Automatic C-plane detection in pelvic floor transperineal volumetric ultrasound

Helena Williams[1,2,4], Laura Cattani[1], Mohammad Yaqub[3], Carole Sudre[2], Tom Vercauteren[2], Jan Deprest[1], and Jan D'hooge[4]

[1] Department of Obstetrics and Gynaecology, University Hospitals Leuven, Belgium
[2] School of Biomedical Engineering & Imaging Sciences, King's College London, UK
[3] Department of Computer Vision, Mohamed bin Zayed University of Artificial Intelligence, Abu Dhabi, United Arab Emirates
[4] Department of Cardiovascular Sciences, KU Leuven, Belgium

Abstract. Transperineal volumetric ultrasound (US) imaging has become routine practice for diagnosing pelvic floor disease (PFD). Hereto, clinical guidelines stipulate to make measurements in an anatomically defined 2D plane within a 3D volume, the so-called C-plane. This task is currently performed manually in clinical practice, which is labour-intensive and requires expert knowledge of pelvic floor anatomy, as no computer-aided C-plane method exists. To automate this process, we propose a novel, *guideline-driven* approach for automatic detection of the C-plane. The method uses a convolutional neural network (CNN) to identify extreme coordinates of the symphysis pubis and levator ani muscle (which define the C-plane) directly via landmark regression. The C-plane is identified in a postprocessing step. When evaluated on 100 US volumes, our best performing method (multi-task regression with UNet) achieved a mean error of 6.05 mm and 4.81 degrees and took 20 seconds. Two experts blindly evaluated the quality of the automatically detected planes and manually defined the (gold standard) C-plane in terms of their clinical diagnostic quality. We show that the proposed method performs comparably to the manual definition. The automatic method reduces the average time to detect the C-plane by 100 seconds and reduces the need for high-level expertise in PFD US assessment.

1 Introduction

The significance of the levator hiatus Pelvic Floor Disease (PFD) refers to a group of conditions affecting the anatomy and function of pelvic organs. PFD consists of urinary and fecal incontinence, sexual dysfunction and pelvic organ prolapse (POP) [11, 3]. PFD is assessed by observing the levator hiatus (LH) using 3D/4D ultrasound (US), in the plane of minimal hiatal dimensions (defined as the C-plane) as shown in Fig. 1. The C-plane is defined as the angled axial plane of shortest distance between the symphysis pubis (SP) and the pubovisceral muscle at the anorectal angle, referred to as the levator ani muscle (LAM)[7]. Assessments follow International Urogynecological Association (IUGA) and The American Institute of Ultrasound in Medicine (AIUM)

standards [1]. Manual C-plane detection is prone to errors, labour intensive, and requires a high level of expertise. Thus, an automatic detection of the C-plane may reduce measurement variability, improve the clinical workflow, speed-up the detection process and possibly allows low-skilled clinicians to perform the procedure and hence may help a wide range of patients. Unfortunately, to date, no automatic solutions exist.

Automatic plane detection Several studies investigated the extraction of 2D *standard* planes from 3D fetal and cardiac US volumes [10, 4, 14, 9, 15]. Chykeyuk et al. and Yaqub et al. used Regression Random Forests (RRF) to extract planes from US volumes [4, 14]. Yaqub et al. [14] adopted an RRF and feature asymmetry to highlight informative voxels during training. Li et al.[9] proposed an iterative transformation network (ITN) for standard plane detection in 3D volumes. The ITN used a CNN to learn relationships between images of 2D planes and the transformation required to reach the ground-truth. These methods utilise clinical knowledge but not necessarily clinical guidelines. Zhu et al. [15] proposed a guideline-driven learning method to determine 2D planes using anatomical regularities according to guidelines. The use of guidelines improved results and running time on a variety of images. All methods performed well, but without the explicit use of clinical guidelines, the output of the networks will not necessarily produce an output that is clinically interpretable.

