired US images from 110 patients. Using a 10-fold cross-validation, we quantify the segmentation accuracy of both 3D volumes (to support segmentation-based registration) and 2D segmentation of sagittal slices (to support intra-procedural motion tracking).

2 METHODS

2.1 Data and Pre-processing

The TRUS images used in this work were acquired from the SmartTarget Biopsy Trial [10], consisting of 110 patients requiring targeted transperineal biopsies. For each patient, 38-177 para-sagittal slices were acquired with a digital stepper in a continuous rotation to avoid translational movement of the prostate gland during acquisition. This study used 4055 2D slices presenting a subset of each acquired volume: up to 59 slices per volume sampled at 3-degree intervals, with a pixel size of 0.18×0.16mm and an image size of 576×720 pixels.

Manual segmentation of the prostate (excluding seminal vesicles) was carried out using ITK-SNAP [11] by two experienced medical imaging research staff. Segmentation took between 20-30 minutes per volume to complete. These segmentations were used for training the proposed neural network as well as for evaluating the automatic segmentation results in the cross-validation experiment, where these manual segmentations were considered as the ground truth.

2.2 Automatic Segmentation using CNN Algorithm

The proposed algorithm uses a fully-convolutional neural network, based on an adapted U-network architecture [12] shown in **Figure 1**. The network takes an ultrasound slice of size S_0 as input and propagates to feature maps of the same size and n_0 (n_0=32 here) initial channels, by a convolution (Conv), a batch normalisation (BN) and a nonlinear rectified linear unit (ReLU). The feature maps are then down-sampled to K different resolution levels by *down-sampling* blocks and each followed by a residual network unit (Resnet) block. At each k (k=1….K) level, the number of channels n_k is doubled and size S_k is halved. Each down-sampling block consists of a troika of Conv, BN and ReLU, followed by a max-pooling layer with stride 2; while each Resnet block has two Conv layers with BN and ReLU, with an identity shortcut over these layers. The up-sampling blocks reverse the down-sampling process using transpose convolution

layers with stride 2, replacing the max-pooling layers, to output an image sized logits layer to represent the segmentation. Summation shortcuts are added before each down-sampling block to the output feature maps from each compatibly-sized up-sampling block.
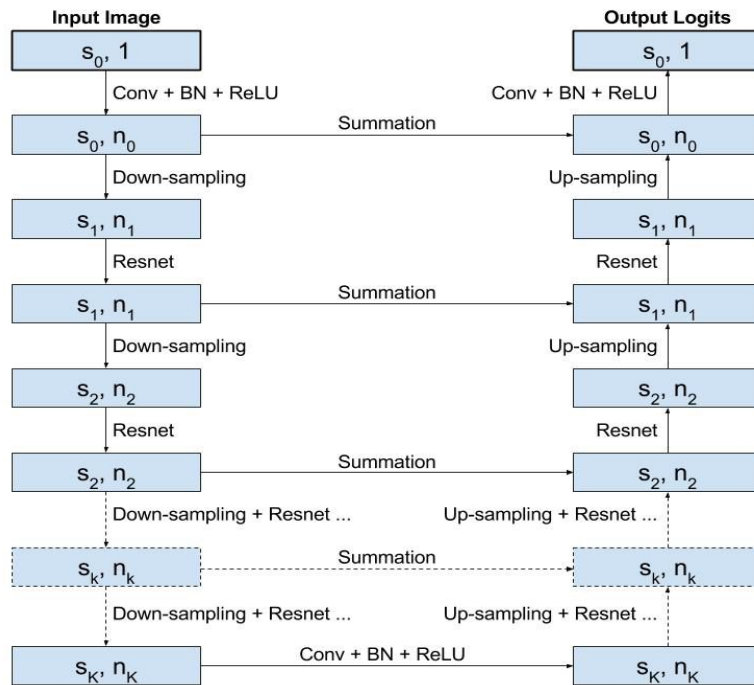

Figure 1: Architecture of the convolutional neural network.

2.3 Validation

The network were implemented in Tensorflow [TM] and trained on a NVIDIA® TITAN Xp GPU with 12 GB memory, using the Adam optimiser with 64 images in each minibatch. The results presented in this work were obtained by minimising a negative probabilistic Dice score that is differentiable with an added L^2-norm weight decay on the trainable parameters, the weighting parameter being set to 1e-6.

