

Automated Analysis of Ultrasound Images for the Assessment of Pelvic Floor Dysfunction



Helena Cynthia Williams

KU Leuven
Biomedical Sciences Group
Faculty of Medicine
Department of Development and Regeneration
Herestraat 49, B-3000 Leuven (Belgium)



DOCTORAL SCHOOL
BIOMEDICAL SCIENCES

AUTOMATED ANALYSIS OF ULTRASOUND IMAGES FOR THE ASSESSMENT OF PELVIC FLOOR DYSFUNCTION

Helena WILLIAMS

Promoters:

Prof. Dr. Jan D'hooge
Prof. Dr. Jan Deprest
Prof. Dr. Tom Vercauteren

Jury:

Prof. Dr. Kathleen Freson (Chair)
Prof. Dr. Aly Youssef
Prof. Dr. Chris de Korte
Prof. Dr. Frederik Maes
Dr. Dominique Van Schoubroeck

Dissertation presented in
partial fulfilment of the
requirements for the
degree of Doctor in
Biomedical Sciences

October 2022

KU Leuven
Groep Biomedische Wetenschappen
Faculteit Geneeskunde
Departement Ontwikkeling en Regeneratie
Herestraat 49, B-3000 Leuven (Belgium)



GEAUTOMATISEERDE ANALYSE VAN ECHOGRAFISCHE BEELDEN VOOR DE BEOORDELING VAN BEKKENBODEMDISFUNCTIE

Helena WILLIAMS

Promoters:

Prof. Dr. Jan D'hooge
Prof. Dr. Jan Deprest
Prof. Dr. Tom Vercauteren

Jury:

Prof. Dr. Kathleen Freson (Chair)
Prof. Dr. Aly Youssef
Prof. Dr. Chris de Korte
Prof. Dr. Frederik Maes
Dr. Dominique Van Schoubroeck

Proefschrift voorgedragen
tot het behalen van de
graad van Doctor in de
Biomedische
Wetenschappen

©Katholieke Universiteit Leuven - Groep Biomedische Wetenschappen Herestraat 49, 3000 Leuven (België)

Alle rechten voorbehouden. Niets uit deze uitgave mag worden vermenigvuldigd en/of openbaar gemaakt worden door middel van druk, fotokopie, microfilm, elektronisch of op welke andere wijze ook, zonder voorafgaande schriftelijke toestemming van de uitgever.

All rights reserved. No part of this publication may be reproduced in any form by print, photoprint, microfilm or any other means without written permission from the publisher.

ISBN 9789080569843

*To my parents,
and in loving memory of
Nanny, Opa,
Auntie Faye and Oom Aad.*

Data is useful. High-quality, well-understood, real-world data is priceless, but biased data can be dangerous.

Declaration

I, Helena Williams, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Helena Williams
November 2022

Acknowledgements

From conversations with supervisors, mentors, reviewers, colleagues, friends, family and even strangers, inspiration, motivation, and support can manifest in equal measures in all situations. This can be so impactful that the person ends up completing a body of work that they may have never dreamed of accomplishing before. I am honoured and humbled to have been a part of many great conversations during my last four years, which have led me to move onto this next stage of my career and the successful completion of this thesis.

First, I must thank **my supervisors**, Jan D'hooge, Jan Deprest and Tom Vercauteren. **Jan D'hooge**, your guidance, kindness, discussions and teachings have been extremely impactful over the last four years, and they have transformed me into a better scientist. You had emphasised to me that an important factor in achieving happiness is freedom and that a key advantage of pursuing a career in academia, is one's freedom to explore an area of science that one is passionate about. I hope to always emulate the curiosity and passion you have shown as a supervisor in my future career. **Jan Deprest** your passion, support, kindness and encouragement has been invaluable to me. The discussions and thesis corrections have taught me so much, and have ensured that I achieved my best. I have learnt a lot about the clinical problem, and the clinical impact of this work through your expertise and experience. It is extremely motivating to work with a clinical leader and pioneer in this field, who is passionate, optimistic, and can see the real-world positive impact that Artificial Intelligence can have in the clinic.

I'm deeply grateful to **Tom Vercauteren**, who since my first e-mail, prior to this position, has provided me with guidance, support, and kindness. I would not have been able to achieve all the exciting work presented in this thesis without your supervision. I am very fortunate to have had the opportunity to learn so much about deep learning and medical imaging analysis from you and your team at King's College London.

Thank you, Frederik Maes and Dominique Van Schoubroeck, my **internal jury members**, for your continuous guidance and support. I would like to thank Frederik for your detailed comments. Thank you, Chris de Korte, and Aly Youseff, my **external jury members**, for the time dedicated to critically reading my thesis and the support I received.

To all my **colleagues and friends at the Louvre** (old and new), thank you for the support, coffee breaks and time together. To Katya, first my colleague and now a dear friend, I am lucky to have met you in the lab. To Andrea, Konstantina, Paolo and Marcus you are all pure joy. To

Ahmed, Michelle, Jeroen and Gwen, thank you for your support, it's been a pleasure getting to know your group. To all of you, Marta, Somayeh, Sophie, Dylan, Sjeord, Annette, Stéphanie, Nitin, Mihnea, Alaa, Hasti, Jasleen, Masha, Kate thank you for your friendship in the lab.

To my **colleagues and friends from the clinical team**, thank you for your support, kindness, and enthusiasm. Thank you to Laura Cattani, Susanne Housmans, Dominique Van Schoubroeck, Anne-Sophie Page, Adéla Samesova and Bram Packet. An extra huge thank you and mention must go to Laura Cattani, who has played a fundamental role in this work, from segmenting endless structures, trialling software, and answering hundreds of questions over the years. I could not have wished for a better, more patient, and kind clinical partner. Also, a big thank you, to Leen Mortier and Petra Stevens for all the administrative help over the years.

To all my **colleagues and friends at King's College London**, thank you for welcoming me into your team. Thank you everyone in the CAI4CAI group, also to Sebastien Ourselin, Pritesh Mehta, Muhammad Asad, Kerstin Kläser, Sarah Jeffrey and so many more who made me feel so welcome. To everyone from **GE Healthcare** that I have met and worked with, to Jos Stas, Anupriya Gogna, Martin Mienkina, Cindy Smrt, Laura Berg, thank you for your support and feedback over the years. During my PhD, I have had the great fortune to work with many outstanding **collaborators** from many institutes, thank you for your input, help and contribution to the work presented in this thesis. These collaborators include KU Leuven, Belgium (Mahdi Tabassian), UZ Leuven, Belgium (Laura Cattani, Susanne Housmans, Dominique Van Schoubroeck), UCL, UK (Carole Sudre), KCL, UK (Wenqi Li, Muhammad Asad), INESC TEC, Portugal (the BEAS expert - João Pedrosa), MBZUAI, UAE (Mohammad Yaqub).

As well as technical support, support from friends and family was fundamental in the completion of this body of work. I will now share some personal thank you's and notes to those who helped me through the last four years. I would like to thank the **friends that I have met in Leuven**. Diletta and Dante, you have both shown me so much kindness and support (and fed me a lot of pasta over the years). It's been a pleasure having friends like you in Leuven. Thy, thank you for all the support and artistic inspiration. Roxy, hearing your British accent was music to my ears in Leuven, thank you for making Leuven feel more like home.

To my **friends across the channel**, I am extremely grateful for the love, support, celebrations, and kindness you have shown me over the years. Although we may not see each other as often as we would like, as having a friend overseas can be difficult logically speaking, your friendship means the world to me. Matt, thank you for all your support and kindness, and for always celebrating my achievements. Jana, thank you for always championing me and my endeavours. Caleb, thank you for your support and your faith in me during the difficult times I faced finishing this body of work. A heart-felt thank you to Elle, Patrick, Laurence, Penni, Andreea, Laura and Chloe. I am extremely grateful for our friendship. Although, I will be happy to not hear "Are you still studying?" again.

My loving **family**, from the Netherlands, England and Wales, thank you for your unconditional love and support. To my sister Becki, thank you for making London feel like home. To my Uncle Derek, thank you for your deep interest in everything I do, and to my dear Auntie Faye, thank you for always being a strong role model to me.

Finally, to my **dear parents**. Thank you for all the support, opportunities, kindness, 5 o'clock gin and tonics, laughter, phone calls, and unconditional love over the years. With endless science kits given to me throughout my childhood, you have always encouraged me to be curious, and my love and passion for learning and intellectual curiosity are because of you.

To my mum, thank you for your warmth, our country walks and retail-therapy when enjoyment was deeply needed. You have always championed my accomplishments and you have taught me to enjoy life and everything it has to offer. From our dancing in the kitchen, you have always encouraged me not to take life too seriously, to have fun and, above all, to do what makes me happy. To my dad, thank you for instilling in me that nothing is impossible, from watching you learn how to plumb, tile, fix electrics from YouTube (or from your own intuition), I am extremely inspired by your ability to learn and find solutions to problems. Thank you for always offering to check my work for any rogue spelling (even if admittedly you don't fully understand the content). As you say, "You need to do something that makes a difference in the world and more importantly something you enjoy".

Growing up and experiencing a loving, caring, laughter-filled environment with incredible people gave me a chance to pursue and achieve my dreams. Thank you to everyone, for every encounter and every conversation that has led me to where I am today.

Abstract

Analysis of transperineal ultrasound (TPUS) images plays an important role in the assessment of pelvic floor disorders. The analysis tasks include the localisation of pelvic floor landmarks, and the localisation and extraction of a two-dimensional (2D) diagnostic ultrasound image for the diagnosis of a pelvic floor disorder called pelvic organ prolapse. In addition, they include the 2D manual outlining of a pelvic floor structure called the levator hiatus to determine the severity of pelvic organ prolapse. Finally, the localisation and extraction of a sequence of 2D diagnostic ultrasound images used for visual diagnosis of anal sphincter tearing.

Currently, these ultrasound imaging analysis tasks are performed manually, and are time-consuming, prone to human error and inter-observer variability, and require high levels of expertise and training. Therefore, there is a strong clinical need for automation, to standardise ultrasound imaging analysis. In addition, such tools may also contribute to the wider use and acceptance of this imaging modality by clinicians. Automation aims to solve the limitations of manual analysis, however, currently no such automation is available in clinical practice. For an automatic ultrasound imaging analysis pipeline to be implemented in a clinical setting, it should achieve high accuracy, minimise user time, perform robustly on clinical data, and be clinically acceptable. To increase clinical acceptance of algorithms, the framework should closely follow the clinical workflow and allow clinicians to edit a sub-optimal result.

Exploiting recent advances in deep learning and active contours, the following areas are explored: semantic segmentation methods of pelvic floor structures from TPUS images and volumes (i.e., in two-dimensional and three-dimensional), landmark regression methods of plane defining landmarks from TPUS volumes, and interactive segmentation methods of pelvic floor structures from ultrasound images and volumes (i.e., in two-dimensional and three-dimensional). The aim of incorporating segmentation and landmark regression into the proposed pipelines is to increase the interpretability of the pipelines, as visual information is used to perform ultrasound imaging analysis tasks, similar to a clinician.

In this thesis, several deep learning-based pipelines are developed and evaluated that automate several ultrasound imaging analysis tasks for the assessment of pelvic floor disorder. These address gaps identified in the literature. First, a novel plane extraction framework is introduced, which uses a landmark regression convolutional neural network (CNN) to identify landmarks that define the diagnostic plane used by clinicians to assess pelvic organ prolapse. The framework

uses post-processing to extract the desired plane from a TPUS volume without user interaction. Experimental results show that the proposed framework performed with a lower error than the measured inter-observer error and was significantly faster.

A two-step approach is then introduced to extract the hiatal area from a TPUS volume, where the hiatal area is a biomarker used to diagnose hiatal ballooning (an indicator of pelvic organ prolapse). The first step uses the plane extraction framework mentioned, and the second step uses a segmentation 2D CNN to outline the border of the levator hiatus. Experimental results show that the pipeline performed within the measured inter-observer error and significantly reduced the time required by 120 seconds. Unfortunately, despite the promising results, the automatic segmentation of the levator hiatus was not optimal in all cases. In another study the clinician wanted to edit 98% of the CNN's automatic levator hiatus segmentations, despite the computer-observer error being lower than the measured inter-observer error.

To address this limitation, interactive segmentation methods have been widely used to allow clinicians to achieve more accurate and robust results, which is important when the segmentation is used for diagnosis. However, this has not been utilised yet in pelvic floor disorder imaging. Therefore, a 2D interactive segmentation framework was proposed, the method uses a CNN to initialise the active contour method called B-spline Explicit Active Surfaces (BEAS). User-defined points are used to explicitly edit the contour in real-time. This gives clinicians liability for the diagnosis of hiatal ballooning. Experimental results show that the interactive framework improved the clinical acceptability of the automatic segmentation result. In addition, it requires half the perceived workload and 13 seconds less user time than manual segmentation, which is significant.

After focusing on the automated analysis of the levator hiatus, there was a gap in the literature for automated analysis of the anal sphincter. Analysis of the anal sphincter is clinically important to measure anal sphincter tearing, caused by vaginal child birth. Segmentation methods of the external anal sphincter are explored, however, due to the limitations of automatic segmentation a three-dimensional (3D) interactive segmentation method for ultrasound images is proposed. This also addresses a gap in the literature for a robust 3D interactive segmentation method. The 3D interactive segmentation method is based on the 2D interactive segmentation method, which allows real-time explicit editing of an automatic segmentation of the external anal sphincter. Experimental results show the proposed framework was significantly quicker (by 170 seconds per segmentation) and had a significantly lower perceived workload than tools used within the clinic.

Finally, this thesis proposes a guideline-driven automatic framework that extracts the tomographic ultrasound imaging (TUI) sequence of the anal sphincter, used by clinicians to visually assess anal sphincter tearing. The method uses a 3D segmentation CNN to identify the external anal sphincter. The TPUS volume is then rotated to align the external anal sphincter, and eight equally spaced slices are extracted, similar to manual TUI extraction. Experimental results show

the proposed framework was accurate and significantly faster (by 52 seconds per TPUS) than manual analysis.

Impact Statement

Pelvic floor disorders are a major health problem that more than a third of women will experience in their lifetime. TPUS imaging can give clinicians access to diagnostic planes that can be used for the assessment of pelvic floor disorders. The algorithms and frameworks developed in this thesis for the automated analysis of pelvic floor disorders from TPUS volumes can lead to commercial valorisation and advances in academia and the clinical world in various ways.

First, they may benefit the research field of medical and ultrasound imaging analysis in academia. The interactive frameworks developed in this thesis in Chapters 6 and 7 address a general problem of interactive segmentation methods not giving explicit real time control of the surface boundary. In addition, the work presented in Chapter 7 addresses a common problem of active contour models, as most active contour models require good quality initialisation and knowledge to fine tune hyper-parameters for a specific segmentation problem. In Chapter 7, the optimisation algorithm suggests certain hyper-parameters that provide accurate segmentation and easy user-interaction. These algorithms may benefit researchers in medical imaging analysis of other medical fields and imaging modalities. In addition, these algorithms may increase the applicability of B-spline explicit active surfaces in automatic and interactive segmentation frameworks. The future work outlined in Chapter 9 based on the extension of the 3D Slicer module for interactive and automatic segmentation may help researchers in the medical image computing community curate and edit medical segmentations and increase the clinical acceptance of their work.

