Program Digest, 2019 IEEE International Ultrasonics Symposium (IUS)
Glasgow, Scotland, October 6-9, 2019

3D Convolutional Neural Network for Segmentation of the Urethra in Volumetric Ultrasound of the Pelvic Floor.

Helena Williams*‡§, Laura Cattani*, Wenqi Li†§, Mahdi Tabassian‡,
Tom Vercauteren§, Jan Deprest*, Jan D'hooge‡
* Department of Obstetrics and Gynaecology
Department of Development and Regeneration, Urogenital, Abdominal and Plastic Surgery
KU Leuven, Leuven, Belgium
{helena.williams | laura.cattani | jan.deprest}@uzleuven.be
†NVIDIA Corporation, Cambridge, United Kingdom
wenqil@nvidia.com
‡Department of Cardiovascular Sciences, KU Leuven, Leuven,Belgium
{mahdi.tabassian | jan.dhooge}@uzleuven.be
§School of Biomedical Engineering & Imaging Sciences, King's College London, United Kingdom
{tom.vercauteren|helena.williams}@kcl.ac.uk

***Abstract*—Pelvic organ prolapse (POP) decreases the quality of life for many women. To assess POP, the levator hiatus is segmented in a 2D plane of minimal hiatal dimensions, known as the C-plane. In order to automate plane detection, landmark information of key structures should be given to a plane detection algorithm. In this work, we present a fully automatic method to segment the urethra from a 3D transperineal ultrasound volume using a convolutional neural network (CNN). A dataset with 35 volumes from 20 patients during the Valsalva manoeuver (i.e. Valsalva, contraction and rest) labelled by an expert, was used for training and evaluation in a 5-fold cross-validation process. The 3D CNN model yielded an average robust Hausdorff distance of 4.68mm (95 percentile) which was comparable to intra-observer results.**

***Index Terms*—Pelvic floor, 3D Convolutional neural network, Semantic segmentation**

I. Introduction

Pelvic organ prolapse (POP) is a type of pelvic floor disorder that decreases quality of life of many women. POP refers to the downward descent of pelvic floor organs such as bladder, vagina, small bowel and rectum, through the genital hiatus. To assess the severity of POP, the levator hiatus is segmented in a 2D plane of minimal hiatal dimensions (referred to as the C-plane) from a 3D transperineal ultrasound volume and the biometrics extracted from the levator hiatus are analysed [1]. Detection of the C-plane, however, is a manual and labour-intensive task, prone to observer variability. Hereto, the aim of this study was to segment the urethra in a 3D volume in order to generate context for a future C-plane detection algorithm.

To our knowledge the urethra has not been automatically segmented in transperineal ultrasound volumes prior to this study. However, the levator hiatus has been segmented from the C-plane extracted from a transperineal ultrasound volume in the work by Bonmati et al. [2]. In this paper they used convolutional neural networks (CNNs) for semantic segmentation. The method achieved a Dice score of 0.90 and results were comparable to previous 'state-of the-art' work using active shape models from B-splines by N. Sindhwani et al. [3].

Given the fact that CNNs have become the gold standard for semantic segmentation without the need for manual initialisation, a 3D CNN was used in this study for automatic segmentation of the urethra. Unlike the segmentation of the levator hiatus, in this work we have a strongly unbalanced task, i.e. the urethra volume foreground pixels in comparison to the background pixels is extremely small and unbalanced. Thus, in this work, we used a label-based approach to weight sampling during training. Due to transperineal ultrasound having relatively low image quality, we also employed pre- and post-processing techniques to enhance results such as specific data augmentation and connected component analysis. We used a 5-fold cross validation to train and validate the performance of the CNN using transperineal ultrasound volumes. Each volume was segmented by a clinical expert and results yielded were compared with intra-observer results.

II. Methods

A. Data

A dataset of 35 transperineal ultrasound volumes from 20 patients was collected. The ultrasound volumes were acquired during the Valsalva manoeuvre (i.e. full Valsalva, rest and contraction). The dataset consisted of 16 volumes acquired at rest, 14 volumes at contraction and 5 at Valsalva. The small sample of Valsalva volumes was due to the exclusion criteria of not considering volumes where a prolapsed bladder was present. A prolapsed bladder severely impacts the shape of the urethra and would make training difficult with a small dataset, thus we start our research by concentrating on 'healthy' urethras. All 35 ultrasound volumes were manually segmented using GE Vocal software by one clinical expert.

B. Network architecture

The network used in this work to segment the urethra was HighRes3DNet by Li et al. and implemented using NiftyNet – a Tensorflow based package for medical imaging segmentation [5].

