*Ultrasound Obstet Gynecol* 2019; **54**: 270–275
Published online 26 June 2019 in Wiley Online Library (wileyonlinelibrary.com). **DOI:** 10.1002/uog.20181.
This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

# Deep learning enables automatic quantitative assessment of puborectalis muscle and urogenital hiatus in plane of minimal hiatal dimensions

F. VAN DEN NOORT[1,2], C. H. VAN DER VAART[2], A. T. M. GROB[3], M. K. VAN DE WAARSENBURG[2], C. H. SLUMP[1] and M. VAN STRALEN[4]

[1] *Robotics and Mechatronics, Faculty of Electrical Engineering, Mathematics and Computer Science, Technical Medical Center, University of Twente, Enschede, The Netherlands;* [2] *Department of Reproductive Medicine and Gynecology, University Medical Center, Utrecht, The Netherlands;* [3] *Multi-modality Medical Imaging, Faculty of Science and Technology, Technical Medical Center, University of Twente, Enschede, The Netherlands;* [4] *Imaging Division, University Medical Center Utrecht, Utrecht, The Netherlands*

**KEYWORDS:** convolutional neural network; deep learning; puborectalis muscle; segmentation; transperineal ultrasound; urogenital hiatus

## ABSTRACT

***Objectives*** *To measure the length, width and area of the urogenital hiatus (UH), and the length and mean echogenicity (MEP) of the puborectalis muscle (PRM), automatically and observer-independently, in the plane of minimal hiatal dimensions on transperineal ultrasound (TPUS) images, by automatic segmentation of the UH and the PRM using deep learning.*

***Methods*** *In 1318 three- and four-dimensional (3D/4D) TPUS volume datasets from 253 nulliparae at 12 and 36 weeks' gestation, two-dimensional (2D) images in the plane of minimal hiatal dimensions with the PRM at rest, on maximum contraction and on maximum Valsalva maneuver, were obtained manually and the UH and PRM were segmented manually. In total, 713 of the images were used to train a convolutional neural network (CNN) to segment automatically the UH and PRM in the plane of minimal hiatal dimensions. In the remainder of the dataset (test set 1 (TS1); 601 images, four having been excluded), the performance of the CNN was evaluated by comparing automatic and manual segmentations. The performance of the CNN was also tested on 117 images from an independent dataset (test set 2 (TS2); two images having been excluded) from 40 nulliparae at 12 weeks' gestation, which were acquired and segmented manually by a different observer. The success of automatic segmentation was assessed visually. Based on the CNN segmentations, the following clinically relevant parameters were measured: the length, width and area of the UH, the length of the PRM and MEP. The overlap (Dice similarity index (DSI)) and surface distance (mean absolute distance (MAD) and Hausdorff distance (HDD)) between manual and CNN segmentations were measured to investigate their similarity. For the measured clinically relevant parameters, the intraclass correlation coefficients (ICCs) between manual and CNN results were determined.*

***Results*** *Fully automatic CNN segmentation was successful in 99.0% and 93.2% of images in TS1 and TS2, respectively. DSI, MAD and HDD showed good overlap and distance between manual and CNN segmentations in both test sets. This was reflected in the respective ICC values in TS1 and TS2 for the length (0.96 and 0.95), width (0.77 and 0.87) and area (0.96 and 0.91) of the UH, the length of the PRM (0.87 and 0.73) and MEP (0.95 and 0.97), which showed good to very good agreement.*

***Conclusion*** *Deep learning can be used to segment automatically and reliably the PRM and UH on 2D ultrasound images of the nulliparous pelvic floor in the plane of minimal hiatal dimensions. These segmentations can be used to measure reliably UH dimensions as well as PRM length and MEP. © 2018 The Authors. Ultrasound in Obstetrics & Gynecology published by John Wiley & Sons Ltd on behalf of the International Society of Ultrasound in Obstetrics and Gynecology.*

*Correspondence to:* Ms F. van Limbeek-van den Noort, University of Twente, Carre 3.526, Drienerlolaan 5, 7522NB, Enschede, The Netherlands (e-mail: f.vandennoort@utwente.nl)

*Accepted: 15 November 2018*

© 2018 The Authors. *Ultrasound in Obstetrics & Gynecology* published by John Wiley & Sons Ltd
on behalf of the International Society of Ultrasound in Obstetrics and Gynecology.