Contribution We propose a novel guideline-driven methodology, that utilises CNNs to extract the C-plane automatically from a 3D US volume. This is a novel application of CNN landmark regression for plane detection, and, to the authors' knowledge, the first paper to present a fully automatic C-plane detection pipeline. The CNN identifies two coordinates (directly via landmark regression), that are sufficient to define the C-plane. These coordinates are well-known to clinicians; thus, the output of our proposed method is interpretable, unlike other plane detection methods. We determine the vector of minimal hiatal dimensions within the mid-sagittal (MS) plane, a direction transformation matrix and the mid-point of the plane. Our contributions consist of: a) a new methodological approach to identify the C-plane; b) the novel application of advances in landmark location for plane detection with clinical context; and c) a clinically viable blinded evaluation of C-plane detection.

2 Materials and methods

2.1 Data and manual detection of the C-plane

A dataset containing a training subset of 25 transperineal US volumes from 15 patients, and a validation subset of 100 transperineal US volumes from 43 patients was collected. The small training subset was due to the large clinical overhead required to generate the ground truth data used for training. The dataset was curated at UZ Leuven, Belgium, over the course of 6 months. It includes a

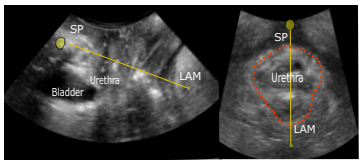

Fig. 1. US at rest, with identification of the C-plane in the MS (left) and oblique axial (right) planes. The yellow line shows the vector of minimal hiatal dimensions and the red contour delineates the hiatal area.

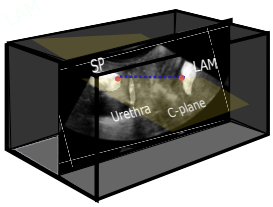

Fig. 2. Extreme coordinates shown as red dots define the C-plane bisecting the MS plane. The vector of minimal hiatal dimensions is shown as a blue line.

high proportion of pathology cases, representing the true *clinical world*, with a variety of pathologies (minor to severe), and a range of image qualities. The volumes were acquired at rest, during Valsalva manoeuvre (i.e. forceful attempted exhalation against a closed airway), and during pelvic floor contraction [1].

To manually determine the C-plane, the acquired volume has been rotated such that the pubic rami are horizontally aligned in the coronal plane, following this stage there are two possible methods: the Multiplanar (MP) and the Omniview (OV) technique (GE Healthcare, Austria). In the MP technique, the MS plane is rotated until the SP and LAM are horizontally aligned. In the OV technique, the clinician draws a line of shortest distance between the SP and LAM to define the C-plane. In our study, volumes were annotated using the MP method, by one expert, resulting in Gold Standard (GS) C-plane orientations used for validation. In practice, manual determination of the C-plane from an US volume can take 2 minutes, which may increase depending on expertise and patient pathology.

2.2 Overall framework

In this work we propose a pipeline to identify the C-plane from an US volume. This is the first time, to the author's knowledge, that localisation of anatomical landmarks has been used directly in the application of plane detection. The process utilises guidelines and automatically detects two extreme coordinates within the US volume; the LAM and the SP which define the C-plane, shown in Fig. 2. Fig. 3 shows the pipeline that detects the extreme coordinates directly via landmark regression. These coordinates are known to clinicians, thus the output can be easily understood and interpreted as it follows the clinical guidelines.

Formation of the direction transformation matrix and mid-point Once SP and LAM extreme points are identified, as described in the following section, the transformation matrix can be determined. We define the vector of shortest Euclidean distance between the SP and the LAM within the MS plane, $\overrightarrow{AB}$. The C-plane is defined in the IUGA guidelines as the plane orthogonal to the

depth direction of the US volume (i.e. $\left|0\ 0\ 1\right|$) and containing the vector $\overrightarrow{AB}$. The second orthogonal vector representing the plane can be defined as the cross product of the normal, and $\overrightarrow{AB}$, thus we define the final orthogonal vector, b_y, as $-AB_y\mathbf{i} + AB_x\mathbf{j} + 0\mathbf{k}$. $\overrightarrow{AB}$ only has magnitude within the x and y components as we restrict the extreme coordinates to lie within the same z slice, this ensures adherence to clinical guidelines. Thus, we have defined the bases of the C-plane and can formulate the following direction matrix

$$\left\|b_x\right\| \left\|b_y\right\| \left\|b_z\right\| = \begin{Vmatrix} AB_x & -AB_y & 0 \\ AB_y & AB_x & 0 \\ 0 & 0 & 1 \end{Vmatrix}. \tag{1}$$

To re-sample the volume, a midpoint, $\vec{O}$, to the C-plane is defined. We identify, $\vec{O}$, as the mid-point between LAM and SP extreme coordinates.