A 10-fold cross validation was carried out on a patient level in order to avoid bias and overfitting. Specifically, in the first fold, images from 11 patients were used to test the network performance and images from the remaining patients were used to train the network; in the second fold, images from another 11 patients were used for testing, and so on until all 110 patients were used for evaluation once. For each fold, the automatic segmentations were compared to the manually segmented images (ground-truth) using both the Dice similarity coefficient (DSC) and boundary distance. The boundary distance was defined as the mean absolute value of the distances between all the points from the automatically segmented boundary and the closest boundary points found on the left-out ground-truth.

3 RESULTS

3.1 Segmentation outputs-Qualitative Results
Figure 2 shows a comparison of the automatic segmentations with the manually segmented images for four different slices (**A-D**). The results show a good qualitative agreement between these segmentations.
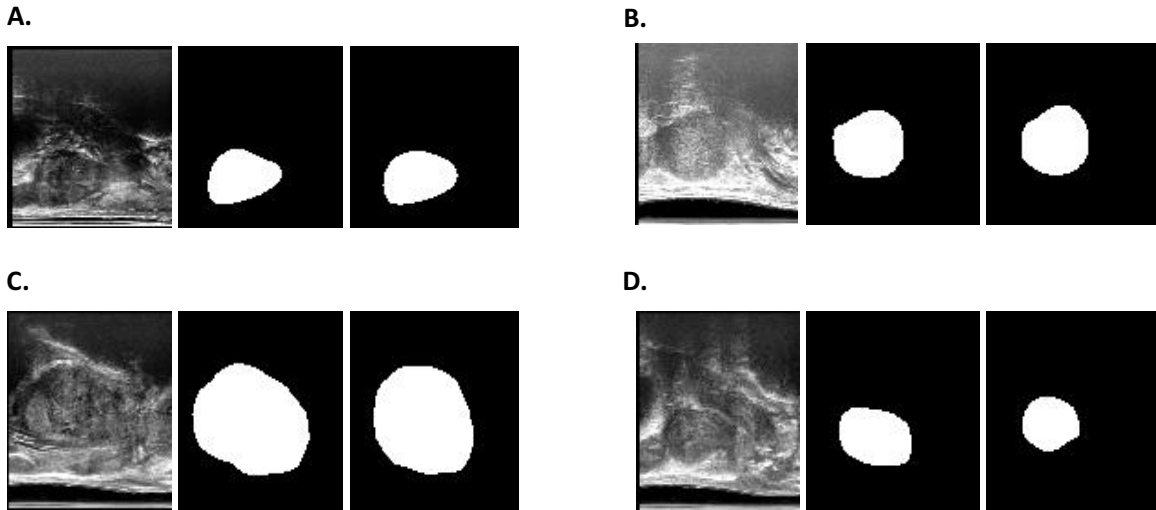
Figure 2: Examples of the manual and automatic prostate segmentations. The manually segmented prostate is shown in the middle, followed by the automatic segmentations on the right-hand side shown in 4 different example slices (A-D) chosen at random. These slices are representative from all the images segmented.