Second, the work in this thesis can benefit urogynaecology, clinical practice, and healthcare. Automatic frameworks developed in Chapters 4, 5, 6 and 8 may help clinicians diagnose pelvic floor disorders more efficiently. The new algorithms can help them achieve accurate diagnosis with less clinical time, reducing the work load within pelvic floor disorder clinics. The new algorithms may also reduce the error between observers, standardise the clinical workflow, and potentially increase the accuracy and use of TPUS imaging of the pelvic floor.

Automation can be expected to decrease the entry level of clinical skill required to perform these difficult ultrasound imaging analysis tasks. By reducing the entry level skill and clinical workload required to perform such tasks, these algorithms may assess ‘at risk’ patients before the development of symptoms. Faster and more accurate assessment of TPUS images, may result in early diagnosis, faster assessments and more time for patient counselling, and hence,

more patients being assessed. This would have a huge clinical impact, as early treatment and physiotherapy can be given to these patients, potentially reducing the risk of surgical interventions and symptoms later in life. A reduction of complicated and expensive therapies and surgery may reduce the financial burden of the pelvic floor clinic. In addition, these algorithms could motivate industry to develop new products that are directly integrated within ultrasound machine software. The work in this thesis is in collaboration with GE Healthcare (Zipf, Austria) with the intention for clinical implementation in the future.

Future work in Chapter 9 also outlines how other pelvic floor ultrasound imaging analysis tasks performed in the clinic may be automated and implemented into the clinical setting. Work has begun in several areas proposed in Chapter 9, specifically the development of algorithms for levator avulsion classification, hiatal area segmentation from a 4D Cine ultrasound clip, and finally the measurement of anal sphincter tearing. All these aspects may improve the standard of care given in pelvic floor disorder clinics by reducing the burden on clinicians and standardising diagnosis between ‘novice’ and ‘expert’ clinicians.

Finally, as well as improving the accuracy of ultrasound imaging analysis, this work could lead to advances in tools to guide the acquisition of good-quality TPUS images. The clinical impact of this would be significant, as the entry level of skill to perform TPUS imaging would reduce, and the accuracy of manual and automatic ultrasound imaging analysis would increase.

Table of contents

List of figures	25
List of tables	27
Nomenclature	29
List of publications	33
1 Introduction	35
1.1 Pelvic floor disorders	37
1.1.1 Pelvic floor anatomy and function	37
1.1.2 Pelvic floor disorders	38
1.1.3 The levator hiatus	41
1.1.4 The anal sphincter complex	41
1.2 Ultrasound imaging assessment of the pelvic floor	42
1.2.1 Ultrasound imaging	43
1.2.2 TPUS imaging of the pelvic floor	48
1.2.3 Imaging the levator hiatus	50
1.2.4 Imaging the anal sphincter	51
1.2.5 TUI sequence of the pelvic floor	52
1.3 Deep learning techniques	53
1.4 Motivation of this work and thesis overview	58
1.4.1 Motivation	58
1.4.2 Thesis objectives	59
1.4.3 Thesis outline and contributions	60
2 Literature review of US imaging analysis	63
2.1 Segmentation techniques for US images	63
2.1.1 Segmentation with low-level feature extractors for US images	64
2.1.2 Segmentation with active contours for US images	64

Table of contents

2.1.3	Segmentation with supervised machine learning for US images	66
2.1.4	Segmentation with machine learning and active contour models for US images	67
2.2	Interactive segmentation methods in US images	68
2.3	Segmentation of pelvic floor structures from US images	70
2.4	Plane detection techniques in US images	71
2.5	Conclusion and addressing gaps in the literature	72
3	Three dimensional segmentation of pelvic floor structures	75
3.1	Single-class urethra segmentation	76
3.1.1	Introduction	76
3.1.2	Material and methods	77
3.1.3	Results	80
3.1.4	Discussion	81
3.1.5	Conclusion	83
3.2	Multi-class segmentation	83
3.2.1	Introduction	84
3.2.2	Material and methods	84
3.2.3	Results and discussion	86
3.2.4	Conclusion	87
4	Automatic extraction of the C-plane from a transperineal ultrasound volume	89
4.1	Introduction	89
4.2	Material and methods	91
4.2.1	Data and manual detection of the C-plane	91
4.2.2	Overall framework	92
4.2.3	Regression of extreme coordinates	93
4.2.4	Evaluation methodology	94
4.2.5	Implementation details	95
4.3	Results and discussion	95
4.4	Conclusion	98
5	Automatic segmentation of the hiatal area from a transperineal ultrasound volume	101
5.1	Introduction	102
5.2	Material and methods	103
5.2.1	Manual C-plane detection	103
5.2.2	Proposed biomarker extraction pipeline	104
5.2.3	Description of the biomarker pipeline	106
5.2.4	Implementation details	107

Table of contents

5.2.5	Data collection	108
5.2.6	Quantitative metrics for evaluation	110
5.2.7	Computer-observer, intra-observer and inter-observer differences	111
5.2.8	Statistical analysis	111
5.3	Results	112
5.4	Discussion	115
5.5	Conclusion	118
6	Interactive two-dimensional segmentation of the levator hiatus	119
6.1	Introduction	119
6.2	Material and methods	121
6.2.1	Proposed pipeline	121
6.2.2	Data collection	125
6.2.3	Experimental details	126
6.2.4	Implementation details	127
6.3	Results	128
6.4	Discussion	130
6.5	Conclusion	131
7	Interactive three-dimensional segmentation of the external anal sphincter	133
7.1	Introduction	133
7.2	Material and methods	135
7.2.1	BEAS-based evolution	135
7.2.2	Interactive framework	136
7.2.3	BEAS hyper-parameter tuning algorithm	136
7.2.4	Data collection and experimental details	138
7.3	Results	139
7.4	Discussion	141
7.5	Conclusion	142
8	Automatic extraction of a tomographic ultrasound imaging sequence from a transperineal ultrasound volume for analysis of the anal sphincter complex	143
8.1	Introduction	143
8.2	Material and methods	145
8.2.1	Acquisition protocol	146
8.2.2	The proposed pipeline	146
8.2.3	Data collection	149
8.2.4	Evaluation methodology	150
8.2.5	Implementation details	150

Table of contents

8.3 Results	150
8.4 Discussion	151
8.5 Conclusion	153
9 General discussion and future work	155
9.1 General discussion	155
9.1.1 My technical contributions for automating pelvic floor disorder assessment	158
9.2 Future work	162
9.2.1 Automatic levator hiatus analysis and computer aided diagnostics . . .	163
9.2.2 Interactive segmentation	165
9.2.3 Automatic anal sphincter analysis	166
9.2.4 Clinical implementation and improvement	167
9.2.5 Clinical deployment and product development	169
9.3 Final remarks	176
10 Scientific acknowledgements, personal contribution and conflicts of interest	179
10.1 Scientific acknowledgements	179
10.2 Personal contribution	180
10.3 Conflicts of interest	180
11 Video links	181
References	183

List of figures

1.1	Anatomy of the pelvic floor in the inferior (top) and lateral (bottom) view from Lasak <i>et al.</i>	38
1.2	Life span model causal factors of pelvic floor disorders from DeLancey <i>et al.</i> . .	40
1.3	Schematic diagram of the anal sphincter complex.	42
1.4	Representation of the US wave propagating through tissue.	43
1.5	Incident US wave on the skin barrier and subsequent interactions.	44
1.6	Specular and diffuse reflection of the US wave.	44
1.7	Outlined speckle in several pelvic floor images.	45
1.8	Incomplete view of the levator hiatus in an US image.	46
1.9	Artifacts and acoustic shadowing in TPUS volumes.	47
1.10	4D View Software (GE Healthcare; Zipf, Austria) multi-planar view of the pelvic floor.	49
1.11	Transducer placement to assess the hiatal area from Dietz <i>et al.</i>	50
1.12	Transducer placement to assess the anal sphincter complex from Dietz <i>et al.</i> . .	50
1.13	The plane of minimal hiatal dimensions and corresponding axial plane of the levator hiatus.	51
1.14	TUI for levator avulsion assessment from Dietz <i>et al.</i>	52
1.15	TUI for anal sphincter assessment from Guzmán Rojas <i>et al.</i>	53
1.16	A generic CNN architecture trained to perform multi-class segmentation from a TPUS volume.	57
3.1	HighRes3DNet network architecture used for volumetric image segmentation. .	77
3.2	Effect of post-processing a CNN segmentation output.	79
3.3	Automatic segmentation of the urethra superimposed on the input TPUS volume.	81
3.4	Automatic urethra segmentation of three separate TPUS volumes (specifically the lowest, highest and average performing result).	82
3.5	Example of traditional <i>mixup</i> on two TPUS images from a patient at rest and at pelvic floor contraction.	85

List of figures

3.6 Automatic segmentation result for multi-class segmentation of the levator ani muscle, the urethra and the symphysis pubis.	86
4.1 Desired 2D C-plane and corresponding mid-sagittal (MS) plane.	90
4.2 Visualisation of the steps to generate the ground-truth heatmaps used in automatic C-plane detection.	91
4.3 3D visualisation of the C-plane.	93
4.4 C-plane extraction pipeline.	93
4.5 The Euclidean distance and angular difference measured for automatic C-plane extraction.	97
4.6 Visualisation of the automatic and gold standard C-planes extracted.	97
5.1 Typical TPUS acquisition and evaluation screen on Voluson systems (GE Healthcare; Zipf, Austria).	103
5.2 Overall levator hiatus analysis pipeline.	105
5.3 Examples of the automatically and manually extracted C-plane position within the TPUS volume and the corresponding extracted hiatal segmentations.	113
6.1 2D interactive segmentation pipeline.	121
6.2 Screenshot of "Beyond"'s GUI for 2D interactive segmentation.	124
6.3 Visual representation of the hiatal area segmentation obtained with clinical tools and the proposed interactive segmentation tool.	128
6.4 Time taken to delineate the hiatal area to a clinically acceptable level using clinical tools and the proposed interactive segmentation tool.	129
7.1 Interactive segmentation of the external anal sphincter using DeepBEAS3D.	134
7.2 Number of BEAS knots vs the ease of user interaction.	137
7.3 Visualisation of segmentations of the external anal sphincter obtained with VOCAL and DeepBEAS3D.	139
7.4 Time taken to delineate the external anal sphincter.	140
8.1 A TUI sequence of a <i>normal</i> anal sphincter.	145
8.2 Proposed pipeline of the automatic TUI extraction algorithm.	147
8.3 Visual representation of the automatic and gold-standard TUI sequences.	151
9.1 High quality and low quality acquisition of the external anal sphincter.	168
11.1 QR code for a video of the 2D interactive segmentation framework.	181
11.2 QR code for a video of the 3D interactive segmentation framework.	182

List of tables

3.1	Average performance of HighResNet for urethra segmentation by employing comparison with the manual segmented labels for each TPUS volume.	80
3.2	Average intra-observer variability between five manually segmented urethras on a test-retest basis.	81
4.1	The Hausdorff distance and <i>visual Turing test</i> results for automatic C-plane extraction.	96
5.1	Characteristics of the study population for automatic hiatal area extraction.	109
5.2	Turing test score per transperineal ultrasound volume for the hiatal area extraction pipeline.	112
5.3	COD, IAOD and IEOD of C-plane detection metrics from the hiatal area extraction pipeline.	114
5.4	COD, IAOD and IEOD errors of hiatal segmentation metrics from the hiatal area extraction pipeline.	114
5.5	COD, IAOD and IEOD limits of agreement of all pipeline metrics for the hiatal area extraction pipeline.	115
6.1	Average perceived weighted workload score from the NASA-TLX questionnaire and individual sub-scale scores for 2D interactive segmentation.	130
6.2	Perceived weighted workload score from the NASA-TLX questionnaire and individual sub-scale scores mid way and at the end of the experiment for 2D interactive segmentation.	131
7.1	The perceived weighted workload score from the NASA-TLX questionnaire and sub-scale scores for VOCAL and DeepBEAS3D.	140
8.1	Clinical acceptability, time taken and slice separation of automatic TUI extraction.	151
8.2	TUI sequence slice number and corresponding clinical acceptability.	151
9.1	Classification Rule 11a of the MDR.	174

Nomenclature

Other Symbols

\vec{v}_{av} Average eigenvector

X_{cm} Centre of mass

Acronyms / Abbreviations

1D One-dimensional

2D Two-dimensional

3D Three-dimensional

AI Artificial intelligence

AIUM The American Institute of Ultrasound in Medicine

BEAS B-spline explicit active surfaces

BMI Body mass index

C-plane Plane of minimal hiatal dimensions

Cine loop 4D volumetric clip

CNN Convolutional neural network

CNNs Convolutional neural networks

COD Computer observer difference

CPU Central processing unit

CT Computed tomography

DEXTER Deep Extreme Cut

Nomenclature

EAS External anal sphincter

FDA Food and Drug Administration

GDPR General Data Protection Regulation

GPU Graphical processing unit

GS Gold Standard

GUI Graphical user interface

HD Hausdorff distance

IAOD Intra observer difference

IAS Internal anal sphincter

IEOD Inter observer difference

ITN Iterative transformation network

IUGA International Urogynecological Association

LAM Levator ani muscle

MDR Medical Device Regulation

MP Multiplanar

MRI Magnetic resonance imaging

MS Mid-sagittal

NASA TLX NASA Task Load Index

OASIs Obstetric anal sphincter injuries

OV Omniview

PCA Principle component analysis

POP Pelvic organ prolapse

PFD Pelvic floor disorder

ReLU Rectified linear unit

RMSprop Root mean squared propagation

RRF Regression Random Forests

SD Standard deviation

SP Symphysis pubis

TPUS Transperineal ultrasound

TUI Tomographic ultrasound imaging

UGIR Uncertainty-guided efficient interactive refinement

US Ultrasound

VOCAL Virtual Organ Computer-Aided Analysis

List of Publications

First Author Peer-Reviewed Journal Papers

1. (PUBLISHED) **Williams, H.**, Cattani, L., Van Schoubroeck, D., Yaqub, M., Sudre, C., Vercauteren, T., D'hooge, J., Deprest, J. Automatic Extraction of Hiatal Dimensions in 3-D Transperineal Pelvic Ultrasound Recordings. *Ultrasound in Medicine & Biology*, 2021.
2. (IN PREPARATION) **Williams, H.**, Pedrosa, J., Asad, M., Vercauteren, T., Deprest, J., D'hooge, J. DeepBEAS3D: Deep Learning and B-spline Explicit Active Surfaces for 3D Interactive Segmentation. *TMI*.

First Author Peer-Reviewed Conference and Workshop Papers

1. (PUBLISHED) **Williams, H.**, Cattani, L., Li, W., Tabassian, M., Vercauteren, T., Deprest, J., D'hooge, J. 3D Convolutional Neural Network for Segmentation of the Urethra in Volumetric Ultrasound of the Pelvic Floor. *IEEE IUS* 2019.
2. (PUBLISHED) **Williams, H.**, Cattani, L., Yaqub, M., Sudre, C., Vercauteren, T., D'hooge, J., Deprest, J. Automatic C-Plane Detection in Pelvic Floor Transperineal Volumetric Ultrasound. *ASMUS workshop at MICCAI* 2020.
3. (PUBLISHED) **Williams, H.**, Pedrosa, J., Cattani, L., Housmans, S., Vercauteren, T., Deprest, J., D'hooge, J. Interactive Segmentation via Deep Learning and B-Spline Explicit Active Surfaces. *MICCAI* 2021.
4. (PUBLISHED) **Williams, H.**, Cattani, L., Vercauteren, T., D'hooge, J., Deprest, J. Automatic Tomographic Ultrasound Imaging Sequence Extraction of the Anal Sphincter. *ASMUS workshop at MICCAI* 2021.