2.3 Regression of extreme coordinates

The method is based on 3D CNN landmark regression, in which we directly regress smooth, distance maps of the SP and LAM extreme coordinates to define the C-plane (see Fig. 3). The rationale for this approach is that regression of one coordinate from an average US volume of 450×500×450 voxels is hard to train. Hence, to make the regression model robust, we propose to regress a distance map of the SP/LAM extreme coordinates, and to use a combined distance map as a weighted sampler. We thus tackle plane detection via regression as a multi-task learning problem and compare two different CNN network architectures. The CNN architecture was adapted to have 2 inputs (i.e. SP and LAM distance maps) and multi-task learning ensuring that features are shared during training.

Determining extreme coordinates The proposed method is based on following the IUGA guidelines. While the regression is performed in 3D, clinicians use 2D MS planes to determine these coordinates. Therefore, unlike other plane detection tasks, a 2D approach is also followed in the determination of the extreme coordinates. As the maximum pixel of the SP and LAM, may be in different MS plane positions within the volume, the combined local sub-pixel maxima of the SP and LAM distance maps are calculated for a small range of 2D MS planes. The range of MS planes remains limited to reduce computational load and is dependent on the overall pixel maxima position of SP and LAM distance maps. Hereto, the following are determined: SP and LAM combined overall maximum, corresponding SP and LAM extreme coordinates, the minimal hiatal vector, $\overrightarrow{AB}$, and mid-point, $\vec{O}$, of the C-plane.

2.4 Evaluation methodology

Intra- and inter-rater variability To evaluate the proposed method, we compare results against inter and intra-rater variability. The expert who determined

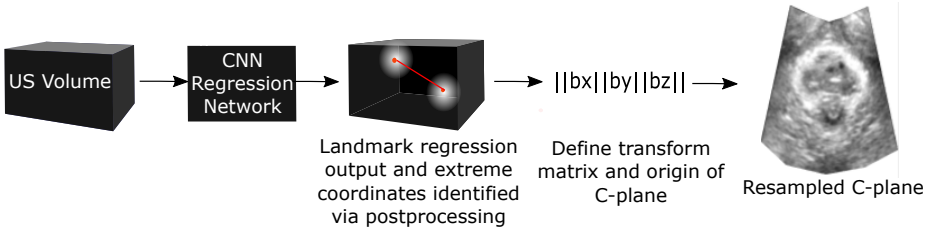

Fig. 3. Pipeline of the landmark regression approach to C-plane extraction.

the GS C-planes (Expert 1), determined the C-plane again a month later, on a subset of 37 volumes. We obtained intra-rater variability using the MP (same method as the GS) and the OV method. Inter-rater variability was obtained by a second expert (Expert 2) using the MP method on a subset of 23 volumes.

Visual Turing test To evaluate clinical relevance, we performed what may be considered a visual *Turing Test* on 10 US volumes. We asked both experts to blindly rate each C-plane (GS and automatically detected), and the corresponding MS plane from 0-5 (5 being excellent, 4, above average, 3, average, 2, below average, 1, poor and 0, no clinical use). Based on the criteria: the minimal hiatal dimension line is placed in the MS plane (Test 1) and the C-plane is of sufficient quality for clinical diagnosis (Test 2). We did a paired Wilcoxon test to compare the proposed method performance against the GS.

Bio-marker analysis The LH identified in the C-plane is a bio-marker used for PFD analysis. We manually segmented the LH in 35 C-planes extracted via the proposed method, 35 GS C-planes and 15 Expert 1 and 15 Expert 2 C-planes. Segmentations were checked and clinically accepted by Expert 1. The Hausdorff Distance (HD) and Robust 95^{th} percentile HD between the segmentation of the GS and predicted C-planes were evaluated to assess the bio-marker quality.

3 Experiments, Results and Discussion

3.1 Implementation details

The network was trained with 25 US volumes, from 12 patients, manually annotated for the SP and LAM positions. For validation, 100 US volumes were used in which C-planes had previously been extracted during clinical routine.