The network architecture, as shown in Fig. 1, uses multiple layers of dilated convolutions and residual connections to map the input volume to a voxel-level segmentation. The network architecture allows for multi-scale visual feature extraction due to the dilation factor of dilated convolutions that is increased as layer depth increases, which leads to an increase of the receptive field size. Residual blocks with identity mapping are used to group every two convolutional layers. This design improves training speed and ensures the propagation of information during training is smooth. Thus, the network has a relatively large effective receptive field; can generate high resolution feature maps; and can be trained efficiently. For more information on this architecture please refer to the original paper [4].

C. Training

Data augmentation is necessary when the training dataset is of limited size. This process increases the number of training volumes for the CNN to learn from. Therefore, the data augmentation must be representative of possible 'new' data. In this work we used randomised intensity scaling and whitening as the data augmentation. The maximum and minimum scaling percentage was set to 10%.

The training of our model was implemented using NiftyNet on a desktop with a 24GB NVIDIA Quadro P6000. NiftyNet is a Tensorflow-based program that is designed to facilitate patch-based medical image analysis [5]. Due to GPU limitations, the original transperineal ultrasound volume was too large to load into one sampling window without heavy down sampling. As such, a patch-based segmentation was used and we had to make sure that the urethra was sampled more than the background voxels as to ensure the network learnt key features related to the urethra.

To accomplish this we used weighted sampling where the ground truth mask provided information for sampling. The patches supplied as input to the network were weighted on the urethra ground truth position during training. Weighted sampling was used to make the problem more balanced than the original highly unbalanced problem, thus it improved network performance and efficiency [6-8]. The loss function used was Dice with the definition given by Milletari, et al. this Dice loss layer performed well for unbalanced tasks. For more information please refer to [9].

D. Evaluation

The best performing model was found by selecting the epoch with the best Dice loss on the validation set. Evaluation was performed in a 5-fold cross validation, in which the 5 models were obtained by training the network 5 times, with a different set of volumes for training, validation and inference. Thus, after all 5 cross validation models had been trained, 35 automatic 3D urethra segmentations were obtained.

E. Post-processing

For each automatic segmentation obtained, segmentation post-processing morphological operators were applied. To fill holes a flood-fill operation on background pixels that are enclosed by foreground pixels was applied and unconnected collections of pixels were removed by selecting the largest component using connected component analysis as shown in Fig. 2.

F. Metrics

Segmentation results were evaluated using a region-based metric (i.e. the Dice coefficient) and a surface distance-based metric (i.e. robust Hausdorff distance). The Hausdorff distance, h, was used to measure the upper limit of the incorrect positioning, as it measures a distance between two sets of edge points from the ground truth 3D mask, **A** and the output 3D mask generated from the network, **B**.

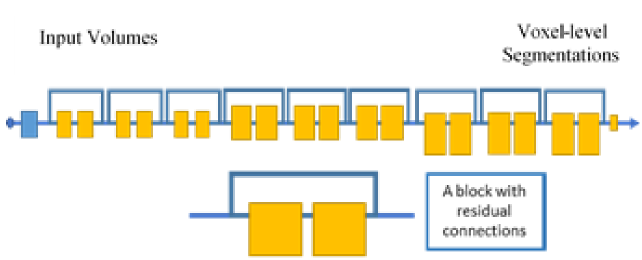

Fig. 1. The network architecture used for volumetric image segmentation in this paper. The network consists of dilated convolutions and residual connections. The architecture allows for multi-scale visual feature extraction as the dilation factor of dilated convolutions increases as layer depth increases. For more detail please refer to work by Li et al. [4].

Fig. 2. Post processing effect on network output, A) shows the initial output of the network and B) shows the filtered result after connected region filtering and filling holes operations are applied.

The original Hausdorff distance is highly sensitive to noise and outliers and is defined as:

$$h(\mathbf{A}, \mathbf{B}) = \max_{\forall p \in \mathbf{A}} \min_{\forall q \in \mathbf{B}} \parallel \mathbf{p} - \mathbf{q} \parallel \tag{1}$$

Having a small Hausdorff distance indicates a good approximation i.e. for a point in **A** there is a point on **B** within radius defined in (1) [10]. Unlike Hausdorff distance, Robust Hausdorff distance is less sensitive to outliers. 'Robust' means it uses the 'Kth' percentile of the distances and not the maximum distance [10]. It is defined as:

$$h(\mathbf{A}, \mathbf{B}) = K^{th}_{a \in \mathbf{A}} d_{\mathbf{B}}(a), \tag{2}$$

where $d_{\mathbf{B}}(a)$ denotes the minimum distance at position a to the position set **B**, and $K^{th}_{a \in \mathbf{A}}$ represents the Kth ranked value of $d_{\mathbf{B}}(a)$ [11]. In our evaluation we use the 95th percentile. Dice expresses the overlap between the ground truth 3D mask, **A** and network output 3D mask, **B**. This metric is on a pixel-wise basis and defined as:

$$D(\mathbf{A}, \mathbf{B}) = \frac{2|\mathbf{A} \cap \mathbf{B}|}{|\mathbf{A}| + |\mathbf{B}|}, \tag{3}$$

where $|\mathbf{A} \cap \mathbf{B}|$ is the overlap of pixels between the ground truth 3D mask, **A** and network output 3D mask, **B** and $|\mathbf{A}| + |\mathbf{B}|$ is the union of pixels (i.e. the total sum of pixels from the ground truth 3D mask, **A** and network output 3D mask, **B**).