First Author Peer-Reviewed Conference Abstract

1. (PUBLISHED) **Williams, H.**, Cattani, L., Yaqub, M., Vercauteren, T., Deprest, J., D'hooge, J. Data augmentation to aid 3D convolutional neural network segmentation of landmarks in a small volumetric ultrasound dataset of the pelvic floor. *IEEE IUS* 2020

Chapter 1

Introduction

Pelvic floor disorders (PFD) include pelvic organ prolapse (POP), urinary and faecal incontinence, defecatory disorders (constipation), pelvic pain and sexual dysfunction [1, 2]. More than a third of women will experience at least one PFD during their lifetime, their prevalence is only expected to increase in the next 10-30 years, due to aging of the population [3]. Normal function of the pelvic floor is dependent on the integrity of its support apparatus (bony pelvis, pelvic floor muscles and fascias). These hold the pelvic organs in their normal position. The assessment of patients with pelvic floor complaints begins with a thorough history and physical examination. However, structural defects in the pelvic floor support structures and anal sphincter may not be visible during inspection or even not detected through clinical examination. Therefore, patients may benefit from further assessment with adjunctive tests, including medical imaging. There are various imaging modalities that can visualise the pelvic floor, of which ultrasound (US) and magnetic resonance imaging (MRI) are most commonly used [4].

In this thesis, the focus is on US, as it allows for real-time three-dimensional (3D) imaging of the pelvic floor. It is readily available in a clinical setting, safe for patients, cheap in comparison to other modalities, well tolerated by patients, and allows access to both diagnostic two-dimensional (2D) planes as well as 3D and 4D assessment [5]. To assess patients with PFDs with US (specifically POP and/or anal sphincter tearing), the clinician must perform several imaging analysis tasks. Currently, these are performed manually within the clinic or in complex cases off-line, but manual analysis is time-consuming, prone to errors, and requires high levels of clinical training and expertise. A potential solution to manual analysis is the development of automated tools to assist in these tasks. To develop these tools, the use of artificial intelligence (AI) and deep learning methodology is being explored.

AI has shown to accurately automate US imaging analysis tasks such as segmentation and classification [6], sometimes achieving a human-level performance (i.e., when the AI algorithms computer-observer error is lower than the known clinical inter-observer variability) [7]. AI

Chapter 1. Introduction

has also shown to perform superior to traditional imaging analysis methods in other fields of medicine [8].

AI has been described as a disruptive technology and has proven to perform with impressive accuracy and sensitivity in identifying imaging abnormalities, organ and/or landmark locations and segmentation. AI is predicted to bring about a significant change in the way and where healthcare is delivered to the patient in the future (i.e., helping clinicians perform difficult US imaging analysis tasks) [9]. When shown to be effective, AI could make US imaging and analysis more accessible to clinical settings around the world. One weakness of AI and automatic US imaging analysis is that it can cause medical liability problems, such as when it would fail to diagnose or over-diagnose an abnormality that in extreme cases could lead to unnecessary interventions if that decision would not be recalled by a clinician.

It is important to note that doctors have a duty to apply the standard of care for their medical field independently, regardless of the output of an AI algorithm [10, 11]. Therefore, how an AI-based algorithm is integrated into the clinical setting is important. The Food and Drug Administration (FDA) requires a level of explanation for the development of algorithms that inform medical treatment (i.e., information for traceability, transparency and an explanation for the development of machine learning and deep learning models) [12]. Any ambition of clinical use would also in Europe qualify the AI tool as a medical device, to be developed under the current Medical Device Regulation (MDR) and in the future all AI technology in Europe will have to be developed under the draft AI act. In addition, the data used in developing such a tool will have to satisfy the requirements of General Data Protection Regulation (GDPR).

In this thesis, the emphasis is on the creation of automatic pipelines that follow the clinical approach closely, while being interpretable to the clinician. The approach was to develop algorithms that use visual landmark information, similar to the decision-making process used by clinicians today when analysing transperineal ultrasound (TPUS) images. In addition, the developed pipelines have multiple steps that can be extended to include user interaction tools, allowing clinicians to make corrections if necessary to increase clinical acceptance of this work. Assessment of PFD can be identified as a medical field that would benefit from the deployment of AI-driven technologies. AI-driven technologies may reduce clinical time, reduce the need for high-level expertise, allow more clinical centres and hence their patients to benefit from US assessment of the pelvic floor, and standardise practice between clinicians and medical centres.

Motivated by these factors, work presented in this thesis is based on several deep learning-based, computer-aided, automated US imaging analysis tasks for the assessment of PFD. The rest of this chapter summarizes the principal topics covered in this thesis. First, the basic concepts of pelvic floor anatomy and PFD are introduced in **Section 1.1**, with a strong emphasis on the levator hiatus and the anal sphincter complex. **Section 1.2** is dedicated to US imaging, starting with the fundamental concepts, and going into further detail on TPUS imaging and pelvic floor imaging techniques to help in the assessment of PFD. A section is dedicated to a general introduction to

deep learning techniques (1.3). Finally, the motivation and thesis objectives and contributions are outlined in **Section 1.4**.

1.1 Pelvic floor disorders

In this thesis, the emphasis will be on female pelvic anatomy and their pelvic floor, and pelvic floor disorders. The female pelvic floor includes all structures within the bony pelvis: those between the symphysis pubis and the coccyx, and from one pelvic sidewall to the contralateral pelvic sidewall. The pelvic floor supports most of the lower urinary tract, the reproductive organs and the distal end of the gastrointestinal tract and contains also neuromuscular components of their support [13, 14].

1.1.1 Pelvic floor anatomy and function

The pelvic floor is a complex system formed from a collection of muscles, ligaments and fasciae (i.e., a sheet of connective tissue), that are attached directly and indirectly to the bony ring of the pelvis. Please refer to Fig. 1.1 for a schematic diagram of the inferior (top) and lateral (bottom) view of the pelvic floor system. The pelvic floor acts as a support to the abdominal and pelvic organs precluding them from descending through the opening of the pelvis. The urethra, vagina and anorectum traverse the pelvic floor. Passive and active components of the pelvic floor system provide pelvic floor support, maintain continence, and facilitate urination and bowel movement. In women, the pelvic floor also facilitates sexual function and forms part of the birth canal [15]. Specifically, the fascia and ligaments provide passive elastic support, while the muscles (mainly the levator ani muscle) provide active support [16].

The levator ani muscle is the largest component of the pelvic floor. It is a dome-shaped broad muscular sheet that supports the pelvic and abdominal organs, helps maintain intra-abdominal pressure, assists in breathing and sexual function, and can adjust to allow defecation and urination [17]. Typically, the levator ani muscle is divided into three collections of muscle fibres which are named after their insertions: the m. pubococcygeus, m. iliococcygeus, and m. puborectalis. The puborectalis forms a sling around the lower rectum and works with the internal and external anal sphincter to support defecation and maintain anal continence. Maintenance of urinary continence is, amongst others, controlled by the pubococcygeus [18]. The pubovaginalis muscle forms the medial part of the pubococcygeus and supports the vagina. It provides core stability, and in childbirth helps position the fetal head. Finally, the iliococcygeus provides a secure anchor for the pelvic floor [19]. Therefore, any weakening or structural defect of the levator ani muscle, ligaments and fasciae can cause failure of pelvic floor support, in turn potentially leading to symptoms referred to as PFD.

Chapter 1. Introduction

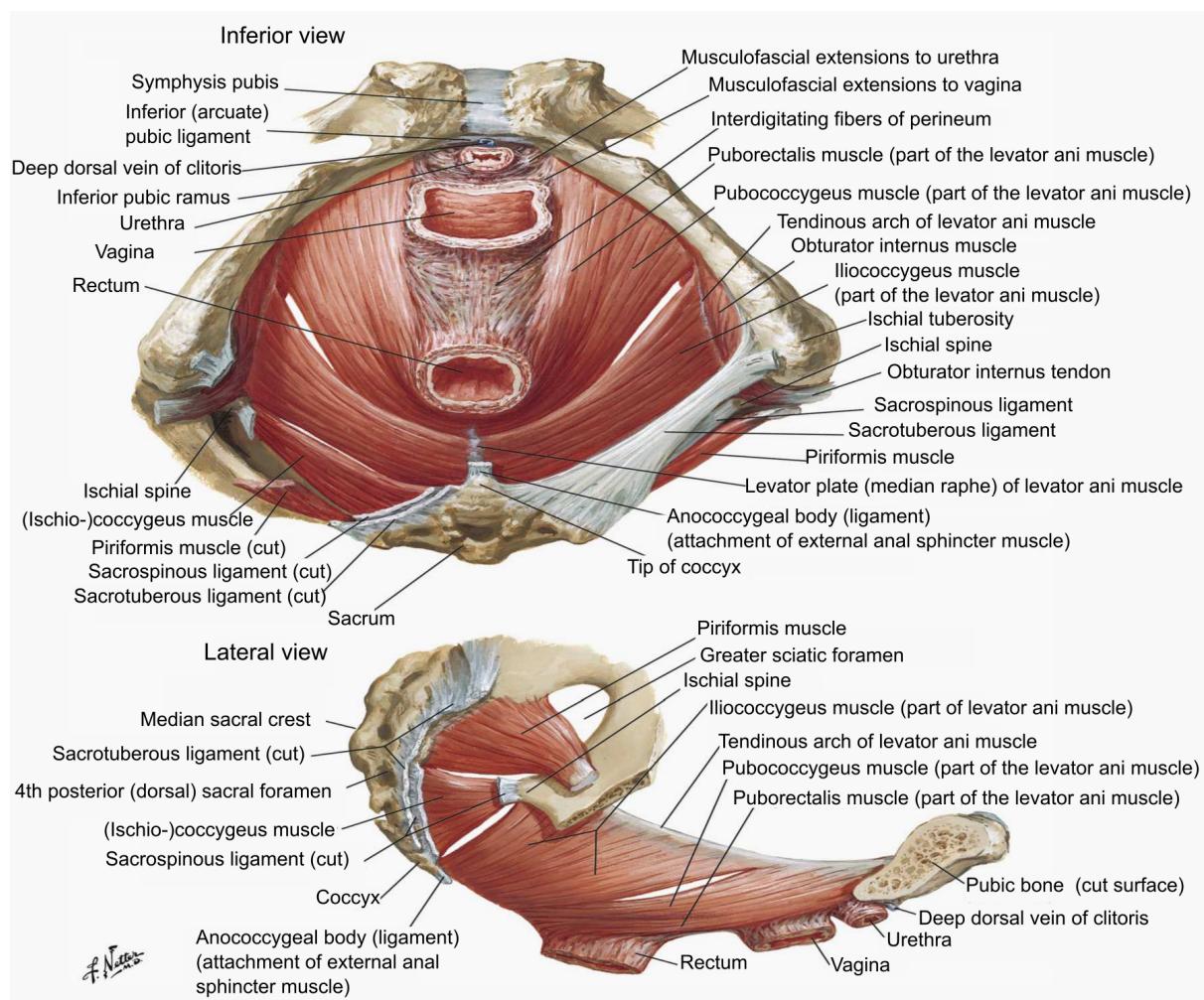


Fig. 1.1 Adapted and reprinted with the permission of John Wiley & Sons [20] (original image from Netter, all rights reserved, www.netterimages.com). Inferior (top) and lateral (bottom) view of the pelvic floor muscles that are traversed by the urethra, vagina and anal canal.

1.1.2 Pelvic floor disorders

What are pelvic floor disorders? Pelvic Floor Disorders (PFD) are a group of conditions including POP, urinary and anal incontinence, pelvic pain, constipation and/or sexual dysfunction [21–23, 14]. PFD are a major health problem that affects 25% of women over 20 years, and 50% of women over 80 years [24]. PFD can affect women at several stages in their life cycle, such as during pregnancy, post-partum or during menopause [25].

Most PFD find their origin in weakening of the pelvic floor, which can be caused by stretch or tears to the pelvic floor muscles and their attachments, and damage to the pudendal nerve [13]. Since the levator ani muscle supports all three compartments (i.e., urethra, vagina and anus), its weakness and tears can impair function of any or all structures it supports. Many surveys suggest there is a coexistence of symptoms of urinary, genital and fecal dysfunction. Patients

with fecal incontinence also have urinary incontinence in 24% to 53% cases, and POP in 7% to 22% [13]. POP occurs when pelvic organs descend from their normal position. Anatomically, POP is categorised as urethrocele, cystocele, utero-vaginal prolapse, enterocele and rectocele - the classification is dependent on the anatomical structure or structures that protrude in the vagina or beyond [26].

Prevalence of PFD PFD are currently under-reported and under-treated, most likely because of social taboos and personal shame regarding these common symptoms [27]. Studies suggest that the prevalence of one or more PFD range from 37 to 67.5% [3, 28], with the prevalence of individual PFDs being: 17.1% for urinary incontinence [29], 30-50% for sexual dysfunction [30], 2.9% for POP [29] and 9.4% for fecal incontinence [29]. PFD are a major public health problem in women, who run a lifetime risk of 11.1% to undergo a single operation for POP and/or urinary incontinence [31]. A study projected that the number of American women with at least one PFD will increase from 28.1 million in 2010 to 43.8 million in 2050 (mainly due to aging of the population), with the number of women with POP expected to increase by 46%, urinary incontinence by 55%, and fecal incontinence by 59% [32]. The projected increase in the prevalence and the surgical risk has important implications for public health and the field of gynaecology [32]. It is logical to assume that this will increase the burden on health care systems and budgets. The above underscores the relevance of accessibility of US imaging and analysis, and the need to standardise the clinical diagnosis of prolapse and anal sphincter tearing. This could lead to early detection of prolapse and anal sphincter tearing in ‘at-risk’ patients, in order to introduce preventive strategies [33].

Causes of PFD The pathophysiology of PFD is multi-factorial. DeLancey graphically summarised several causal factors in a ‘life-span model’, shown in Fig. 1.2. The life-span model plots a woman’s pelvic floor function in three main life phases. In Phase I, the woman develops her maximum pelvic floor functional capacity, which is dependent on her genetic background, nutrition and environment. In Phase II, the pelvic floor function may deteriorate due to the amount of injury and potential recovery that occurs during pregnancy and after vaginal birth. Finally, in Phase III, the pelvic floor function further deteriorates due to the aging process [24]. To note, many other factors may increase the rate of deterioration further.