## INTRODUCTION

In the last 5 years, deep learning has had an increasing impact on automating the analysis of medical imaging data[1]. Deep learning is a class of computer algorithms that mimics the learning of the human brain and performs well on tasks formerly thought of as primarily human. For example, in image analysis, convolutional neural networks (CNNs), which are a type of deep learning, outperform state-of-the-art algorithms[2–4]. In medicine, CNNs are used for diagnosis, for example in discrimination between images of benign and malignant skin lesions[5] and the detection of Alzheimer's disease on magnetic resonance imaging (MRI)[6]. Segmentation can also be learned by CNN and was found to be successful in, for example, segmenting brain structures on MRI and ultrasound[7] and segmentation of the urogenital hiatus (UH)[8].

Transperineal ultrasound (TPUS) is used in the field of urogynecology to diagnose and understand pelvic floor biometrics and problems. UH dimensions[9,10] and mean echogenicity (MEP)[11–13] and global strain of the puborectalis muscle (PRM)[14] provide information about PRM composition and function, and are therefore parameters of potential clinical relevance[15,16]. These parameters are measured manually, making them observer dependent and time consuming, limiting their introduction into general practice. To automate these measurements, Bonmati *et al.*[8] used a CNN for UH segmentation. We showed previously automatic segmentation of the PRM[17] using active appearance models. However, this was three-dimensional (3D) and the clinical relevance still needs to be investigated.

In this study, we aimed to segment automatically the PRM and UH in the plane of minimal hiatal dimensions, using a CNN. In addition, we aimed to use these segmentations to measure the corresponding relevant clinical parameters (width, length and area of the UH, the length of the PRM and MEP) automatically and to test the reliability.

## METHODS

### Data

In this study, the dataset acquired by van Veelen *et al.*[18] was used. This dataset contains three- and four-dimensional (3D/4D) TPUS data of 280 nulliparae at 12 and 36 weeks of gestation and 6 months postpartum. The data were acquired using the GE Voluson 730 Expert ultrasound system (GE Medical Systems, Zipf, Austria), using the RAB4-8L probe. During the examinations, the women were asked to relax and contract their pelvic floor muscles and to perform a Valsalva maneuver. Only data on the intact pelvic floor, acquired at 12 and 36 weeks of gestation, were included in this study. In a previous study[12], ultrasound images from the 253 women had been segmented manually in the plane of minimal hiatal dimensions. A total of 1318 two-dimensional (2D) TPUS images in the plane of minimal hiatal dimensions from these women were used in this study.

The ultrasound data were loaded into 4D View 7.0 (GE Medical Systems) and rotated manually to obtain the plane of minimal hiatal dimensions which was then exported as a 2D bitmap image. This was done for images obtained at rest, on maximum pelvic floor muscle contraction and on maximum Valsalva. Afterwards, the PRM and UH were segmented manually for further study[11,12] and inter- and intraobserver agreement were investigated[19]. These segmentations were used in this study to train and test the CNN as follows: the dataset was split randomly into a training set (122 women, 649 images), a validation set (12 women, 64 images) and a test set (test set 1, TS1). Four images obtained on Valsalva maneuver in the test set did not capture fully the PRM and were excluded, giving a total of 601 images obtained in 119 women (Figure 1). The training set was used to train different CNNs. In order to decide which network performed best on these data, the validation set was used. The best performing CNN was then applied to the test set in order to analyze its performance independently.

The developed CNN was additionally tested on an independently acquired dataset that contained TPUS images obtained at 12 weeks' gestation in pregnant nulliparae. This dataset was acquired using the same ultrasound system and settings as those used to acquire the previous dataset[18], but was obtained and processed by a different independent sonographer. The TPUS images of 40 subjects were selected randomly (test set 2, TS2) and the PRM and UH were segmented manually at rest, on contraction and on Valsalva, following the same procedure as that used for the first data set. Two images obtained on Valsalva maneuver did not capture fully the PRM and were therefore excluded, giving a total of 117 images obtained in 40 women (Figure 1). The remaining manual segmentations were used as the gold standard for evaluation of the performance of the CNN.

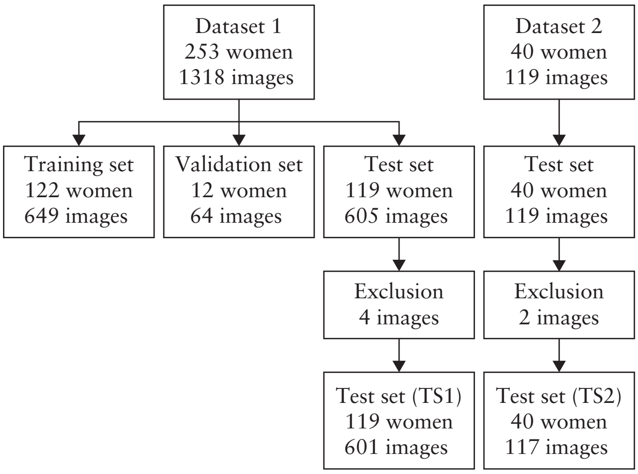

**Figure 1** Flowchart summarizing division of two datasets of transperineal ultrasound images used in training, validation and testing of convolutional neural network for automatic segmentation of urogenital hiatus and puborectalis muscle.