G. Intra-observer analysis

Evaluation was recorded by measuring the difference between computer and observer. The manual and automatic segmentation results were compared to retrieve computer-to-observer performance metrics. These results were compared to our 'gold' standard of the expert intra-observer variability. To attain metrics for intra-observer variability we asked the expert to segment the urethra a second time on a random selection of five transperineal ultrasound volumes.

III. Results

Results of the performance metrics averaged over all folds can be seen in Table I. Intra-observer variability using the same performance metrics as Table I is shown in Table II. Fig. 3 shows an example of an automatic segmentation overlaid on the corresponding input transperineal ultrasound volume. Fig. 4 Shows the corresponding overlap between the manually segmented urethras and the automatic segmented urethras for three different transperineal ultrasound volumes.

TABLE I
TABLE I. AVERAGE PERFORMANCE OF THE HIGHRES3DNET BY EMPLOYING COMPARISON WITH THE MANUAL SEGMENTED LABELS FOR EACH ULTRASOUND VOLUME. STANDARD DEVIATION OF RESULTS GIVEN IN SQUARE BRACKETS.

Robust Hausdorff 95 percentile (in mm) [SD]	*Standard Hausdorff Distance (in mm) [SD]*	*Dice [SD]*
4.68[0.49]	7.56[1.65]	0.65[0.08]

TABLE II
TABLE II. AVERAGE INTRA-RATER VARIABILITY BETWEEN 5 MANUALLY SEGMENTED URETHRAS ON A TEST-RETEST BASIS. STANDARD DEVIATION OF RESULTS GIVEN IN SQUARE BRACKETS.

Robust Hausdorff 95 percentile (in mm) [SD]	*Standard Hausdorff Distance (in mm) [SD]*	*Dice [SD]*
6.84[9.68]	9.25[10.65]	0.60[0.12]

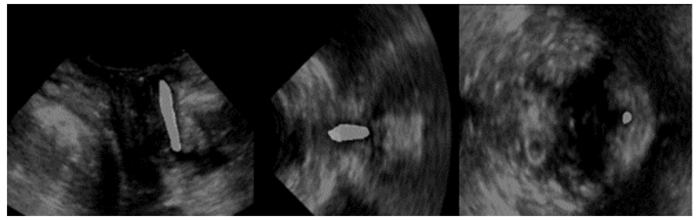

Fig. 3. Shows a deep learning network result (white) superimposed on top of the input transperineal ultrasound volume.

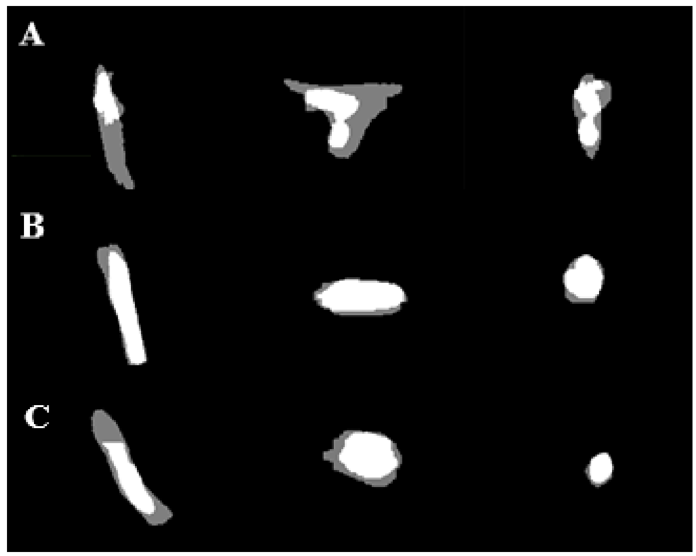

Fig. 4. Shows the deep learning network result (white) superimposed over the corresponding ground truth mask for three separate volumes. A) represents the lowest performing result (Hausdorff 95th percentile- 11.4mm and Dice loss – 0.40) B) represents the highest performing result (Hausdorff 95th percentile- 2.6mm and Dice loss- 0.78) C) represents an average performing result (Hausdorff 95th percentile- 5.09 mmm and Dice loss- 0.62).