Women will experience symptoms when their pelvic floor function depletes past a threshold. When, and if, this happens depends on several factors: the maximum functional capacity of the pelvic floor in Phase I, the rate of decline, stress their lifestyle places on the pelvic floor (i.e., chronic constipation, occupational lifting, obesity, chronic cough), and the events in Phase II (i.e., vaginal childbirth, obstetrical interventions, muscle avulsion, among others) [24]. Vaginal childbirth is a major risk factor, the increased risk of PFD can be caused by delivery trauma, in particular levator ani muscle injury and levator avulsion (the detachment of the levator ani muscle

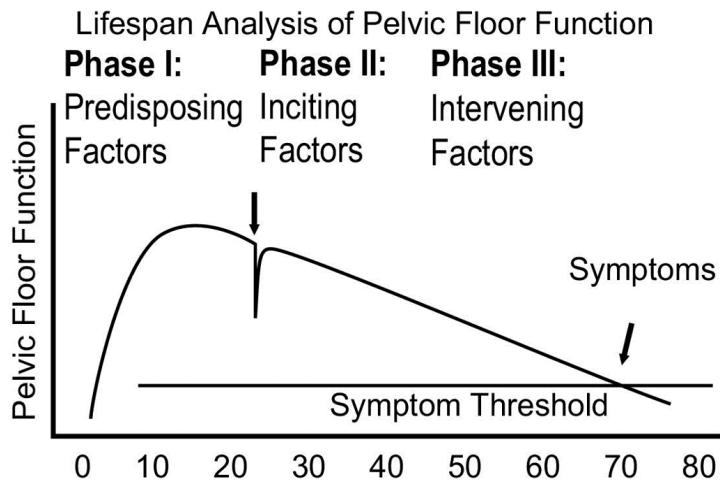


Fig. 1.2 Reprinted with the permission from Elsevier [24]. Integrated lifespan analysis of pelvic floor function. This graphical display of pelvic floor function tracks the functional reserve throughout several phases of woman's life. The pelvic floor function grows to a maximum capacity in Phase I. Vaginal birth affects pelvic floor function in Phase II, and age-related deterioration occurs in Phase III. Once pelvic floor function declines past the symptom threshold the woman will experience symptoms, when, and if, this occurs depends on many factors (i.e., genetics, chronic constipation, heavy-lifting, BMI, etc.).

from the pubic ramis). One study suggests that levator avulsion occurred in approximately 36% of women who deliver vaginally [34].

Major and minor trauma of the levator ani muscle caused by vaginal childbirth, can lead to a larger levator hiatus [34] and POP. Simulations can explain levator ani muscle injury during vaginal childbirth. Several studies showed the pelvic floor muscles can stretch up to 300% of its normal limit during vaginal childbirth [35, 36], which exceeds the physiologic limits of striated muscles [37]. Finally, women can also experience obstetric anal sphincter injuries (OASIs) during vaginal childbirth. Endoanal US imaging studies have shown that structural sphincter injuries occur in up to a quarter of all women after first vaginal delivery, however not all are symptomatic [38].

There are several factors, such as: chronic constipation, high BMI, chronic respiratory diseases, that can accelerate the deterioration of pelvic floor function (i.e., in Phase III of the life-span model) by increasing the stress and strain over time [24]. Women with PFD are more likely to have a higher BMI than asymptomatic women [25]. A higher BMI is associated with increased intra-abdominal pressure, which causes excessive loading of the pelvic floor muscle. This can weaken the pelvic floor over time [39, 25]. Chronic repository disease can cause repetitive motion trauma that challenges the pelvic floor, and hence, it may accelerate the decline of pelvic floor function [24].

Ageing plays a role in the deterioration of existing PFDs and development of POP, as the pelvic floor tissues weaken with age. Each year of life increases the risk of developing POP by

10% [37]. Ageing may further increase muscle loss of important pelvic floor structures (i.e., muscle loss of the urethra with ageing can reduce the urethra closure pressure and increase the risk of incontinence) [40]. This is supported by a study showing that 9.7% of women aged between 20 to 39 years report at least one PFD, 26.5% of women aged 40 to 59 years, 36.8% of women aged 60 to 79 years, and 49.7% of women aged 80 years or older [23]. As women live longer and lead longer active lives, one can expect that PFD will become an increasingly important issue for public health. Menopause may also be a contributing factor, due to the relationship between hypo-estrogenism and the quality of pelvic floor muscle and connective tissues, yet the relationship between menopause and POP has not been experimentally validated [37].

1.1.3 The levator hiatus

The levator ani muscle supports the pelvic viscera, and micro and macro trauma of the levator ani muscle and levator avulsion can predispose women to POP. The biomechanical properties of the levator ani muscle have also been shown to influence the likelihood of POP [41]. Specifically, the biomechanical properties of the levator ani muscle are related to the levator hiatus. The levator hiatus is part of the birth canal, and it defines the ‘hernial portal’ through which POP can develop. The area of the levator hiatus is called the ‘hiatal area’, and its dimensions are a diagnostic biomarker used to assess the presence and/or extent of POP. If the hiatal area at maximum Valsalva (i.e., when the patient forcefully attempts to exhale against a closed airway) is larger than 25cm^2 , the hiatus is referred to as ballooning, and may predict the later occurrence of POP [41]. Therefore, imaging the levator hiatus to measure its area is a fundamental US imaging analysis task focused on in this thesis, and described in more detail in Section 1.2.3. A ballooning hiatus and a healthy hiatus imaged with US, are shown in Section 1.2.3, Fig. 1.8 A (healthy) and Fig. 1.8 B and C (hiatal ballooning).

1.1.4 The anal sphincter complex

The anal sphincter complex, which surrounds the anal canal, is key in maintaining anal continence. It consists of three muscles: the smooth longitudinal muscle layer called the internal anal sphincter, the striated external anal sphincter, and the puborectalis muscle (which is part of the levator ani muscle) [42]. The external anal sphincter muscle provides active control of the anal canal, and the internal anal sphincter acts involuntarily. A schematic drawing of the anal sphincter complex is shown in Fig. 1.3. The most common cause of fecal incontinence in women are obstetric perineal tears associated with anal sphincter injuries [43]. Common symptoms of anal incontinence include flatus incontinence, passive soiling, incontinence of liquid or of solid stool. These symptoms can occur shortly after delivery, present many years later or a

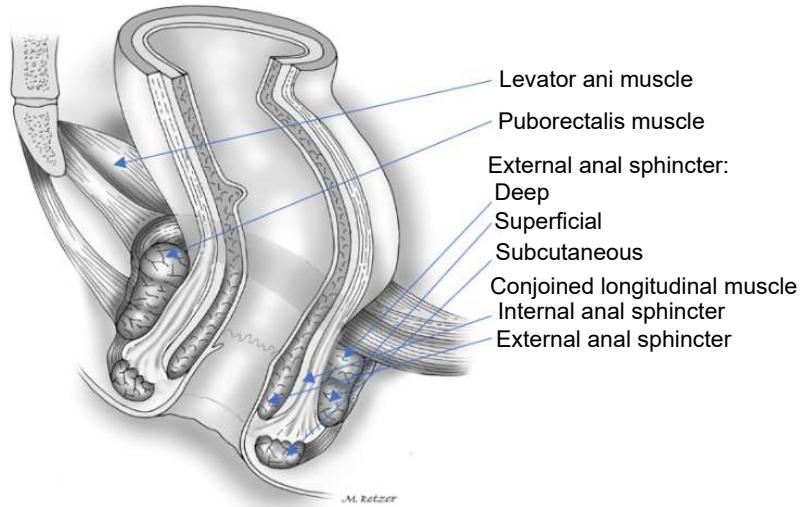


Fig. 1.3 Reprinted with the permission from AME Publishing Company [46]. Schematic diagram of the anal sphincter complex with key muscles outlined. The rectum passes through the pelvic floor muscle and becomes the anus, surrounded by the anal sphincter complex.

combination of both [43]. Studies suggest 10% of women with OASIs develop symptoms at least 5 years after delivery [44], and up to 30% of women develop symptoms after 10 years [45], severely compromising women's quality of life [43]. The detection of anal sphincter tearing can be achieved through US imaging, and this is explored in more detail in sections 1.2.4 and 1.2.5.

1.2 Ultrasound imaging assessment of the pelvic floor

The pelvic floor can be considered a mechanical three-dimensional system that acts in harmony and allows for urinary and anal continence, sexual satisfaction, and vaginal delivery [13]. Pelvic floor imaging can be used as a diagnostic tool to explain the cause of symptoms a patient may present, by identifying what are believed to be causative mechanisms (i.e., anatomical defects such as hiatal ballooning, levator avulsion and/or anal sphincter tearing; or functional problems, such as co-contraction) [4, 47].

Today several modalities are clinically available to image the pelvic floor system, including: 2D and 3D/4D US and pelvic floor MRI. Many guidelines are based on pelvic floor US, because it is relatively cheap and therefore widely accessible within clinics, relatively comfortable for the patient, non-ionising, and can image the pelvic floor system in 3D and in real-time [48]. This thesis uses images obtained by 3D TPUS, and the present section therefore covers US imaging of the pelvic floor in more detail.

The following ideas are covered: the basic principles of the physics of US and the effect it has on pelvic floor US, 3D TPUS imaging, tomographic ultrasound imaging (TUI), and finally several acquisition tasks used in this thesis for data collection.

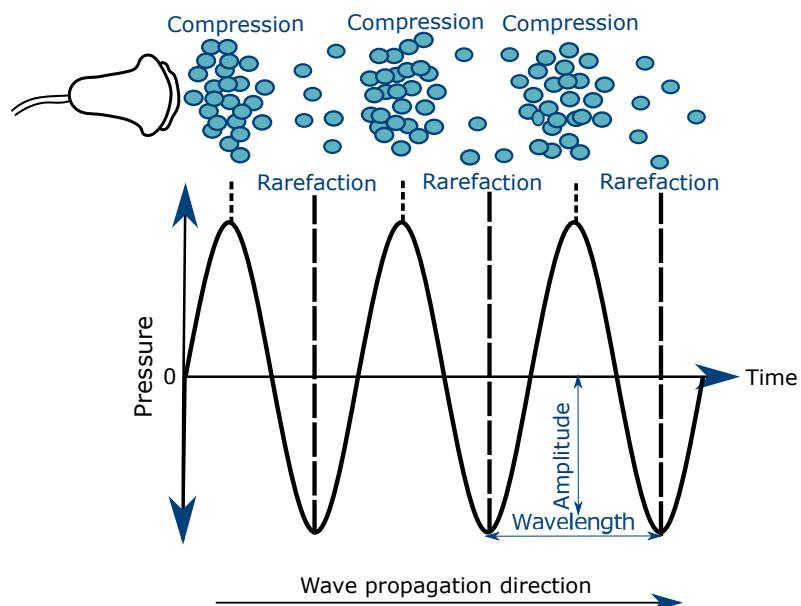


Fig. 1.4 A schematic of the US wave travelling through tissue, creating areas of compression (increased density and pressure) and rarefaction (reduced density and pressure).

1.2.1 Ultrasound imaging

Basic principles of ultrasound

To produce an US image of the pelvic floor, the US machine first insonates sound waves, subsequently receives the echoes of these waves, and finally reconstructs an image of those echoes. In diagnostic imaging of the pelvic floor, piezoelectric transducers produce sound waves with frequencies of 1 to 10 MHz. The specific frequency depends on the target object being imaged. Ultrasonic waves are progressive longitudinal compression waves that displace particles in the medium parallel to the direction of wave motion [49]. A schematic of an US wave is shown in Fig. 1.4, with compression and rarefaction of the US wave visualised regarding the amplitude of the US pressure.

Ultrasound interactions with tissue (reflection, transmission, scattering and attenuation).

US imaging depends on the interactions between the tissue and the propagating US wave, as visualised in Fig.1.5. When the US wave travels through the human body, it traverses many structures and tissues. These tissues and organs have unique properties (i.e., tissue densities and elasticities) that make up their ‘acoustic impedance’. At the boundaries of tissues with different densities and/or elasticities, part of the US wave is reflected, and the remaining part is refracted through the boundary into the tissue distal to the source. The reflection and transmission coefficients depend on the acoustic impedances of the two tissues. Large differences in acoustic impedance lead to a higher reflection fraction of the US wave (e.g., at tissue/bone and air/tissue

Chapter 1. Introduction

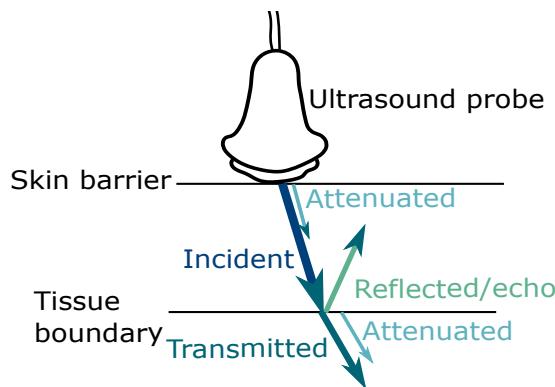


Fig. 1.5 Incident US wave on a skin barrier with US interactions with the tissue boundary shown.

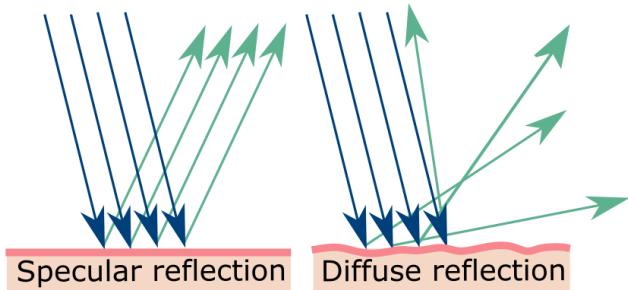


Fig. 1.6 Two US wave fronts incident upon a *perfectly smooth* surface and on a *rough* surface causing specular and diffuse reflection respectively.

transitions), and small differences in acoustic impedance lead to a higher transmission fraction of the US wave (i.e., at similar tissue/tissue transitions). In terms of wave reflection, specular reflections occur on ‘perfectly’ smooth surfaces (i.e., when the imperfections in smoothness of the surface are small in size with respect to the US wavelength), and the reflected wave returns in a single direction. However, tissues are imperfect and can contain irregularities (i.e., when imperfections in smoothness are large in size with respect to the US wavelength). Therefore, diffusion reflection occurs, and the wave is reflected in various directions, as shown in Fig. 1.6.

Reflections do not only occur at tissue boundaries, there are also scattering events throughout the tissue, as tissues are inhomogeneous due to local deviations in tissue density and compressibility. The returning overlapped echoes coming from scattering events may experience interference effects that are constructive or destructive, dependent on the relative phase between the echoes [50]. This interference phenomenon leads to a grainy appearance in the US image, referred to as **speckle**. Speckle reflects the inhomogeneous nature of the tissues. Experienced clinicians can use the appearance of speckle to determine anatomical and functional information. Please refer to Fig. 1.7, which shows an outlined region of speckle in US images of the A) anal sphincter complex, B) the levator hiatus and C) mid-sagittal view of the pelvic floor. Note that the speckle can reduce the visibility of the pelvic floor boundaries, such as the hiatal border, making it difficult for an inexperienced clinician to outline structures. Some US imaging software also includes smoothing post-processing filters, to reduce the level of speckle from US images, as shown in Fig. 1.7 B and C.