© 2018 The Authors. *Ultrasound in Obstetrics & Gynecology* published by John Wiley & Sons Ltd on behalf of the International Society of Ultrasound in Obstetrics and Gynecology.

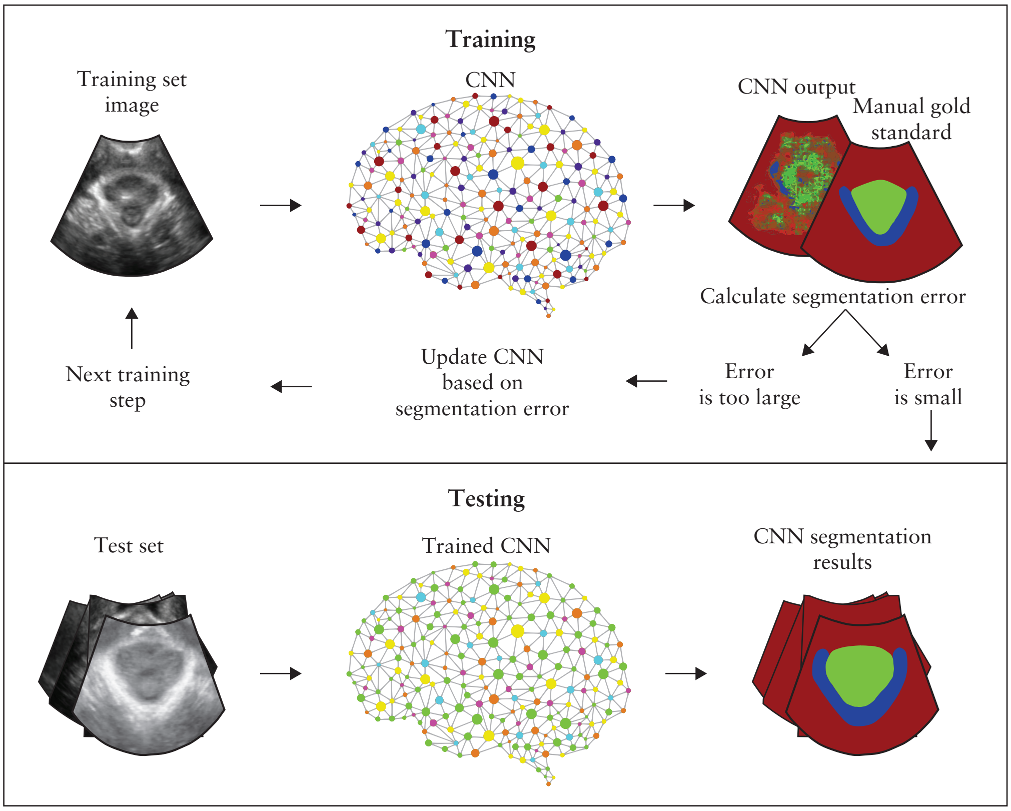

**Figure 2** During training of convolutional neural network (CNN) for automatic segmentation, image from training set is fed into CNN and output of CNN is compared to manual segmentation of this image. If error is too large, CNN is updated and process starts again. If error is sufficiently small, CNN can be tested on test set of data that it has not seen during training. Quality of CNN segmentations can thus be evaluated independently.

## Convolutional neural networks

CNNs can learn to segment the PRM on TPUS images from example segmented data[1]. They consist of a number of layers (which can be adjusted, creating different CNN architectures), each filtering the previous layer and subsequently creating a higher level of abstraction of the input image in the final layer, creating a pixel-wise segmentation of the PRM. Within the layers, feature detectors are learned (the kernels), based on the example training data. Sufficient and representative data need to be provided during the learning phase to be able to generalize over the input examples. During training, the output of the CNN is compared to the desired output (the manual segmented gold standard) and the kernels are optimized mathematically to minimize the error between the CNN output and the ground truth segmentations for the training set only. After optimization, the CNN is tested on an independent dataset that it has not seen before to evaluate its performance. Figure 2 illustrates the training and testing phases. The specifics of the implementation of our CNN are available in Appendix S1 and Figure S1.