Figure 3 shows the results of the automatic segmentation for two slices (**A-B**) over the different iterations during training. In the first slice shown (**A,** original image) the shape of the prostate is quite visible on the original image and therefore the network started to converge to ground-truth segmentation after around 16400 iterations. For the second slice shown however (**B,** original image), there is a small part to the left of the prostate which the network seems to associate with being prostate even up to 18600 iterations. This evolution can also be seen from **Figure 4** which shows the loss function against the number of iterations for the first fold of the testing. Even though the loss function stabilizes after around 5000 iterations, there is still some peaks and troughs due to these very different evolutions of the automatic segmentations.
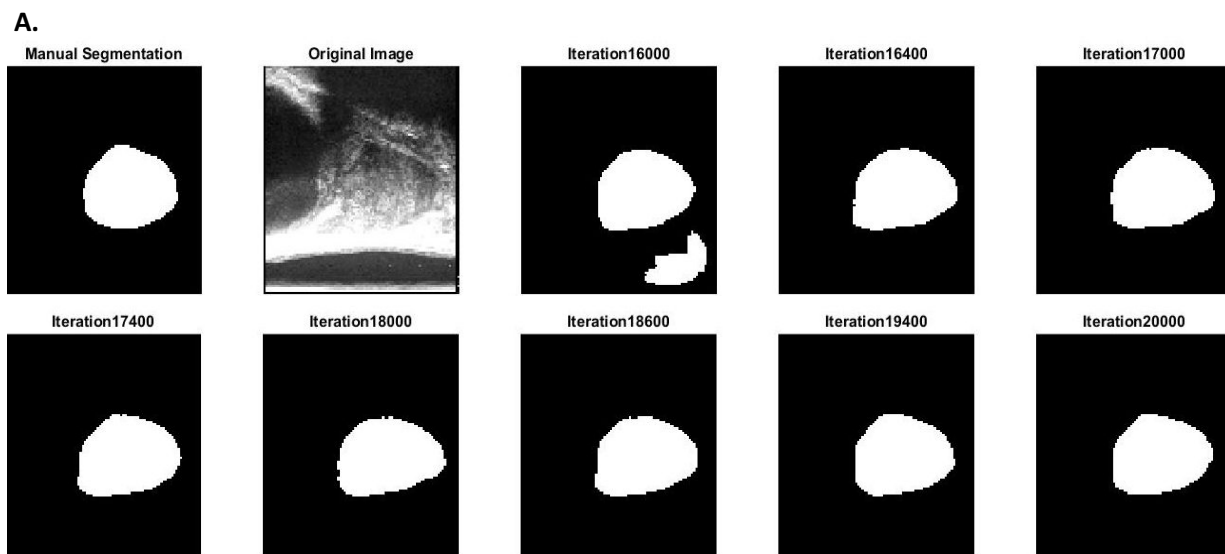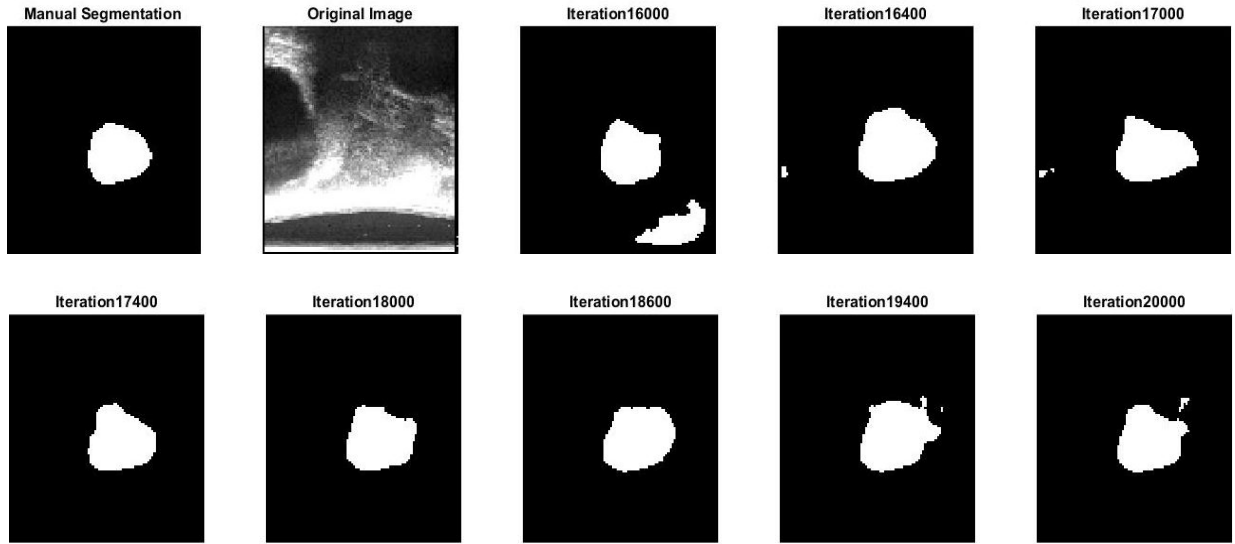
B.