Besides scattering, reflection and transmission, tissues also attenuate the US wave as it travels through the body. The initial acoustic energy of the US wave dissipates during propagation, as the acoustic energy is converted into heat. This is due to the absorption of US waves into the tissue, and the US wave’s amplitude decays exponentially. The rate of attenuation depends on the US frequency and the properties of the tissue. Therefore, the US frequency will vary

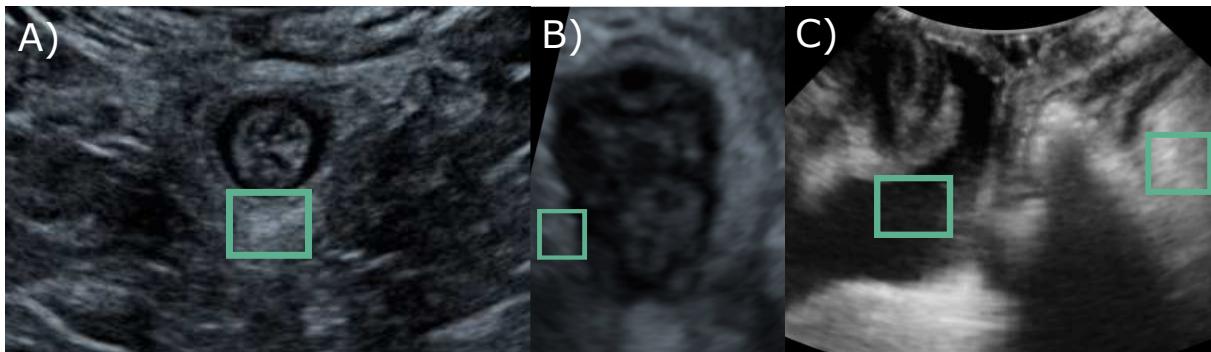


Fig. 1.7 Speckle in an US image of the A) external anal sphincter, B) levator hiatus, and C) the mid-sagittal view of the pelvic floor. Image: UZ Leuven, Belgium, with permission of patient, Software: 4D View (GE Healthcare; Zipf, Austria).

depending on the target anatomical structure imaged and the distance of the structure to the probe. Superficial structures (i.e., structures close to the transducer) are imaged at a higher frequency for a higher axial resolution, while deeper structures (i.e., structures further from the transducer) are imaged at a lower frequency with greater tissue penetration [51], due to increased attenuation of the wave. A high axial resolution is dependent on a short spatial pulse length, which is dependent on a broad transducer bandwidth. Typically, the bandwidth is directly related to the frequency, therefore, a high frequency has a higher axial resolution.

Forming an ultrasound image

The US transducer contains piezoelectric crystals, which generate US waves, when excited with an oscillating electric field. The electric field causes the crystal to expand and contract at the same frequency, which creates a compression/acoustic wave (i.e., the US wave), and the direct opposite phenomenon is used to receive the echo signal reflected from tissues and tissue boundaries. Note, this is a simplified explanation of how medical US transducers generate US waves. Most often, a rectangular pulse drives electrical excitation in the transducer. The frequency response of the transducer then determines the frequency contents of the acoustic wave.

Typically, the same transducer, which is composed of an array of piezoelectric crystals, is used to first transmit the ultrasonic pulse and then receive the reflected echo US waves. Ultrasound echo waves generate signals at the transducer and are electronically processed into a digital image. A digital image is formed by determining the strength of each reflected echo, which determines the intensity of the US image at every pixel/voxel location. The strength depends on the attenuation of the US wave, and on the depth/time duration between the echo received and transmitted by the transducer. Finally, the US machine attempts to give an indication of the echo strength/brightness independent to the depth of structure being imaged.

Chapter 1. Introduction

Ultrasound modes

Several US modes can be used to image the pelvic floor. The B-mode or 2D mode scans a plane through the body, which is not the preferred option for assessing PFD. The most common and informative mode of US imaging for the assessment of PFDs is 3D or 4D (= 3D + time) US. This allows the operator to obtain views that may not be available using B-mode imaging (e.g., the plane of minimal hiatal dimensions used to measure the hiatal area) [52]. 3D imaging can also provide volumetric measurements of the pelvic floor organs. However, this is not yet routinely used in the assessment of PFDs, and 2D contouring remains the standard practice. Finally, 4D imaging allows the clinician to observe the three-dimensional pelvic floor system functioning in real-time.

Ultrasound challenges

Despite the advantages, US has its limitations. US imaging is prone to artifacts, noise, acoustic shadows, and there can be a low contrast between soft tissues and a small field of view compared to other medical imaging modalities. Furthermore, the quality of an US image is operator and patient dependent, and good quality acquisition requires high levels of training and expertise. Specifically, a pathological, ballooning levator hiatus imaged at Valsalva may not be fully captured by the US transducer, so the border of the levator hiatus cannot be fully outlined. As reference, Fig. 1.8A shows a healthy intact levator hiatus. Fig. 1.8B and Fig. 1.8C show a ballooning hiatus that is not fully visible in the US image. Fig. 1.8D also shows an example of the levator hiatus that is not fully captured within the US frame, due to a sub-optimal US acquisition. Which underlines the importance of adequate US acquisition for assessing PFD.

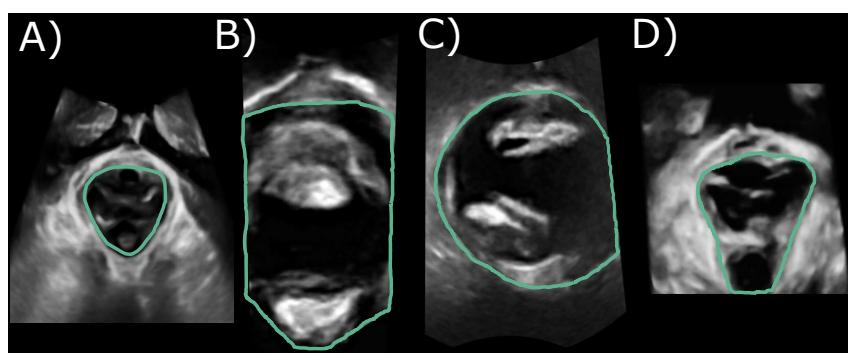


Fig. 1.8 A) Complete view of a healthy levator hiatus in the axial view of the plane of minimal hiatal dimensions, B), C), D) incomplete views of the levator hiatus, due to: hiatal ballooning B,C) and sub-optimal US acquisition D). The green outline highlights the levator hiatus which is used to calculate the hiatal area. Image: UZ Leuven, Belgium, with permission of patients, Software: 4D View (GE Healthcare; Zipf, Austria).

1.2 Ultrasound imaging assessment of the pelvic floor

Acoustic shadowing is a common US artifact caused by adjacent tissues with significantly different acoustic impedances. It can occur when there are strong reflectors or absorbers of ultrasonic waves [53]. Specifically, to pelvic floor US, there are several common causes of US imaging artifacts. Fig. 1.9 shows several examples of acoustic shadowing and multiple reflections. Fig. 1.9 A) and B) both show shadowing caused by stool or air in the anorectum, C) shows shadowing caused by pelvic floor contraction or air in the glove around the transducer, D) shows shadowing due to the paramedian view of the symphysis pubis which casts a shadow behind the bone, E) shows a transobturator tape attachment for urinary stress incontinence and F) shows multiple reflections at the boundary of the symphysis pubis. The challenges and limitations of US listed can affect the ease of US imaging analysis and increase the skill and expertise required to navigate the US volume and make appropriate acquisitions and an accurate diagnosis.

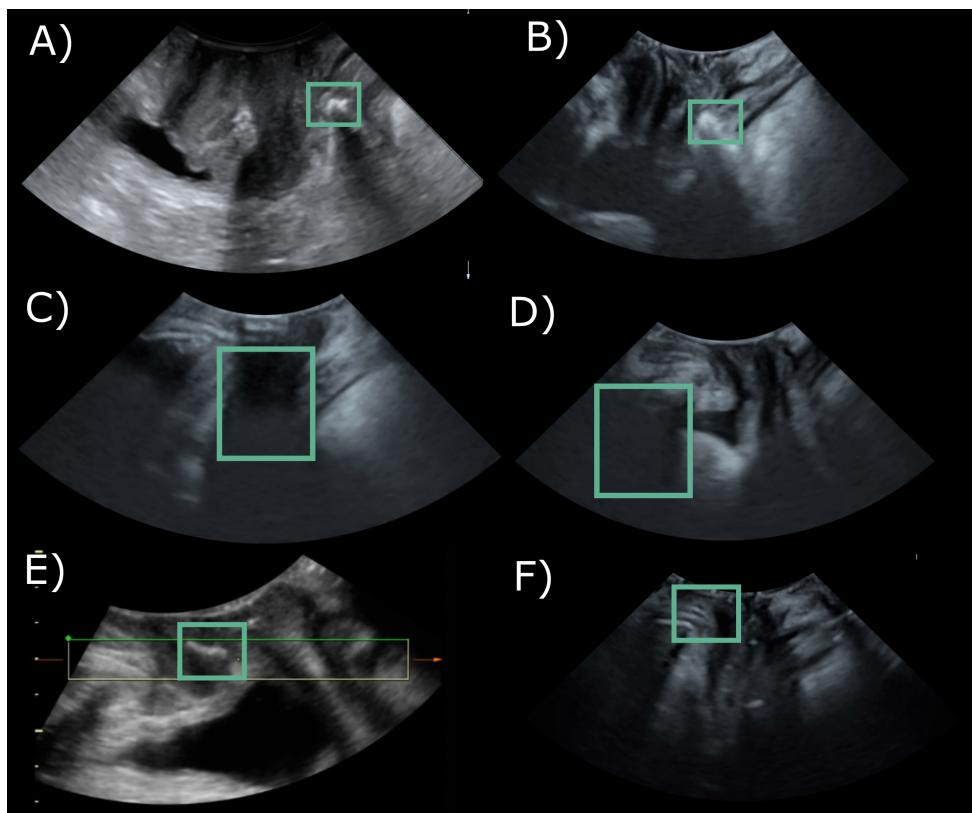


Fig. 1.9 Several examples of acoustic shadowing and artifacts in TPUS images in the mid-sagittal plane. A) and B) highlight stool or air present, C) acoustic shadowing caused by pelvic floor contraction or air in the glove around the transducer, D) acoustic shadowing due to the paramedian view of the symphysis pubis, E) transobturator tape attachment for urinary stress incontinence and F) multiple reflections at the boundary of the symphysis pubis. Image: UZ Leuven, Belgium, with permission of patients, Software: 4D View (GE Healthcare; Zipf, Austria).

Properties of a good quality ultrasound image

A high-quality US image should contain information with a high spatial resolution (i.e., can distinguish between anatomical objects in space), a high contrast resolution (i.e., can distinguish between structures that reflect US differently, for example, due to a differing scattering cross section), high temporal resolution (i.e., ability to separate between events in time), and a large field-of-view (i.e., ability to capture more anatomical structures within the US frame). In addition, an US image should be free from avoidable artifacts (i.e., shadowing, reflections from stool and/or urine in the bladder). As described, the noisy pattern in US images is speckle. However, it is not defined as pure noise, as it contains useful visual information that allows a clinician to distinguish between different tissues and anatomical properties [54].

1.2.2 TPUS imaging of the pelvic floor

To have a full scope of diagnostic capabilities for pelvic floor imaging, 3D and/or 4D imaging is necessary to assess the anal sphincter and levator complex and genital or levator hiatus. All 4D capable US machines with abdominal 4D transducers in an obstetric imaging unit should be suitable for pelvic floor imaging, provided the aperture and acquisition angle is at least 70° [55]. This may explain the popularity of TPUS imaging in clinics and academic literature, as technically all volumetric trans-abdominal probes developed for obstetric imaging can be used for three and four-dimensional TPUS imaging [55].

In this thesis, a Voluson-type system (i.e., the European market leaders in the field of 3D/4D US in gynaecology and obstetrics) is used to obtain US volumes of the pelvic floor. Off-line analysis of TPUS images is possible on the actual US system or on a computer with US software. In this thesis, 4D View (GE Healthcare; Zipf, Austria) is used to perform US imaging analysis tasks. The US software must allow the clinician to perform post-processing techniques on the acquired US volume, such as: zooming, rotating, translating, contrasting, sharpening and smoothing.

The standard representation of female pelvic floor structures acquired with TPUS is shown in Fig. 1.10. The upper left image shows the midsagittal plane, upper right the axial plane, and the lower left the coronal plane. 4D TPUS imaging involves the real-time acquisition of volume US data that can be instantly visualised in orthogonal planes or rendered volumes. This allows the assessment of functional anatomy, as 3D data can be archived as a cine loop (4D volumetric clip), which includes manoeuvres such as a pelvic floor contraction and the Valsalva manoeuvre.

Acquiring a TPUS volume

Pelvic floor imaging is performed with the patient at rest, during maximal Valsalva manoeuvre (i.e., when the patient is asked to first inhale and then exhale with a closed mouth to increase

1.2 Ultrasound imaging assessment of the pelvic floor

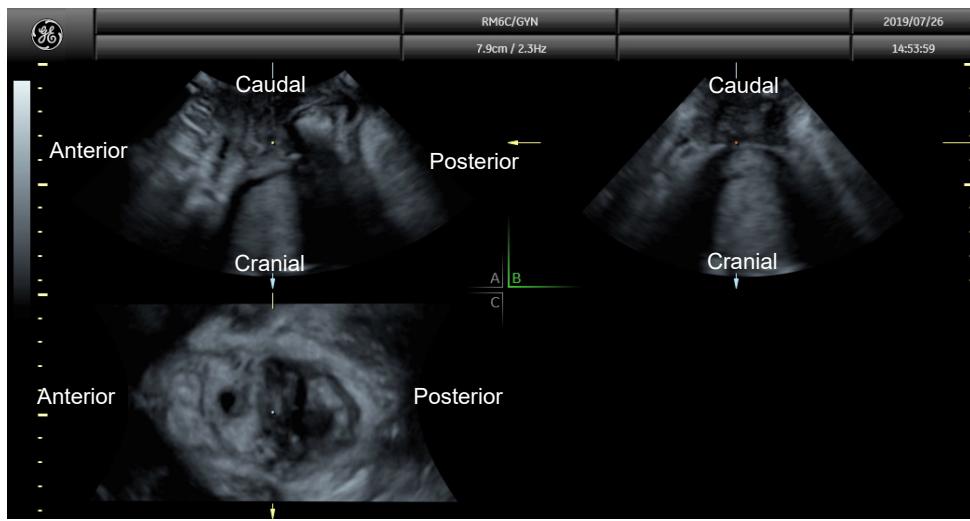


Fig. 1.10 Multi-planar view of the pelvic floor on 4D View software (GE Healthcare; Zipf, Austria). Showing the mid-sagittal (plane A and plane of acquisition), coronal (plane B) and axial (plane C) views. The caudal/cranial and posterior/anterior directions are labelled. Image: UZ Leuven, Belgium, with permission of patient, Software: 4D View (GE Healthcare; Zipf, Austria).

internal abdominal pressure) and during pelvic floor contraction (i.e., when the patient is asked to squeeze the pelvic floor muscles). This allows for dynamic assessment of the position and anatomical relationship between the pelvic floor organs [55].

Dietz *et al.* [55] describes that to acquire a TPUS, the patient should first be asked to empty their bladder and/or bowel, as this can prevent the full development of POP and reduce the image quality, as well as it often inhibits the patient because of fear for incontinence. The patient is placed in the dorsal lithotomy position with hips flexed and abducted, the probe is covered with a powder-free glove, and the transducer surface with US gel to increase the transmission rate of US waves from the probe into the body [55]. The US image is acquired by placing the probe on the perineum, as shown in Fig. 1.11 [55]. This can simultaneously image the urethra and the anal canal [55]. For 3D/4D TPUS, there is a small region of interest (i.e., acquisition angle), so it may not be possible to image all pelvic floor organs in patients with high-grade prolapse and in patients with high BMI [55]. 3D and 4D imaging gives access to the axial plane, allowing imaging of the caudal part of the levator ani muscle and, most importantly, the levator hiatus. In addition, 3D/4D imaging allows the clinician to access the coronal plane, which has attracted interest for its excellent ability of imaging the anal sphincter complex, as shown in Fig. 1.12 [55].