## Validation

To validate the performance of our trained CNN, it was applied to TS1 and TS2. We first determined CNN segmentation success by assessing the completeness of the segmentations, rejecting obvious errors when the CNN segmented only half or less of the PRM. From the successful CNN segmentations, the UH height, width and area, PRM length and MEP were calculated automatically. Figure 3 shows how these parameters are defined in the plane of minimal hiatal dimensions. For these measurements, mean ± SD and intraclass correlation coefficients (ICCs) with 95% CI were calculated using Python and the ICCs were evaluated according to the subgroup definitions of Landis and Koch[20]. The Bland–Altman method[21] was used to investigate the mean difference and limits of agreement (LOA) between CNN and manual segmentations.

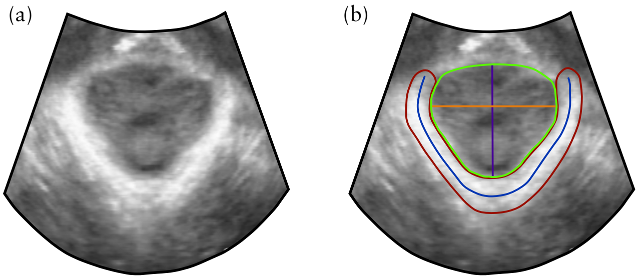

**Figure 3** (a) Transperineal ultrasound image of pelvic floor in plane of minimal hiatal dimensions, in pregnant nullipara. (b) Same image, showing parameters measured in this plane: width (orange), length (purple) and area (green) of urogenital hiatus and length (blue) and mean echogenicity of puborectalis muscle (i.e. of red area).

© 2018 The Authors. *Ultrasound in Obstetrics & Gynecology* published by John Wiley & Sons Ltd on behalf of the International Society of Ultrasound in Obstetrics and Gynecology.