IV. Discussion

Manual segmentation of transperineal ultrasound volumes in 3D can be challenging due to noise, low resolution and observer variability. We have presented a fully automatic method using a CNN, to segment the urethra in 3D from a transperineal ultrasound volume. This work will be used to develop an automatic plane detection algorithm by supplying landmark knowledge via semantic segmentation. As manual C-plane detection is a time-consuming task with observer variability, this future work will improve and speed up the pelvic floor assessment process for many women.

In this paper, we used HighRes3DNet, a powerful network for semantic segmentation [4]. It is to our knowledge the first time this network has been used on transperineal ultrasound volumetric data. The implementation of the network and training procedure had to be adapted for this strongly unbalanced problem (i.e. small foreground to background pixel ratio). To solve this, we sampled our network during training in the region of the ground truth foreground and used a Dice loss layer.

In this work we compare our automatic segmentation approach against the observer's own variability calculated by having the observer re-segment 5 random transperineal ultrasound volumes. Comparing the computer-to-observer difference (Table I) with intra-observer differences (Table II), we show that the automatic approach performs better than the manual segmentation retest.

Surprisingly the standard deviation for intra-observer variability is 10 times larger than the CNNs standard deviation. When all re-tests from the observer were investigated it was highlighted that this result was due to one volume having a particular high standard Hausdorff distance of 32.01mm and 95th percentile Robust Hausdorff distance of 26.45 mm. Therefore when this result was excluded from the intra-observer variability performance metrics, the new standard Hausdorff distance was 4.56 [1.91] mm and 95th percentile Robust Hausdorff distance was 2.86 [1.51] mm. The high Hausdorff Distances were due to an error during manual segmentation of one volume using GE Vocal software, this highlights an issue with the urethra 3D segmentation protocol, thus in future work we aim to check the quality of all manual segmentations before training and have several observers performing manual segmentation on a proportion of the data to generate inter-observer variability performance metrics.

V. Conclusion

In this research, we present an automatic approach using a convolutional neural network to segment the urethra from a 3D transperineal ultrasound volume. The task was strongly unbalanced and to overcome this we implemented several techniques to ensure features from the urethra were learnt during training. The performance metrics show the automatic approach is comparable to manual segmentation by comparison of observer-to-computer and observer-to-observer performance metrics.

Acknowledgment

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the P6000 GPU used for this research and GE for their support during research and contribution towards creating the dataset.

References

[1] AIUM/IUGA Practice parameter for the performance of urogynecological ultrasound examinations: developed in collaboration with the ACR, the AUGS, the AUA, and the SRU. Journal of Ultrasound in Medicine. 2019;38(4):851-64.

[2] E. Bonmati, Y. Hu, N. Sindhwani, H. P. Dietz, J. D'hooge, D. Barratt, J. Deprest, T. Vercauteren, 'Automatic segmentation method of pelvic floor levator hiatus in ultrasound using a self-normalising neural network' Journal of Medical Imaging, vol. 5, 2017.

[3] N. Sindhwani, D. Barbosa, M. Alessandrini, B. Heyde, H. Dietz, J. D'hooge, J. Deprest. 'Semi-automatic outlining of levator hiatus' Ultrasound in Obstetrics & Gynecology, vol. 48, Sep 2015.

[4] W. Li, G. Wang, L. Fidon, S. Ourselin, M. J. Cardoso, T. Vercauteren. 'On the Compactness, Efficiency, and Representation of 3D Convolutional Networks: Brain Parcellation as a Pretext Task', 2017

[5] E. Gibson, W. Li, C. Sudre, L. Fidon et al. 'NiftyNet: a deep-learning platform for medical imaging', Computer Methods and Programs in Biomedicine, vol.158, Sep 2017

[6] K. Kamnitsas, C. Ledig, V. F. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, 'Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation', Medical image analysis, vol. 36, pp. 61–78,2017

[7] L. Berger, E. Hyde, M. J. Cardoso, S. Ourselin, 'An adaptive sampling scheme to efficiently train fully convolutional networks for semantic segmentation', 2017.

[8] C. Sudre, W. Li, T. Vercauteren, S. Ourselin, M. J. Cardoso ' Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations' 2017.

[9] F. Milletari, N. Navab, S. Ahmadi, 'V-net. fully convolutional neural networks for volumetric medical image segmentation', 2016

[10] Y. Kang, M. Kyung, S. Yoon, M. Kim, 'Fast and robust hausdorff distance computation from triangle mesh to quad mesh in near-zero cases', Computer Aided Geometric Design, vol. 62, March 2018.

[11] D. Sim, O. Kwon, R. Park ' Object matching algorithms using robust hausdorff distance measures', IEEE transactions on image processing, vol. 8 March 1999.