Experiments were implemented using NiftyNet [5] on a desktop with a 24GB NVIDIA Quadro P6000. Two CNN architectures were compared. First an implementation of 3D UNet [16] by Isensee et al. [6], almost identical to the original 3D UNet, but due to high memory consumption of 3D convolutions with large patch sizes, the number of filters was reduced before up-sampling. This network was chosen due to its performance in LH segmentation [2]. The second network

implemented was HighRes3DNet, proposed by Li et al. [8]. It has a large contextual field of view with little effect to parameter cost and was chosen due to its performance in 3D segmentation of the urethra from US volumes [12]. HighRes3DNet, has less trainable parameters than UNet, suggesting a faster inference time, which is beneficial for an *in-clinic* solution. Both multi-task Regression with UNet (mRegU) and multi-task Regression with HighRes3DNet (mRegHR) have a combined L2 loss of the SP and LAM distance maps with initial learning rate of 0.001. A smoother version of the combined distance maps was used for weighted sampling during training. Methods were optimised until network convergence, and data augmentation such as rotation and scaling were used.

3.2 Results and discussion

Fig. 4a compares the Euclidean distance of the mid-point of (manually and automatically) identified C-planes against the GS. From the proposed methods, mRegU showed to have the smallest distance to the GS, highlighting that UNet outperforms other CNN architectures in pelvic floor US landmark localisation tasks. The mean Euclidean distance of mRegU, mRegHR, Expert 1 (OV), Expert 1 (MP) and Expert 2 (MP), were 6.05mm, 7.18mm, 6.95mm, 4.44mm and 5.85mm, respectively. This suggests that mRegU, is within a clinically acceptable error range, as the error is lower than inter-rater variability between experts. The outliers of mRegU and mRegHR were due to the CNN not identifying the true 'extreme coordinate'. In all cases, the network had correctly located a coordinate within the LAM and SP structures. However, it was not the true 'extreme coordinate' position. A contributing factor to the error was the severity of patient pathology. In misclassified cases, patients had one or more conditions such as severe hiatus ballooning, avulsion, bladder and/or vagina prolapse. In addition, poor US acquisition, stool and urine presence, poor acoustic coupling, and patient movement may have contributed to misclassification. In Fig. 4b the angular error is presented, showing that mRegU's interquartile range is smaller than the inter-rater variability. mRegU's mean angular difference was lower than Expert 1's (OV) and Expert 2's (MP), suggesting it is within the range of clinical acceptability. The worst angular mismatch of the detected C-plane was obtained in a patient with severe hiatal ballooning resulting in the LH not being fully within the axial plane due to its size, making it difficult for the network, as the SP was only partially visible in the MS plane. In extreme cases, the SP may not be present at all, clinicians face this problem but use a 4D cine-loop to estimate the SP projection and thereby its location. This is not exploited by the proposed regression network and thus a limitation of the proposed solution.

Fig. 5 shows a visual comparison between the GS and the detected C and MS planes. Qualitatively, there is minimal difference; all show the same clinical diagnosis (symmetric and intact LH); and all structures are visible. The results of the *Turing test*, presented in Table 1, show that mRegHR outperformed all methods including the GS. In one case, the C-plane detected by mRegHR was voted the true minimum hiatal dimension on the MS plane, and the GS

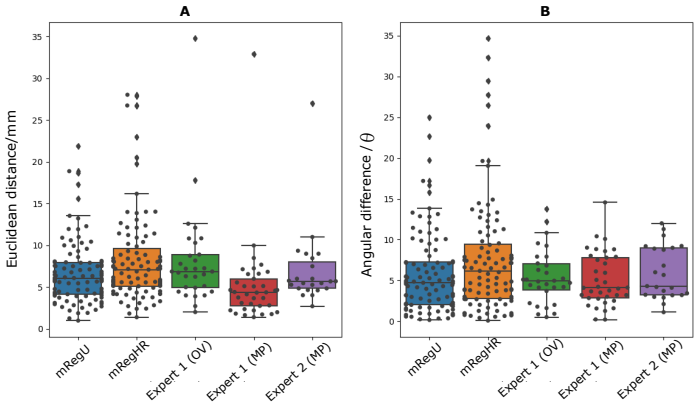

Fig. 4. a) Euclidean distance of the mid-point b) angular difference of the planes determined by the automatic method, Expert 1(MP), Expert1(OV), and Expert 2(MP) against the GS (generated using MP).