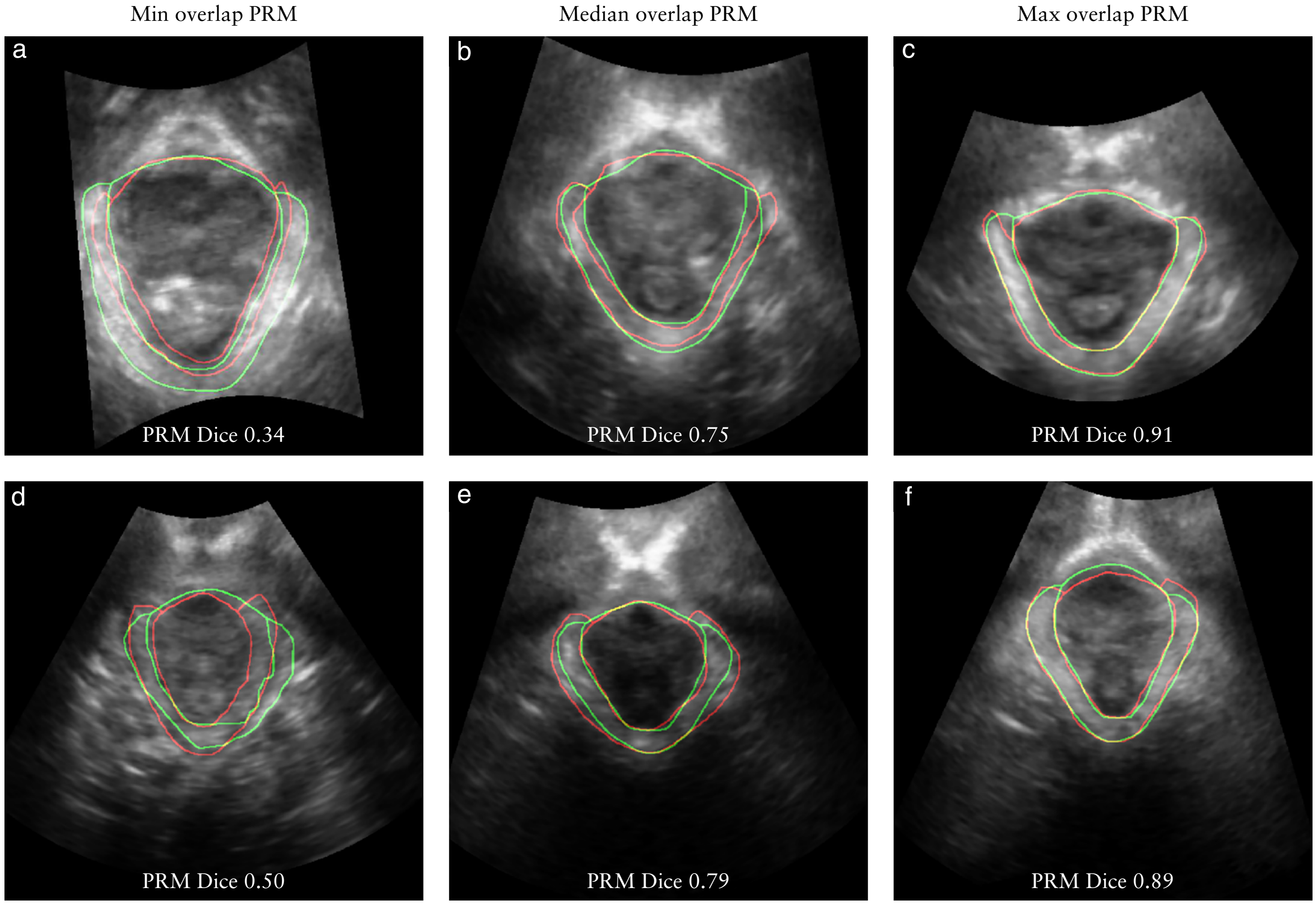

**Figure 4** Convolutional neural network (CNN) automatic segmentation results of urogenital hiatus and puborectalis muscle (PRM) on transperineal ultrasound in plane of minimal hiatal dimensions in pregnant nulliparae for images in test sets 1 (a–c) and 2 (d–f), with minimum (min; a,d), median (b,e) and maximum (max; c,f) overlap between manual (red) and CNN (green) PRM segmentations. Yellow denotes overlap between green and red lines.

In order to analyze the similarity of the manual and CNN segmentations, the Dice similarity index (DSI), mean absolute distance (MAD) and Hausdorff distance (HDD) were calculated. DSI is a measure of overlap between two segmentations, X and Y, calculated as follows: $\mathrm{DSI} = (2\,|X \cap Y|)/(|X| + |Y|)$. Here, $\mathrm{DSI} = 0$ indicates no overlap and $\mathrm{DSI} = 1$ indicates maximum overlap between the segmentations. MAD is the average distance between the borders of the two segmentations and HDD is the maximum distance between the two segmentations.

## RESULTS

After applying the trained CNN to both test sets (TS1 and TS2), the results were checked visually. There were six (1.0%; three at rest, three on Valsalva) unsuccessful segmentations of images in TS1 and eight (6.8%; two at rest, three on contraction, three on Valsalva) in TS2.

In Figure 4, the automatic segmentations with the worst, median and best results for PRM overlap are shown for each test set. Table 1 displays the results for DSI, MAD and HDD between CNN and manual segmentations, providing an indication of how similar the segmentations are. In Table 2, the performance of the CNN in measuring the corresponding clinically relevant parameters is shown in comparison to the manual measurements, showing ICCs, means, SD, mean differences and LOA. The ICCs indicate good to very good agreement between manual and automatic measurements.

## DISCUSSION

In this study, we developed a CNN for segmenting the UH and PRM in the plane of minimal hiatal dimensions on TPUS images in pregnant nulliparae. From these segmentations, we were able to measure UH and PRM dimensions. We showed in both test sets that these automatic measurements have good agreement with manual measurements.

Our results can be compared to the (semi) automatic segmentation results obtained for the UH, reported previously by Sindhwani *et al.*[22] and Bonmati *et al.*[8]. With respect to overlap and distance measures, the automatic segmentations in the current study performed similarly or slightly better. However, it should be noted that Sindhwani *et al.*[22] explicitly also included avulsion patients, while we included nulliparae with no known

© 2018 The Authors. *Ultrasound in Obstetrics & Gynecology* published by John Wiley & Sons Ltd on behalf of the International Society of Ultrasound in Obstetrics and Gynecology.

**Table 1** Dice similarity index (DSI), mean absolute distance (MAD) and Hausdorff distance (HDD) values between manual and convolutional neural network segmentations of puborectalis muscle (PRM) and urogenital hiatus (UH) on transperineal ultrasound in plane of minimal hiatal dimensions in pregnant nulliparae, for test sets 1 (TS1) and 2 (TS2)

| | *DSI* | | *MAD (mm)* | | *HDD (mm)* | |
|---|---|---|---|---|---|---|
| *Test set* | *UH* | *PRM* | *UH* | *PRM* | *UH* | *PRM* |
| TS1 | 0.94 ± 0.02 | 0.73 ± 0.09 | 1.19 ± 0.43 | 1.23 ± 0.33 | 3.68 ± 1.52 | 4.85 ± 1.70 |
| TS2 | 0.93 ± 0.03 | 0.77 ± 0.06 | 1.22 ± 0.43 | 1.35 ± 0.36 | 3.52 ± 1.18 | 5.85 ± 1.69 |

Data are given as mean ± SD.