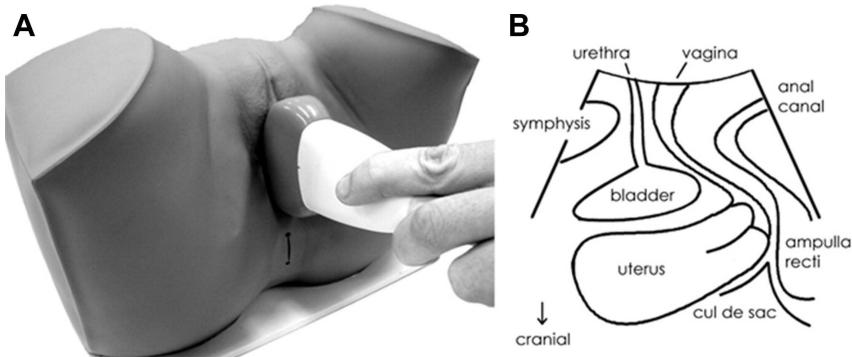


Fig. 1.11 Reprinted with the permission from Elsevier [5]. A) Transducer placement on the perineum, parallel to the labia. B) Schematic diagram of the imaged pelvic floor structures in the mid-sagittal / acquisition plane.

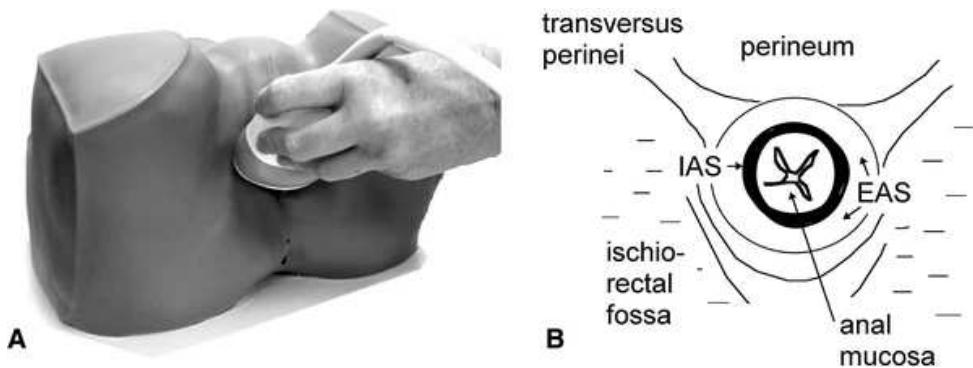


Fig. 1.12 Reprinted with permission from John Wiley & Sons [56]. A) Transducer placement for exo-anal sphincter imaging. The transducer is placed on the perineum, perpendicular to the labia. B) Schematic diagram of the imaged structures in the coronal / transverse plane. IAS = internal anal sphincter and EAS = external anal sphincter.

1.2.3 Imaging the levator hiatus

To image the levator hiatus, the US transducer is placed on the perineum, and the main axis of the transducer is oriented in the mid-sagittal plane. Within the clinic, the protocol defined by Dietz *et al.* is used to image the levator hiatus [57]. The muscle wall is easily defined relative to the surrounding tissues, due to the high echogenicity of muscle fibres through the axial plane, (i.e., the plane perpendicular to the US wave). The area of the levator hiatus is at a depth of 2-4cm from the perineum and can be well visualised with an US transducer operating between 7 and 4 MHz [57].

Fig. 1.13 shows the location of the plane used to determine the hialtal area. The plane is referred to as the plane of minimal hialtal dimensions, and it is identified in the mid-sagittal plane. It is the plane of minimal distance between the hyperechogenic posterior aspect of the symphysis pubis and the hyperechogenic posterior aspect of the anorectal angle, where the puborectalis muscle encircles it, which is part of the levator ani muscle [57]. Both structures

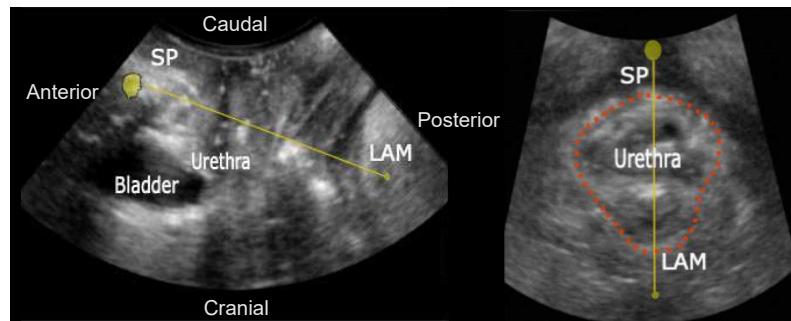


Fig. 1.13 The plane of minimal hiatal dimensions is represented by the yellow line in both views. (Left) Shows the plane position within the mid-sagittal plane, defined as the minimal distance between the symphysis pubis (SP) and the levator ani muscle (LAM) and, (right) shows the cross-section of the US in the axial plane. This plane is used to measure the hiatal area, which is displayed by the dashed red line. The caudal/cranial and posterior/anterior directions are labelled on the mid-sagittal plane. Image: UZ Leuven, Belgium, with permission of patient, Software: 4D View (GE Healthcare; Zipf, Austria).

are highly echogenic and serve as landmarks for the US image. Once the plane is defined in the mid-sagittal plane, the cross-section of the volume is displayed on the axial or ‘C’ plane [57]. To assess levator avulsion a clinician can extract a TUI sequence of the levator hiatus, following the protocol outlined in Section 1.2.5.

1.2.4 Imaging the anal sphincter

To image the anal sphincter complex, the transducer is placed at the perineum in the coronal plane [56]. Once the US volume is acquired, the clinician can extract a TUI sequence of the anal sphincter complex, following the protocol outlined in Section 1.2.5. A tomographic representation of the US volume can effectively image the entire length of the external anal sphincter which permits diagnosing tearing. This technique is repeatable and can be performed on any US system that allows multi-slice volume imaging.

US imaging of the anal sphincter can be imaged at rest, and on anal sphincter contraction. The contraction provides a functional assessment of the anal sphincter. The benefits of exo-anal (i.e., outside of the anal canal) US imaging are that it does not distort the anal canal, allows for a dynamic evaluation of the anal sphincter, and is more comfortable for the patient [56]. Generally, the anal sphincter complex is imaged as a 3D US volume at rest, and as a 4D US volume to dynamically assess the anal sphincter contraction.

Within the US images, the components of the anal sphincter complex can be seen. The mucosa of the anal canal is a star-shaped hyper-echoic area, the internal anal sphincter is a hypo-echoic ring, and the external anal sphincter is an isoechoic to hyper-echoic ring surrounding the internal anal sphincter [56]. Although TPUS imaging (i.e., exo-anal imaging) performs well and is a desirable modality (due to availability within clinic), endo-anal US is currently

Chapter 1. Introduction

considered the standard for assessment of anal sphincter injury [58]. However, endo-anal US requires specific US equipment and expertise [58].

1.2.5 TUI sequence of the pelvic floor

TUI allows the user to obtain a volume dataset that simultaneously displays multiple images at specific distances from a selected view plane (i.e., axial, sagittal or coronal) [59].

TUI sequence of the levator hiatus

To evaluate the entire levator ani muscle and its attachment to the pubic rami, 3D/4D TPUS imaging can be used. The protocol defined by Dietz [60] requires the clinician to obtain a TUI sequence (set of eight slices) in the axial or C-plane. These slices are positioned at intervals of 2.5mm, from 5mm caudal to 12.5mm cranial of the plane of minimal hiatal dimensions [60]. This US imaging analysis task can diagnose levator avulsion. Fig. 1.14 shows a TUI sequence obtained on a patient with right uni-lateral levator avulsion.

TUI sequence of the anal sphincter

A TUI sequence can assess the integrity and health of the entire anal sphincter complex, as described in [62, 56]. A set of slices at pre-determined locations is based on the cranial termination of the external anal sphincter and the caudal termination of the internal anal sphincter,

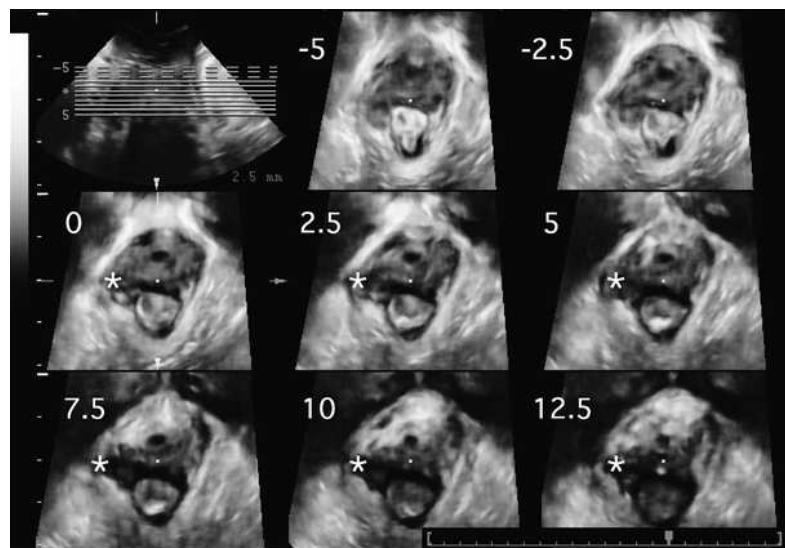


Fig. 1.14 Reprinted with permission of Springer [61]. Complete right-side avulsion (shown by an asterisk) affecting all TUI slices. Slices are marked from 5mm (-5) below to 12.5mm above (12.5) the plane of minimal hiatal dimensions (0) [61]. Top left image shows the corresponding positions of the TUI slices in the mid-sagittal plane.

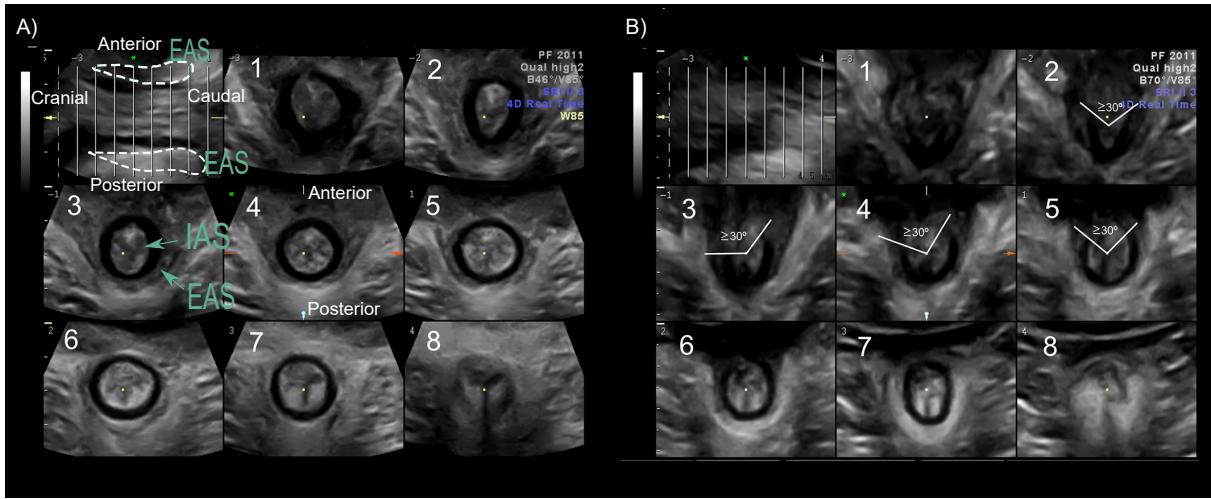


Fig. 1.15 Reprinted and edited with permission from John Wiley & Sons [63]. TUI sequence of the anal sphincter. In both examples the top left image is the mid-sagittal plane and it shows the placement of 8 transverse slices, which cover the entire external anal sphincter (EAS), slices 2 to 7. Slices 1 to 8 show the transverse coronal views used to assess the integrity of the EAS and internal anal sphincter (IAS). A) Shows a healthy anal sphincter with no defects, and B) shows a defected anal sphincter with 4 slices showing a defect greater than 30 degrees circumference. The caudal, cranial, posterior and anterior directions are indicated on the mid-sagittal view of the anal sphincter complex. The anterior and posterior directions of the transverse coronal view are indicated on slice 4. Slice 1 shows the most cranial slice and slice 8 shows the most caudal slice.

as shown in the visual example of the TUI sequence in Fig. 1.15. The integrity of the anal sphincter complex can be manually assessed by measuring the angle of tear in each slice (if present) in the external and/or internal anal sphincter. The tear measurements are used to grade the type of anal sphincter tear, according to the protocol outlined in [62, 56].

In endo-anal imaging an anal sphincter defect has been defined as a defect of 30 degrees or greater in at least 2 of 3 slices [56]. Regarding TPUS imaging, six out of the eight TUI slices are routinely assessed [56]. Therefore, the 2/3 rule of endo-anal imaging is translated into a 4/6 rule for TPUS [64]. Fig. 1.15A shows a healthy anal sphincter TUI sequence, and Fig. 1.15B shows a pathological anal sphincter, where slices 2, 3, 4 and 5 have tears larger than 30 degrees.

1.3 Deep learning techniques

Artificial Intelligence (AI) has shown promising advances in automated US imaging analysis. In particular, it has performed with impressive accuracy and sensitivity in identifying imaging abnormalities [65], tissue-based detection and classification [66], organ segmentation [7] and standard plane detection [67].

AI is a technology that involves computerised algorithms to dissect complicated data, and it has the ability to imitate intelligent human behaviour [9]. ‘Machine learning’ is a term used to

Chapter 1. Introduction

describe a subclass of AI where the algorithm's performance improves as it is exposed to more data. 'Deep learning' is a subset of machine learning, in which artificial neural networks learn from a large amount of data over time. A review article by Liu *et al.* [68] reported over 100 deep learning contributions to US imaging analysis, with a focus on classification, detection and organ segmentation. Deep learning techniques are utilised in this thesis due to their 'state-of-the-art' performance [68, 6].

Deep learning algorithms refer to artificial neural networks with three or more layers. Deep learning can be divided into two categories: unsupervised and supervised learning. *Unsupervised* learning is when the neural network tries to find patterns from the input US image during network training, without guidance. Whereas, *supervised* learning methods require an input of US images and their corresponding expert-annotated label, or ground-truth pairs, during training. The parameters of the neural network are then optimised for the specific label supplied to the neural network during training. Therefore, supervised learning requires a large training dataset of good quality ground truth labels (e.g., pixel-wise annotations for segmentation). This can be difficult and expensive for medical imaging, as a high level of clinical expertise and large volumes of data are required. Supervised learning methods are the current 'state-of-the-art' method for US imaging analysis in terms of performance and accuracy in classification and segmentation tasks. Therefore, they are utilised in this thesis, and will be explored further in the literature review, Chapter 2.