C-plane position was rated sub-optimal. This is due to the nature of pelvic floor US where occasionally the C-plane is placed slightly higher/lower than the true position. Although this results in an inaccurate C-plane position, clinicians obtain a better-quality image of the LH by avoiding the hyperechogenicity of the LAM and its associated shadow. This is a limitation to the proposed solution, as although it may determine the true position on the MS plane, the C-plane quality may be sub-optimal. We use a novel application of data and guidelines, to drive a hybrid approach to determine the C-plane, thus we can achieve relatively low errors with a small training dataset, as we utilise the geometry of the patient.

Results of Table 1 show the HD of LH segmentation compared to the GS, mRegU lies within inter-rater variability, thus clinically acceptable bio-markers can be extracted. mRegU and mRegHD perform C-plane detection in 20s and 18s respectively, independent of pathology and image quality. We believe this might reduce with further code optimisation. The time difference, is due to UNet having more parameters than HighRes3DNet. Manually on average, Expert 1 and Expert 2 determine the C-plane in 2 mintues. However, this can vary greatly due to patient's exhibiting PFD, poor acquisition and limited clinical experience. To

Table 1. Hausdorff distance, *Turing test* results and time taken to determine the plane

Plane approach	HD/mm	Robust 95th percentile HD/mm	*Turing Test* 1	*Turing Test* 2	Time/s
mRegU	8.95±4.20	6.13±3.93	3	4	20
mRegHR	9.81±4.31	6.53±4.33	4	4	18
Intra-rater (MP)	8.30±2.27	5.52±2.99	-	-	120
Inter-rater (MP)	8.68 ±3.83	5.99±3.11	-	-	120

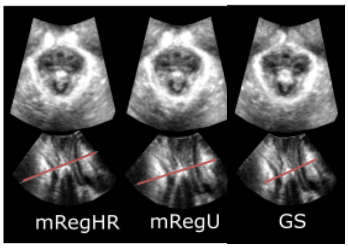

Fig. 5. Visualisation of axial (upper) and MS (lower) plane of minimal hiatal dimensions for automatic methods and GS.

the author's knowledge, landmark regression of only two landmarks has not been used yet to localise a 2D plane from a 3D US volume, even though by practice this is how humans navigate a 3D volume. This method provides interpretability and can be developed to create a fully interactive plane detection application. Other plane detection tasks may be difficult for clinicians to understand why a specific plane was selected if incorrect, and more difficult to integrate into a clinical setting. The results presented in this paper are based on the training dataset of 25 volumes only. It seems plausible that with more training data the results will improve further. In future work, clinicians may be able to edit the C-plane position by re-selecting the extreme coordinate manually. We will expand this pipeline, by including 2D segmentation of the LH, to produce the first, fully automatic PFD bio-marker analysis pipeline.

4 Conclusion

We proposed an automatic plane detection method via CNN landmark regression, that is comparable to experts as demonstrated through validation on 100 US volumes. We automatically detect the plane from an US volume within 18-20 seconds compared to several minutes when performed manually. We discussed limitations to our solution, such as levator ballooning. However, by focusing on anatomy and clinical guidelines, we believe our method follows current clinical workflow, and produces results that are understandable to clinicians. Our hybrid method, which is data and guideline driven, allows our method to learn from a small set of examples, compared to a purely data-driven method. We identified that mRegU performed better than the mRegHR, suggesting the UNet architecture is favourable. In future work, the 2D segmentation of the LH will be included [2] to the pipeline. This will produce the first fully automatic bio-marker analysis pipeline, that can detect the C-plane, segment the LH, and indicate PFD disorders from an US volume [1, 13]. This will reduce the need for high-level clinical expertise and decrease the time needed to perform PFD assessments. Finally, user interaction can be implemented, allowing clinicians to adapt the detected C-plane, by adjusting the SP and LAM coordinates if required.

Acknowledgments We gratefully acknowledge the support of NVIDIA Corporation with the donation of the P6000 GPU and GE Healthcare Women's Health Ultrasound (Zipf, Austria) for their ongoing research and data support.