**Table 2** Comparison of manual and convolutional neural network (CNN) measurements of length, width and area of urogenital hiatus and length and mean echogenicity (MEP) of puborectalis muscle on transperineal ultrasound in plane of minimal hiatal dimensions in pregnant nulliparae, for test sets 1 (TS1) and 2 (TS2)

| | *Urogenital hiatus* | | | *Puborectalis muscle* | |
|---|---|---|---|---|---|
| | *Length* | *Width* | *Area* | *Length* | *MEP* |
| TS1 | | | | | |
| ICC (95% CI) | 0.96 (0.94–0.97) | 0.77 (0.65–0.84) | 0.96 (0.95–0.96) | 0.87 (0.64–0.93) | 0.95 (0.90–0.97) |
| CNN | 4.86 ± 0.85 | 4.03 ± 0.45 | 14.47 ± 3.68 | 11.54 ± 1.70 | 141 ± 21 |
| Manual | 4.93 ± 0.85 | 4.17 ± 0.55 | 14.63 ± 3.83 | 12.08 ± 1.76 | 138 ± 21 |
| Mean difference* | −0.07 ± 0.23 | −0.14 ± 0.31 | −0.16 ± 1.11 | −0.54 ± 0.74 | 3.1 ± 5.8 |
| LOA | −0.53 to 0.38 | −0.75 to 0.47 | −2.35 to 2.02 | −2.00 to 0.92 | −8.2 to 14.5 |
| TS2 | | | | | |
| ICC (95% CI) | 0.95 (0.79–0.97) | 0.87 (0.71–0.93) | 0.91 (0.21–0.97) | 0.73 (0.00–0.92) | 0.97 (0.70–0.99) |
| CNN | 4.46 ± 0.68 | 4.00 ± 0.42 | 12.97 ± 3.01 | 10.64 ± 1.42 | 119 ± 25 |
| Manual | 4.31 ± 0.70 | 3.88 ± 0.47 | 11.90 ± 3.06 | 11.68 ± 1.42 | 114 ± 23 |
| Mean difference* | 0.15 ± 0.18 | 0.12 ± 0.19 | 1.07 ± 0.77 | −1.03 ± 0.57 | 4.2 ± 3.5 |
| LOA | −0.21 to 0.50 | −0.26 to 0.51 | −0.44 to 2.58 | −2.16 to 0.09 | −2.6 to 11.2 |

Data are given as mean ± SD in cm (length, width), cm$^2$ (area) or arbitrary units (MEP), unless stated otherwise. *Mean difference = (CNN − Manual). ICC, intraclass correlation coefficient; LOA, limits of agreement.

pelvic floor problems, which may positively influence our results.

Automatic segmentation of the PRM in the plane of minimal hiatal dimensions has, to the best of our knowledge, not been presented before. The MAD of the UH and PRM segmentations are comparable and indicate only a few pixels' mismatch on average between manual and automatic segmentations. The maximum error, the HDD, was about 1–2 mm higher for the PRM than for the UH. This maximum mismatch is typically found at the site of attachment to the pubic symphysis, as this is the most difficult part in PRM segmentation, which is similarly the case for volume segmentations on 3D ultrasound[17]. Therefore, the maximum segmentation mismatch is expected to be higher for the PRM. The DSI values were lower for the PRM than for the UH because DSI is sensitive to the shape of segmentations; round segmentations (such as the UH) are more likely to have high DSI values as compared with thin segmentations (such as the PRM)[17].

The reliability of the measured clinical parameters with respect to the UH and PRM can be compared to the results obtained manually by Weinstein *et al.*[10], van Veelen *et al.*[18] and Grob *et al.*[19]. The ICC and LOA values for the length, width and area of the UH, when comparing computer and observer measurements, are comparable to or better than those between manual observers[10,18]. The ICC and LOA values for MEP between automatic and manual measurements are comparable to those between manual measurements[19]. Reliability of manual measurements of length of the PRM has, to the best of our knowledge, never been analyzed but, comparing the ICC values in the current study to those for manual measurement of the PRM inner perimeter, which is comparable to PRM length[10], suggests that the automatic measurements are slightly less reliable.

Of the 14 images that were excluded after the CNN segmented only half of the PRM, the majority had poor image contrast. These mismatches can potentially be filtered out automatically as they have a short PRM length. In principle, a PRM length threshold can be defined and used by a human observer to check the segmentation results. Alternatively, in future work, the model could be extended with a discriminator network to check realistic PRM segmentations in all cases[23].