Figure 3: Evolution of the automatic segmentation over number of iterations for two different slices (A-B). We can see that depending on the prostate size and shape, the automatic segmentations stabilize after a different number of iterations. The manual segmentations shown in these images were done by observer 1.
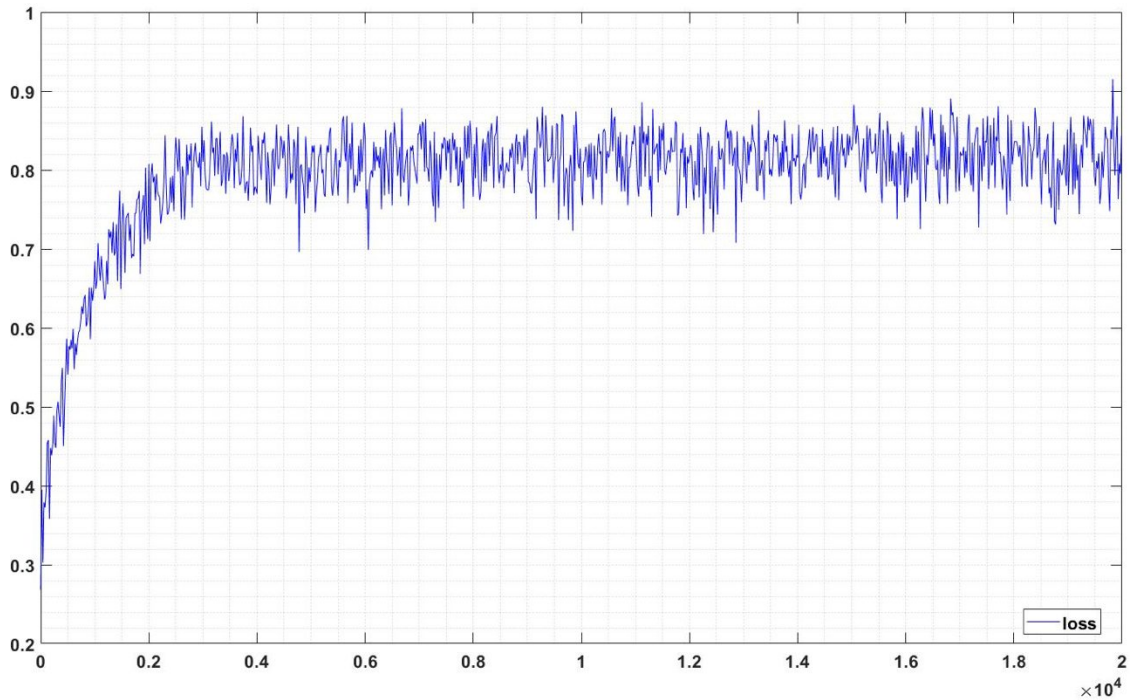

Figure 4: Plot of the loss function through the different iterations. The image shows that after around 5000 iterations the loss function becomes stable. However even after these many iterations there is still some variations in the loss function and this is due to the large differences in the prostate images themselves as shown in the previous figure.

3.2 Similarity measures-Quantitative Results

The DSC and boundary distances were calculated between the automatic and manual prostate segmentations for each fold of the cross validation. The proposed segmentation method achieved mean (±standard deviation (SD)) DSC of 0.91±0.12 over all cross-validation folds. The average boundary distance is 0.38±0.36mm. Dice scores (0.91±0.04) were also calculated for 3D volumes on the patient level. This was done by only taking into account the slices used in the testing in each fold and calculating the average dice score over the patients from which these slices are taken from only. This is computed on volumes reconstructed with the slices from the individual patient, which arguably is more relevant to the registration application of interest. The mean absolute distance appears on the same order of magnitude of other ultrasound segmentation methods [6-7] which reported results of 0.81 and 1.12 respectively. Unfortunately, lacking access of these detailed results and data prevented a direct statistical comparison.

3.3 Comparison between different observers

As mentioned in the methods, two different observers carried out the manual segmentations and the neural network was trained separately on each of the two sets of segmentations. **Table 1** shows the results of the comparison between the manual and automatic segmentations from each of the two observers. Moreover, the manual segmentations from the two observers were also compared with each other to see how much inter-observer variability existed. The DSC between the two observers was calculate to be 0.92±0.06.