References

1. 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*, 38(4):851–864, 2019.
2. E. Bonmati, Y. Hu, N. Sindhwani, H. Dietz, J. D'hooge, D. Barratt, J. Deprest, and T. Vercauteren. Automatic segmentation method of pelvic floor levator hiatus in ultrasound using a self-normalising neural network. *Journal of Medical Imaging*, 5, 12 2017.
3. L. Cattani, D. Van Schoubroeck, S. Housmans, G. Callewaert, E. Werbrouck, J. Y. Verbakel, and J. Deprest. Exo-anal imaging of the anal sphincter: a comparison between introital and transperineal image acquisition. *International urogynecology journal*, 31(6):1107—1113, June 2020.
4. K. Chykeyuk, M. Yaqub, and J. Noble. Class-specific regression random forest for accurate extraction of standard planes from 3d echocardiography. volume 8331, pages 53–62, 04 2014.
5. E. Gibson, W. Li, C. H. Sudre, L. Fidon, D. I. Shakir, G. Wang, Z. Eaton-Rosen, R. Gray, T. Doel, Y. Hu, T. Whyntie, P. Nachev, D. C. Barratt, S. Ourselin, M. J. Cardoso, and T. Vercauteren. Niftynet: a deep-learning platform for medical imaging. *CoRR*, abs/1709.03485, 2017.
6. F. Isensee, P. Kickingereder, W. Wick, M. Bendszus, and K. H. Maier-Hein. No new-net. *CoRR*, abs/1809.10483, 2018.
7. K. Lammers, J. J. Fütterer, J. Inthout, M. Prokop, M. E. Vierhout, and K. B. Kluivers. Correlating signs and symptoms with pubovisceral muscle avulsions on magnetic resonance imaging. *American Journal of Obstetrics and Gynecology*, 208(2):148.e1 – 148.e7, 2013.
8. W. Li, G. Wang, L. Fidon, S. Ourselin, M. J. Cardoso, and T. Vercauteren. On the compactness, efficiency, and representation of 3d convolutional networks: Brain parcellation as a pretext task. *CoRR*, abs/1707.01992, 2017.
9. Y. Li, B. Khanal, B. Hou, A. Alansary, J. J. Cerrolaza, M. Sinclair, J. Matthew, C. Gupta, C. Knight, B. Kainz, and et al. Standard plane detection in 3d fetal ultrasound using an iterative transformation network. *Lecture Notes in Computer Science*, page 392–400, 2018.
10. X. Lu, B. Georgescu, Y. Zheng, J. Otsuki, and D. Comaniciu. Autompr: Automatic detection of standard planes in 3d echocardiography. pages 1279 – 1282, 06 2008.
11. J. L. Whiteside and T. Muffly. Chapter 26 overview of pelvic floor disorders: Epidemiology, diagnosis, and treatment. In M. B. Goldman, R. Troisi, and K. M. Rexrode, editors, *Women and Health (Second Edition)*, pages 389 – 403. Academic Press, second edition edition, 2013.
12. H. Williams, L. Cattani, W. Li, M. Tabassian, T. Vercauteren, J. Deprest, and J. D'hooge. 3d convolutional neural network for segmentation of the urethra in volumetric ultrasound of the pelvic floor. In *2019 IEEE International Ultrasonics Symposium (IUS)*, pages 1473–1476, 2019.
13. Y. Xuan, T. Friedman, and H. Dietz. Does levator ani hiatal area configuration affect pelvic organ prolapse? *Ultrasound in obstetrics and gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology*, 54(1):124—127, July 2019.
14. M. Yaqub, A. Kopuri, S. Rueda, P. Sullivan, K. McCormick, and J. Noble. A constrained regression forests solution to 3d fetal ultrasound plane localization for longitudinal analysis of brain growth and maturation. volume LNCS 8679, pages 109–116, 09 2014.

15. P. Zhu and Z. Li. Guideline-based machine learning for standard plane extraction in 3d cardiac ultrasound. In *MCV/BAMBI@MICCAI*, 2016.
16. Özgün Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger. 3d u-net: Learning dense volumetric segmentation from sparse annotation, 2016.