The large amount of data available in this study provides strong evidence for the reliability of the presented method. The confirmation of these results in an independent data set (TS2) demonstrates its potentially wide applicability as each dataset was acquired and segmented manually by different observers. This also confirms that the performance is on a par with that of human observers.

This study has some limitations. The algorithm was trained using only data on nulliparous women. As the complete dataset (including some postpartum patients from the dataset of van Veelen *et al.*[18]) consists of mostly nulliparous data, the number of TPUS images capturing an avulsion is relatively low. This selection could introduce bias into the CNN, challenging the development of such

© 2018 The Authors. *Ultrasound in Obstetrics & Gynecology* published by John Wiley & Sons Ltd on behalf of the International Society of Ultrasound in Obstetrics and Gynecology.

a method capable of dealing with avulsions, based on these limited data. Avulsion of the PRM influences its appearance on TPUS and there is no complete agreement between experts on the appearance of avulsions[24–27]. Therefore, we decided to focus on the intact pelvic floor by including only the data of nulliparae in our training set. Now that we have shown that a CNN can detect reliably an intact pelvic floor, the next step is to analyze women with PRM avulsion and retrain the CNN.

Another limitation of this study is that the CNN was applied to the plane of minimal hiatal dimensions, the selection of which is still a manual process. Therefore, before our CNN method can be used in clinical practice, there are a few steps that need to be taken. These include training the computer on selecting automatically the relevant frame in the movie (rest, contraction or Valsalva) and finding the plane of minimal hiatal dimensions.

Furthermore, detection of a PRM avulsion or a wide levator–urethra gap as a sign of an avulsion, as proposed by Dietz *et al.*[28], may make our results valuable in clinical practice. Nevertheless, there is value in automatically filtering out images with normal PRM and hiatal dimensions to limit the number of images that need to be examined visually.

In conclusion, we have shown that, using deep learning, it is possible to segment automatically and reliably the PRM and UH on 2D TPUS images of the nulliparous pelvic floor in the plane of minimal hiatal dimensions. From these segmentations, we could measure reliably the width, length and area of the UH, MEP and length of the PRM. These results were confirmed in an independent dataset.

## ACKNOWLEDGMENT

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research. This work is part of the research program Gynius with project number 15301, which is financed by The Netherlands Organisation for Scientific Research (NWO).