To develop accurate AI models to be used to automate the clinical workflow, it is important the training data meets several requirements. Firstly, the data must represent the clinical reality, meaning the acquisition quality, case mix of pathologies and patient demographics are balanced and a true representation of what is observed in the clinic (real world data). The case mix could be identified by a review of the literature to estimate the prevalence of pelvic floor disorders [69] (ideally based on data from multiple centres).

A true clinical representation is important because the AI algorithm will learn visual features of the training data. Therefore, the algorithm needs to be exposed to varied and diverse data [70]. Biased and unbalanced training data is a valid concern in healthcare [71]. For example, an artificial neural network for skin lesion classification was trained on a majority of white patient data, with Black patients accounting for only 5% to 10% of the training data. This resulted in the algorithm achieving half the classification accuracy on Black patients compared to the accuracy initially claimed [72].

Secondly, the ground truth labels need to be of high quality and curated by experts within the specific field of medicine. This is because the algorithm will learn to perform the task (e.g., outlining of anatomical structures) similar to the training data supplied. The ground truth labels also need to be free from diagnostic or personal bias. One study [73] demonstrated that an artificial neural network trained on biased data to classify chest X-ray pathology, was under-diagnosing historically under-served populations (i.e., females, Black patients, Hispanic

patients, and patients of low socio-economic status). To mitigate this, one must ensure that training datasets are diverse and free from bias, and that institutes address the low diversity in health data science [74].

Finally, the training dataset size needs to be of sufficient quantity. The optimal training dataset size is task dependent and could be identified by measuring the performance against training dataset size. In general, the performance will continue to increase with an increasing training dataset size [75], however, the rate of increase may start to decrease when sufficient data is provided [76]. It is important to note, that the overall performance is mainly dependent on how well a training dataset represents the clinical environment rather than the dataset size [77]. In addition, to understand the optimal dataset size and quality, one can refer to other studies that have automated similar tasks with a high-level of accuracy in the literature. Unfortunately, prior to this work there was no literature using TPUS images except for 2D US images of the levator hiatus [78], which was in collaboration with our group. The proposed method used a small dataset size of 91 US images, with three expert labels per image (i.e., 273 labels). An alternative, is to review literature of machine learning techniques for automatic segmentation of other anatomical structures from US images [79, 80]. It can be used to determine the relationship between performance and dataset size. However, it is less optimal, due to the differences in imaging and acquisition quality, and structural and tissue variation between anatomical structures.

Before using a study as an example, it is important to base this judgement on the quality and diversity of the dataset used for evaluation (i.e., is it multi-centre data, is there a wide case mix, is there a wide patient demographic?). Due to the difficulty in curating high-quality medical datasets, access to large public datasets [81] is important and beneficial for the development and training of artificial neural networks. For some tasks large public datasets exist, which makes direct comparisons of performance, training dataset quality and quantity more reliable among institutes. However, unfortunately, no public datasets exist of TPUS images.

Within deep learning, there are several types of artificial neural networks. In computer vision and US imaging analysis, convolutional neural networks (CNNs) are the ‘gold standard’. CNNs are therefore used in this thesis and will be further explored. A CNN is a type of artificial neural network that utilises spatial information from data that has a grid pattern, such as 2D or 3D images [82]. Specifically in this thesis, the images are US images (2D or 3D) of the pelvic floor anatomy. Most CNN’s exploit the property that image features are compositional hierarchies, in which higher-level features can be obtained by combining lower-level features [83]. Some CNN’s also build hierarchy feature trees [84].

In US images, pixels are assembled into edges, edges into motifs, motifs into parts (i.e., for the pelvic floor this can be tissues, organ boundaries, vessels), parts into objects (i.e., urethra, vagina, anal sphincter, bladder), and objects into scenes (i.e., the pelvic floor system in different women, pelvic floor systems of different pathologies and different image or acquisition quality) [85].

Chapter 1. Introduction

A CNN is formed from multiple layers of mathematical operations, where information is selectively passed from one layer to the consecutive layer using a non-linear activation function. Some examples of layers are convolutional, pooling and fully connected layers, and it is the convolutional layer that gives CNN its name [85]. Convolutional layers apply a convolution operation to the input data. The convolution filter slides across the input data, performs a weighted element-wise multiplication, and sums the result into a single output pixel. This filter is applied at every location of the input data until the entire input has had a convolution operation applied. Filters are also referred to as weights, and they transform a matrix of features into a new matrix of features. It is this property that allows the computation of hierarchical features of US images (i.e., low-level features combine to form higher-level features, and the feature complexity increases in each subsequent layer). These filters/weights are first randomly initialised and in the process of *training*, they are optimised for the specific task. Therefore, specific features are learned during *training*.

During CNN training for segmentation, the CNN is shown an US image, and it produces an output as a pixel-wise probability map. For a single-class segmentation, the pixels of the probability map range between 0 and 1 (i.e., a value of 1 being where it is certain the target object is, 0 where it is certain the background pixel, and 0.5 being where it is the most uncertain on whether the pixel is foreground or background). A comparison is made between the output probability map and the ground truth label, and the error or distance between the two is computed (referred to as ‘loss’). The algorithm modifies the feature maps of the layers, and this process is repeated many times on the ‘training’ dataset, until the ‘training’ error tends to zero. CNN training is demonstrated in Fig. 1.16 as a multi-class segmentation problem of the symphysis pubis, levator ani muscle and urethra. The CNN is validated during training on a validation dataset that shows the CNN’s performance on data on which it is not specifically trained. It is common practice to use 15-20% of the total dataset size (training, validation and testing combined) for validation [86].

Once the CNN has been trained and the ideal validation model iteration has been found, the performance of the CNN can be measured on a dataset of ‘unseen’ US images (referred to as the ‘testing’ dataset). This is an accurate reflection of the CNN’s performance on new US images, and measures the accuracy, performance and generalisability of the CNN model on ‘unseen’ data. The testing dataset is the dataset used to calculate computer-observer metrics, to compare to the clinical inter-observer and intra-observer variability. A ‘high-quality’ evaluation dataset should ideally contain data from multiple centres, a variety of US imaging machines, acquisition and imaging qualities and patient demographics. CNN’s have different architectures (i.e., the layers are arranged differently, and some layers differ between CNNs). Different architectures have shown various performance regarding computational cost (i.e., ‘training’ and ‘testing’ time) and segmentation and/or classification accuracy.

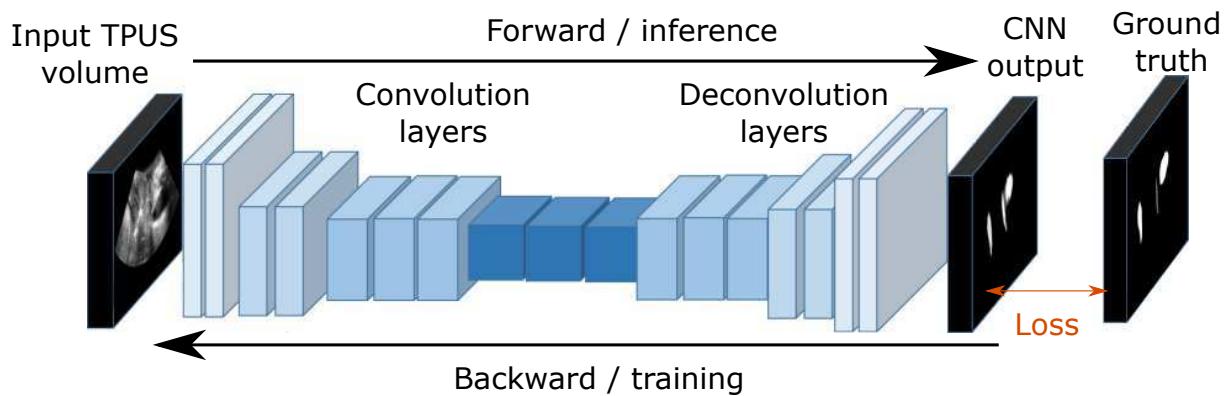


Fig. 1.16 A generic CNN architecture (composed of convolution and deconvolution layers) trained to segment the levator ani muscle, urethra and symphysis pubis, as a multi-class segmentation task. The arrows show the forward pass of an input TPUS image through the CNN, and the backward pass of the loss between the CNN output and ground truth segmentation. The backward pass is otherwise known as ‘learning’. This process is repeated (on the training dataset) and the loss will tend towards zero as the features and weights adapt to minimise the difference between the CNN output and ground truth segmentation.

The following text, describes several key categories of artificial neural network architectures. In 1998, Yan Le Cun [87] developed the first artificial neural network similar to the CNN’s used in US imaging analysis today. LeNet-5 recognised handwritten characters by aggregating simple features into more complex ones. In 2012, Krizhevsky proposed AlexNet [88], which increased the popularity of CNNs in US imaging analysis. It achieved ‘state-of-the-art’ classification performance on a large visual recognition challenge called ImageNet. Several features are commonly used today: a rectified linear unit (ReLU) activation, data augmentation, increased network depth, and overlapped pooling. Since AlexNet, several CNN architectures have been developed that have achieved human-level performance. In 2015, GoogLeNet was proposed and used multiple sized feature filters to overcome CNN over-fitting of training data [89]. ResNet was proposed in 2015 [90] and used ‘residual mapping’ to solve the ‘degradation problem’, which refers to a point where an increase in network depth saturates CNN accuracy, followed by a rapid decline in accuracy. ResNet also contains residual connections that improve training speed, a common limitation of CNNs.

In 2015, Ronneberger *et al.* developed U-Net [91]. U-Net is referred to as an encoder-decoder neural network. The encoder/contracting path collects pixel location features and context from the US image, and the decoder/expansive path restores the spatial dimension and pixel location features, enabling precise localisation. This gives U-Net its U-shape architecture. U-Net was trained with few images (by CNN standards) and outperformed prior ‘state-of-the-art’ methods on the ISBI challenge for segmentation of neuronal structures from electron microscopic stacks. Since the publication, U-Net has been cited over 39,000 times, and is commonly used in US

imaging analysis, especially in segmentation tasks. U-Net has been extended into a 3D version for 3D semantic segmentation [92].

Milletari *et al.* proposed V-Net, which is another encoder-decoder neural network, and an adaptation of U-Net [93]. V-Net was proposed for volumetric segmentation integrated with 3D convolution and residual connection blocks (as proposed in ResNet to improve training speed). Finally, HighRes3DNet was developed in 2017, it is a high-resolution compact CNN for volumetric segmentation [94]. It used dilated convolution to preserve the resolution of 3D feature maps and uses residual connections (as proposed in ResNet) to improve training speed. To note, there are many other network architectures that have been proposed in the literature, which are derivations of the architectures described [95, 96].

1.4 Motivation of this work and thesis overview

1.4.1 Motivation

Assessment of the anatomy of the pelvic floor, as well as the function in women with pelvic floor complaints is a crucial part of clinical practice. Assessment can help clinicians determine the cause of pelvic floor disorders, estimate their impact, guide the management of symptoms, defined treatment options, and in particular be useful in planning surgical interventions. Among several pelvic floor imaging modalities, US is to be preferred because of its safety, relatively low cost compared to other medical imaging modalities, real-time 3D imaging capabilities, clinical availability and portability [48]. In particular, 3D TPUS has advantages, such as the potential of imaging of 3D structural geometry and the movement of pelvic floor muscles in real time while being comfortable to the patient. Another advantage is that it gives clinicians access to planes within the pelvic floor that are not accessible via 2D US probes. In addition, the required probe for TPUS is commonly available, and can be used for several pelvic floor analysis tasks [55].

Studies suggest that TPUS imaging of the levator hiatus is a reliable technique that can be learned in a short time [97]. However, standardisation is still needed, as TPUS imaging and the analysis of the acquired US volumes (for the calculation of the hiatal area, avulsion detection and anal sphincter tear observation) can still be challenging, time-consuming and prone to human error [98]. Therefore, TPUS imaging may not always be used efficiently and accurately in many clinics. A deep learning technique known as CNNs has shown that it performs comparable to experts for several automatic US imaging analysis tasks, while reducing user-time [68]. This has motivated the use of CNNs in this thesis to automate several manual US imaging analysis tasks for pelvic floor disorders.

At the beginning of the work outlined in this thesis, there were limited solutions for the analysis of pelvic floor US images [99], including some developed by our team [100, 78]. However, the research area has expanded over the project period [101–104, 99, 105], thereby

highlighting the timeliness of this work. The aim of this thesis was to develop several automated pipelines based on those carried out within the clinic. The work aimed to closely follow the clinical workflow to increase clinical acceptance, reduce user time, reduce the need for high-level expertise, and perform within or better than the measured inter-observer error between clinicians.

1.4.2 Thesis objectives

The aim of this thesis was to develop automatic pipelines and novel techniques for automating several TPUS imaging analysis tasks carried out within the clinic. The work aimed to closely follow the clinical workflow to increase clinical acceptance, reduce user-time, reduce the need for high-level expertise, and perform within or better than the measured inter-observer error between clinicians. Automatic US imaging analysis tasks should:

- be developed with clinical implementation and acceptance in mind. Therefore, they should closely follow the clinical workflow to ensure that the algorithm pipeline is interpretable to the clinician;
- be validated on a challenging, clinical dataset that reflects the true clinical world scenario;
- reduce user time and perform faster than manual US imaging analysis;
- not have an error greater than the measured inter-observer error, as we do not want to reduce performance with automation, but to match or improve and standardize the current performance;
- allow less experienced clinicians to perform difficult US imaging analysis tasks;
- allow easy manipulation and editing of the result, which gives clinicians liability for diagnosis (i.e., the pipeline can be extended and user interaction could be introduced to make corrections).

In particular, the tasks of US imaging analysis focused on this thesis are the following:

- automatic localisation and segmentation of the pelvic floor structures of interest;
- automatic detection and extraction of a clinically used plane, the ‘plane of minimal hiatal dimensions’, also known as the ‘C-plane’ from a TPUS volume;
- automatic segmentation of the hiatal area from a TPUS volume;
- 2D interactive segmentation pipeline to allow robust, quick editing and recovery from poor levator hiatus segmentation;

Chapter 1. Introduction

- 3D interactive segmentation pipeline to allow robust editing of the automatic segmentation of the external anal sphincter;
- automatic extraction of a TUI sequence of the anal sphincter complex (i.e., eight aligned coronal 2D US slices across the length of the anal sphincter complex) used to assess anal sphincter tearing.

1.4.3 Thesis outline and contributions

The thesis presents the methodology of several automated US imaging analysis tasks to support the assessment of pelvic floor disorder (Chapters 3-8). A summary of the chapters is provided below:

Chapter 1 presents an introduction to the anatomy of the pelvic floor and pelvic floor disorders. It then describes how US can assess the integrity and function of the pelvic floor. There is then an introduction to convolutional neural networks used for automatic US imaging analysis.

Chapter 2 presents a literature review of automated US imaging analysis tasks focusing on automatic segmentation, interactive segmentation, plane detection and pelvic floor structure (including the levator hiatus and puborectalis muscle) segmentation. Due to the limited literature in this field (at the start of this work) there are only a few examples of automatic pelvic floor US analysis.

Chapter 3 presents several automatic 3D CNN segmentation tasks developed in this thesis. The segmentation of pelvic floor structures is important for the project, as clinicians visually identify them and use visual information to manipulate a TPUS volume when performing US imaging analysis. The first section focuses on the automatic segmentation of the urethra from a TPUS volume [106]. The second part of this chapter shows multi-class segmentation, in which the urethra, levator ani muscle and symphysis pubis are automatically segmented from a TPUS volume.