Table 1: Comparison of the different similarity measures obtained between automatic and manual segmentations done by two different observers.

Observer	2D Dice Scores	3D Dice Scores	Boundary Distance (mm)
1	0.91±0.12	0.91±0.04	1.23±1.46
2	0.89±0.11	0.89±0.04	1.55±1.48

4   CONCLUSION

In this paper, we propose a deep learning CNN algorithm for fully automatic prostate segmentation from US images. Results obtained from the network showed convergence of the loss function after a few thousand iterations and good agreement with the manually segmented images in terms of both visual assessment and similarity measures, giving competitive quantitative results on both 2D and 3D cases. Our work gives comparable results to the work carried out by Yang et al [8], 0.91 Vs 0.92, validated on a much larger dataset. Also, our method achieves DSC that are very close to the DSC computed in the inter-observer experiment which is very good. These results should be considered in the context of a few limitations of the presented work. First, we did not extensively tune the hyperparameters in the proposed CNN architecture, and therefore our results may underestimate the potential accuracy of the proposed method. Second the data used in this study was acquired at a single center, precluding generalisation about its performance. Finally, our results have not been compared statistically with other existing algorithms.

Future work will focus on improving the accuracy, by adding a temporal or spatial prior to take into account 3D information, e.g. slice positions. When the segmentations are done manually, the observer looks at the slices around the current slice as well in order to determine exactly where the prostate is, and therefore adding this 3D information into the network should also hopefully improve the results. Another extension would be to have the manual segmentations done by more observers, especially more experienced clinicians, to have better reference standard.

5   ACKNOWLEDGEMENTS

6 REFERENCES

1. Klotz, L., & Emberton, M. (2014). Management of low risk prostate cancer-active surveillance and focal therapy. *Nature Reviews Clinical Oncology* , pp. 324-334.

2. Sankineni, S., *et al*. (2015). Posterior subcapsular prostate cancer: identification with mpMRI and MRI/TRUS fusion-guided biopsy. *Abdominal Imaging*, 40(7), pp. 2557-2565.

3. Liu, S., Zheng, H., Li, W. (2017). Prostate cancer diagnosis using deep learning with 3D multiparametric MRI. *Proc. SPIE* 10134, Medical Imaging 2017: Computer-Aided Diagnosis, 1013428.

4. Tsehay, Y., *et al*. (2017). Convolutional neural network based deep-learning architecture for prostate cancer detection on multiparametric magnetic resonance images. *Proc. SPIE* 10134, Medical Imaging 2017: Computer-Aided Diagnosis, 1013405

5. Liao, S., *et al*. (2013). Representation learning: a unified deep learning framework for automatic prostate MR segmentation. *MICCAI 2013, Part II. LNCS*, 8150, pp. 254–261

6. Zhan, Y., Shen, D. (2003). Automated Segmentation of 3D US Prostate Images Using Statistical Texture-Based Matching Method. *MICCAI 2003*. *LNCS*, 2878, pp.688-696.

7. Zhan, Y., Shen, D. (2006). Deformable segmentation of 3-D ultrasound prostate images using statistical texture matching method. *IEEE Transactions on Medical Imaging*, 25(3), pp.256-272

8. Shen, D., *et al*. (2003). Segmentation of Prostate Boundaries From Ultrasound Images Using Statistical Shape Model. *IEEE Transactions on Medical Imaging*. 22(4), pp. 539-551

9. Yang, X., et al. (2016). Fine-Grained Recurrent Neural Networks for Automatic Prostate Segmentation in Ultrasound Images. *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17)*

10. Donaldson, I., et al. (2017). MP33-20 THE SMARTTARGET BIOPSY TRIAL: A PROSPECTIVE PAIRED BLINDED TRIAL WITH RANDOMISATION TO COMPARE VISUAL-ESTIMATION AND IMAGE-FUSION TARGETED PROSTATE BIOPSIES. *The Journal of Urology.*197(4), p.425

11. Paul A., et al. (2006). User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *Neuroimage*. 31(3), pp.1116-28.

12. Ronneberger, O., Fischer, P. and Brox, T., (2015). U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical Image C omputing and Computer-Assisted Intervention* (pp. 234-241). Springer