## REFERENCES

1. Litjens G, Kooi T, Bejnordi BE, Arindra A, Setio A, Ciompi F, Ghafoorian M, van der Laak JAWM, van Ginneken B, Sánchez CI. A survey on deep learning in medical image analysis. *Med Image Anal* 2017; **42**: 60–88.
2. Fukushima K. Biological cybernetics neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol Cybern* 1980; **36**: 193–202.
3. LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proc IEEE* 1998; **86**: 2278–2324.
4. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems*, vol 1, Pereira F, Burges CJC, Bottou L, Weinberger KQ (eds). MIT Press, Cambridge, MA, 2012; 1097–1105.
5. Kawahara J, Hamarneh G. Multi-resolution-tract CNN with hybrid pretrained and skin-lesion trained layers. In *Machine Learning in Medical Imaging*, Wang L, Adeli E, Wang Q, Shi Y, Suk H-I (eds). Springer International Publishing, New York, NY, 2016; 164–171.
6. Hosseini-Asl E, Gimel 'farb G, El-Baz A. Alzheimer's disease diagnostics by a deeply supervised adaptable 3D convolutional neural network. *Front Biosci* 2018; **23**: 584–596.
7. Milletari F, Ahmadi S-A, Kroll C, Plate A, Rozanski V, Maiostre J, Levin J, Dietrich O, Ertl-Wagner B, Bötzel K, Navab N. Hough-CNN: Deep learning for segmentation of deep brain regions in MRI and ultrasound. *Comput Vis Image Underst* 2017; **164**: 92–102.
8. Bonmati E, Hu Y, Sindhwani N, Dietz HP, D'hooge J, Barratt D, Deprest J, Vercauteren T. Automatic segmentation method of pelvic floor levator hiatus in ultrasound using a self-normalizing neural network. *J Med Imaging (Bellingham)* 2018; **5**: 021206.
9. Dietz HP, Shek C, Clarke B. Biometry of the pubovisceral muscle and levator hiatus by three-dimensional pelvic floor ultrasound. *Ultrasound Obstet Gynecol* 2005; **25**: 580–585.
10. Weinstein MM, Jung SA, Pretorius DH, Nager CW, den Boer DJ, Mittal RK. The reliability of puborectalis muscle measurements with 3-dimensional ultrasound imaging. *Am J Obstet Gynecol* 2007; **197**: 68.e1–6.
11. Grob AT, Withagen MI, van de Waarsenburg MK, Schweitzer KJ, van der Vaart CH. Changes in the mean echogenicity and area of the puborectalis muscle during pregnancy and postpartum. *Int Urogynecol J* 2016; **27**: 895–901.
12. Grob AT, Withagen MI, van de Waarsenburg MK, Schweitzer KJ, van der Vaart CH. Association of first-trimester echogenicity of the puborectalis muscle with mode of delivery. *Obstet Gynecol* 2016; **127**: 1021–1026.
13. Van de Waarsenburg MK, van der Vaart CH, Withagen MIJ. Structural changes in the puborectalis muscle after vaginal delivery. *Ultrasound Obstet Gynecol* 2019; **53**: 256–261.
14. Grob ATM, Hitschrich N, Withagen MIJ, van de Waarsenburg MK, Schweitzer KJ, van der Vaart CH. Changes in the global strain of the puborectalis muscle during pregnancy and postpartum. *Ultrasound Obstet Gynecol* 2018; **51**: 537–542.
15. Liu Y-W, Su C-T, Huang Y-Y, Yang C-S, Huang J-W, Yang M-T, Chen JH, Tsai WC. Left ventricular systolic strain in chronic kidney disease and hemodialysis patients. *Am J Nephrol* 2011; **33**: 84–90.
16. Pillen S, Arts IMP, Zwarts MJ. Muscle ultrasound in neuromuscular disorders. *Muscle Nerve* 2008; **37**: 679–693.
17. van den Noort F, Grob ATM, Slump CH, van der Vaart CH, van Stralen M. Automatic segmentation of the puborectalis muscle on three-dimensional transperineal ultrasound. *Ultrasound Obstet Gynecol* 2018; **52**: 97–102.
18. van Veelen GA, Schweitzer KJ, van der Vaart CH. Reliability of pelvic floor measurements on three- and four-dimensional ultrasound during and after first pregnancy: implications for training. *Ultrasound Obstet Gynecol* 2013; **42**: 590–595.
19. Grob ATM, Veen AAC, Schweitzer KJ, Withagen MIJ, van Veelen GA, van der Vaart CH. Measuring echogenicity and area of the puborectalis muscle: method and reliability. *Ultrasound Obstet Gynecol* 2014; **44**: 481–485.
20. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33**: 159–174.
21. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; **327**: 307–310.
22. Sindhwani N, Barbosa D, Alessandrini M, Heyde B, Dietz HP, D'Hooge J, Deprest J. Semi-automatic outlining of levator hiatus. *Ultrasound Obstet Gynecol* 2016; **48**: 98–105.
23. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative adversarial nets. In *Proceedings of the 27th International Conference on Neural Information Processing Systems*, vol 2, Ghahramani Z, Welling M, Cortes C, Lawrence ND, Weinberger KQ (eds). MIT Press, Cambridge, MA, 2014; 2672–2680.
24. Dietz HP, Bernardo MJ, Kirby A, Shek KL. Minimal criteria for the diagnosis of avulsion of the puborectalis muscle by tomographic ultrasound. *Int Urogynecol J* 2011; **22**: 699–704.
25. van Delft KWM, Sultan AH, Thakar R, Shobeiri SA, Kluivers KB. Agreement between palpation and transperineal and endovaginal ultrasound in the diagnosis of levator ani avulsion. *Int Urogynecol J* 2015; **26**: 33–39.
26. van Veelen GA, Schweitzer KJ, van Delft K, Kluivers KB, Weemhoff M, van der Vaart CH. Diagnosing levator avulsions after first delivery by tomographic ultrasound: reliability between observers from different centers. *Int Urogynecol J* 2014; **25**: 1501–1506.
27. Da Silva AS, Digesu GA, Dell'Utri C, Fritsch H, Piffarotti P, Khullar V. Do ultrasound findings of levator ani "avulsion" correlate with anatomical findings: a multicenter cadaveric study. *Neurourol Urodyn* 2016; **35**: 683–688.
28. Dietz HP, Abbu A, Shek KL. The levator–urethra gap measurement: a more objective means of determining levator avulsion? *Ultrasound Obstet Gynecol* 2008; **32**: 941–945.

## SUPPORTING INFORMATION ON THE INTERNET

The following supporting information may be found in the online version of this article:

**Appendix S1** and **Figure S1** may be found in the online version of this article.

© 2018 The Authors. *Ultrasound in Obstetrics & Gynecology* published by John Wiley & Sons Ltd on behalf of the International Society of Ultrasound in Obstetrics and Gynecology.