Chapter 4 presents a novel method that automatically extracts the plane of minimal hiatal dimensions or C-plane from a TPUS volume. This method enables robust extraction of the C-plane by using CNNs that learn to identify the key landmarks (symphysis pubis and levator ani muscle) that define the C-plane. To our knowledge, it was the first solution in the literature for the localisation and extraction of the plane of minimal hiatal dimensions [107].

Chapter 5 presents a method for automatically extracting the hiatal area from a TPUS volume, and is evaluated on a large clinical dataset. The method has a two-step approach, where the first 3D CNN identifies the landmarks defining the C-plane. The second step is a 2D CNN that segments the hiatal area from the extracted C-plane. This work was the first solution to extract the hiatal area from a TPUS volume [108].

Chapter 6 addresses a fundamental issue of automatic segmentation - that ‘state-of-the-art’ automatic segmentation methods are not robust enough for clinical implementation. The lack of robustness is caused by discrepancies between training and testing US images (i.e., US images may differ because of acquisition quality, patient pathology, US quality, artifacts and speckle). For this reason, a novel interactive 2D segmentation pipeline is proposed to edit incorrect segmentation of the levator hiatus. The proposed method utilises CNNs and B-spline explicit active surfaces (BEAS). Therefore, the segmentation is regularised, smoothed, and can be edited in real time due to the nature of BEAS. This pipeline provides liability to the clinician, which is important when segmentation is used for disease diagnosis [109].

Chapter 7 presents a novel interactive 3D segmentation pipeline developed to edit incorrect external anal sphincter segmentations. The method extends from work presented in Chapter 6 and includes a ‘3D Slicer’ module and a novel application of a hyper-parameter training algorithm to select the optimal BEAS hyper-parameters for interaction and segmentation accuracy. In turn, increasing the applicability of this interactive segmentation pipeline.

Chapter 8 presents a novel automatic pipeline that extracts eight equally spaced coronal view slices of the anal sphincter complex used for clinical assessment of anal sphincter tearing. The method uses a CNN to segment the external anal sphincter. This segmentation allows the algorithm to determine the axis of rotation and position of the coronal slices in a similar approach to the manual technique (referred to as TUI extraction) used in the clinic [110].

Finally, **Chapter 9** discusses the contributions of this thesis. It also discusses which US imaging analysis tasks can be automated in the future to aid the assessment of pelvic floor disorders, and how one may translate this research to clinical implementation.

Chapter 2

Literature review of US imaging analysis

This chapter reviews ‘state-of-the-art’ work on automatic and semi-automatic US imaging analysis relevant for this thesis. The main US imaging analysis tasks considered in this thesis are extraction of the plane of minimal hiatal dimensions, levator hiatus segmentation, and TUI sequence extraction of the anal sphincter complex. Therefore, the main US analysis approaches utilised in this thesis are organ segmentation, landmark detection, and standard plane detection.

Due to the difficult nature of 3D manipulation of US volumes and the challenges that US poses, manual US imaging analysis is prone to inter-observer error, labour intensive, and requires a high level of expertise and training. Therefore, automated and semi-automated algorithms may improve the current clinical workflow by reducing clinical time, reducing the entry level of skill required, and reducing the inter-observer error between clinicians, hence standardising the analysis task. Segmentation and plane detection methods of US images will be discussed in more detail, as these are the main tasks utilised in this thesis.

Section 2.1 and Section 2.2 introduce automatic and interactive segmentation methods for US images respectively, Section 2.3 reviews segmentation methods of the puborectalis muscle and levator hiatus from US images, and Section 2.4 reviews methods that extract standard diagnostic planes from US volumes. The gaps in the literature that are specifically addressed in this work are pointed out in Section 2.5.

2.1 Segmentation techniques for US images

Segmentation of US images is the technique that subdivides anatomical objects in an US image [111]; this could be to identify a region of interest (i.e., a tumour or lesion), measure the volume or area of an anatomical structure and for diagnosis [111]. Many segmentation algorithms for US image segmentation have been proposed. In this section, four automated US segmentation approaches are explored: segmentation with low-level features, segmentation with

active contours, segmentation with deep learning, and segmentation with a combined approach of machine learning and active contours.

2.1.1 Segmentation with low-level feature extractors for US images

Low-level feature extraction of pixel intensity, texture, edges and shape are widely used methods for simple segmentation [112]. A low-level feature segmentation method, known as threshold segmentation, assumes the target object has a uniform brightness, and the background pixels have a different pixel intensity. The method segments the image into foreground pixels (i.e., target organ) and background pixels (i.e., rest of the US image), with no consideration of pixel location. This is not suitable for US images, where the intrinsic nature of US means that biological, inhomogeneous tissues are displayed as a speckle pattern (i.e., not with a uniform pixel intensity). Another type of low-level feature extractor segmentation is edge-based segmentation. It utilises discontinuities in pixel intensity of US images to represent boundaries between objects [111]. Unfortunately, these methods are sensitive to image noise [111]. Therefore, due to acoustic shadowing, speckle, pure noise and the low contrast of US images, it is not a reliable approach.

Another method is region growing-based segmentation. This is based on the principle of homogeneity (i.e., pixels with similar properties are grouped together) [111]. Segmentation begins at an origin, and the algorithm checks adjacent pixels of the US image against predefined homogeneity criteria. Pixels that meet the tissue criteria are included in the segmentation, and the segmentation evolves. Noise affects this method less than other low-level methods. However, it can tend to over- or under- segment the region of interest [111]. Shan *et al.* accurately segmented breast masses from US images [113] with a region-based approach. The origin was automatically defined from the breast mass region based on textural and spatial features. Poonguzhal *et al.* segmented liver cysts and kidney masses from 2D US images [114]. The boundaries of segmentation were not smooth due to speckle, and it was sensitive to acoustic shadowing and artifacts.

2.1.2 Segmentation with active contours for US images

In active contour segmentation, an initial contour evolves to the boundary of a target object on an US image [115]. The contour evolves by minimising an energy term. These energy terms can be internal terms (i.e., dependent on the curve itself) or external terms (i.e., dependent on the US image information and context) [116, 117].

The first active contour segmentation method proposed was the snake model. Snake points explicitly express snake contours and they move based on the value of the energy field (i.e., energy between inside and outside the contour) [116]. However, it suffered from many limitations, such as topological changes (i.e., where the curve merges with another curve or splits into two or

more segments), self-intersection (i.e., contour overlapping can occur), sensitivity to accurate contour initialisation, and sensitivity to noise and speckle [117].

To address limitations of the snake model, level-set segmentation was proposed. This implicitly provides a segmentation by propagating an evolving interface. The evolution is derived from a variational formulation, and the segmentation evolves by minimising an energy function that reflects the properties of the target object [117, 118]. There are several advantages, such as the propagation of the contour can handle topological changes, the contour can be initialised automatically or semi-automatically, it is less sensitive to noise and it can be extended to 3D [117].

Energy functions in level-set based methods can be divided into two categories: edge-based or region-based terms. Edge-based uses image gradients to identify boundaries, so they are sensitive to noise. They cannot detect edges without a gradient (i.e., smooth or discontinuous boundaries) [119]. Region-based methods are driven by the properties of regions within and outside the evolving surface and perform better on weak or discontinuous boundaries. However, they can fail if the properties are calculated globally (i.e., using the whole inner or outer region of the curve).

To address this limitation, Lankton and Tannenbaum *et al.* proposed a localised region-based approach [120], which detected objects with heterogeneous properties [120]. It performed better on non-homogeneous objects, which is important for US images [118]. Li *et al.* [121] also proposed a solution for the limitation associated with global region-based methods, by proposing a region-scalable fitting active contour model. This dealt with intensity inhomogeneity by using the intensity information in spatially varying local regions. However, it is sensitive to contour initialisation [121]. Since Li *et al.* [121], several papers have adapted region-scalable fitting active contour models. Wang *et al.* proposed the local and global intensity fitting model, which is robust to contour initialisation and noise [122]. However, it required individual tuning of a hyper-parameter. Zong *et al.* [123] proposed a two-stage automatic segmentation scheme based on the local entropy of an US image, and a global and local level-set region-based active contour model, segmented nine anatomically different US images. However, the local entropy feature that defines the initial contour may not be suitable for complex US images.

Level-set active contour models have achieved promising segmentation performance of US images, despite the challenging nature. Therefore, they have become a popular approach when a large dataset may not be available for ‘state-of-the-art’ methods (i.e., CNNs). Level-set methods evolve an interface that can be represented in the Eulerian framework, so it can deal with topology changes and multiple unconnected segmentation regions. This is achieved by adding a dimension to the problem, which increases the computational load. Unfortunately, this makes real-time 3D applications difficult [118].

For real-time 3D applications, Duan *et al.* proposed an active contour method that represented the segmentation interface as an explicit representation [124]. A disadvantage of this, was that

the method was no longer topologically flexible. However, this is not a noticeable disadvantage in many US applications, where only one connected object is being segmented [118, 124]. Barbosa *et al.* expanded [124] into a B-spline formulation, known as B-spline explicit active surfaces (BEAS) [118]. The benefits of BEAS are that it can use a local or global region-based energy function, typically designed for level-set frameworks. It also gives the user explicit control of the smoothness of the interface through a hyper-parameter (i.e., the scale factor of the underlying B-spline formulation). Furthermore, it provides real-time performance for 3D data, as it allows the derivative of the contour’s energy to be calculated analytically, rather than numerically. Therefore, active contour evolution can be expressed as a succession of simple 1D convolutions [118].

Several works have used the BEAS framework in US image segmentation. Barbosa *et al.* for automatic real-time 3D segmentation of cardiac US images [125]. Torres *et al.* for 3D segmentation of the kidney from renal US images [126], and Pedrosa *et al.* utilised BEAS and statistical shape models as a shape prior, for accurate 3D segmentation of the left ventricle in cardiac US images [127]. Finally, related to this thesis, our group utilised BEAS with template matching to segment the levator hiatus from a pre-defined 2D US image [100]. In summary, active contour models have achieved promising segmentation results in US images. They have shown that US images can be segmented with very limited or no annotated training data. However, they rely on good contour initialisation and hyper-parameter estimations. Therefore, it may not be suitable and robust enough for anatomical structures that vary in shape and texture due to pathology.

2.1.3 Segmentation with supervised machine learning for US images

Semantic segmentation is a task that gives each pixel or voxel in one image a class label. The class labels are classified as background or foreground (i.e., the target structure in single-class segmentation or target structures in multi-class segmentation). Supervised machine learning-based segmentation methods learn image patterns of the target object or objects (referred to as features) from a set of US images and corresponding pixel-wise annotations, or otherwise known as ground-truth labels [128].

CNNs have achieved ‘state-of-the-art’ performance for automatic US image segmentation. Several works have been influential in automatic US segmentation. Those in particular are the encoder-decoder neural networks: Ronneberger *et al.*, (2D U-Net) [91], Cicek *et al.* (3D U-Net) [92] and Milletari *et al.* (V-Net) [93].

2D U-Net [91] has been extended and utilised in various 2D US segmentation tasks, such as breast lesions [129, 130], atrial walls [131], left ventricle [80], kidney [132], fetus’ [79] and tongue contour tracking [133], among other tasks [68].

3D U-Net [92] has been extended, utilised and optimised in various 3D US volumetric segmentation tasks, such as breast tumours [134], cardiac intervention instrument detection [135], neonatal cerebral lateral ventricles [136], catheter localisation for cardiac US images [137], ovary and ovary follicles [138], as well as other tasks [68].

Several works have utilised, extended and optimised V-Net for their specific US segmentation task, such as prostate segmentation [139, 140], fetal head segmentation [141] and kidney and renal tumour segmentation [142].

2.1.4 Segmentation with machine learning and active contour models for US images

As discussed earlier in Section 2.1.2, active contours can accurately segment US images, but they rely on good contour initialisation and hyper-parameter selection. Later in Section 2.1.3 key aspects described how CNNs have achieved high performance in a wide range of US segmentation tasks, as they can estimate complex shapes and appearances by learning annotated training datasets. However, these datasets must be large and contain high-quality annotations.

Acquiring these large datasets of the pelvic floor is a time-consuming and highly skilled task that requires expert knowledge of pelvic floor anatomy and US images. Sub-optimal training data could reduce the CNN's performance. In addition, CNN segmentation is a voxel or pixel-wise process. Therefore, segmentations can fail in terms of anatomical plausibility (i.e., the segmentation may contain holes and be incomplete, and contain noisy boundaries). These limitations could be addressed by integrating the CNN approach into an active contour model [115].

Several works have combined active contour models and machine learning methods for medical image segmentation [115, 143]. Sivanandan *et al.* utilised a CNN as a feature extractor for a clustering and active contour-based segmentation method [144]. The CNN learned features that distinguish between benign and malignant breast tumours from breast US images. The learned features are represented as activation maps, and by using fuzzy c-means clustering, an accurate initial contour was formed for active contour segmentation. The active contour segmentation was based on localised energies implemented by Lankton *et al.* [120], which emphasised small pixel-wise variations to detect challenging echogenicities of the breast tumour [144]. Rong *et al.* proposed using a popular active contour external force, called gradient vector flow, as a reference to train a CNN [145]. The network was trained to derive an external force and its corresponding weights for an active contour segmentation method. The active contour segmentation method was evaluated on fetal head segmentation from US images. This method ensured the initial curve was close to the target boundary and could be used for other segmentation tasks.

Several studies have used machine learning and active contour segmentation methods within other medical imaging modalities [146–148]. The combination of the two segmentation methods has several advantages, such as: machine learning can provide accurate and automatic initial segmentation for an active contour model, active contour models can overcome machine learning limitations by requiring a large training dataset, while employing a geometric structure that is more biologically plausible than a machine learning pixel or voxel-wise segmentation [115].

2.2 Interactive segmentation methods in US images

Although automatic segmentation methods have achieved ‘state-of-the-art’ performance in several US segmentation tasks, there is still limited applicability in the clinical setting. The lack of automation is seen in the pelvic floor dysfunction clinic, as automated segmentation of the levator hiatus is not used despite the advances in research literature [78]. This is because automatic segmentation methods have not reached the ideal level of robustness for clinical implementation and acceptance. There are various reasons for this, such as: the clinician remains responsible for the diagnosis and output of an automatic method, automatic methods are sensitive to domain shifts (i.e., varying acquisition qualities, image qualities, patient pathologies and artifacts) between training and testing data, it is not easy to recover from automatic segmentation failures, as there is no mechanism for smart correction of poor segmentation, it is unlikely to achieve accurate segmentation of structures and pathologies not seen during training, and lack of user interaction can cause frustration to the user, which may lead to a decline in clinical confidence of such automatic tools [149].

Interactive segmentation solves the problems associated with automatic segmentation by using user interactions to improve automatic segmentation results. Several 2D and 3D interactive segmentation approaches used user interaction in the form of bounding boxes, scribbles, or user points. Ardon *et al.* proposed an interactive method based on a support vector machine learning-based algorithm, followed by a model-based deformation technique [150]. This method was evaluated on kidney segmentation from 3D US images. The deformation framework enabled users to interact by indicating points inside or outside the kidney. These points were introduced in the optimisation process as constraints on the sign of the deformed function. This method achieved real-time response interactions, which is beneficial for clinical confidence and usability. Barbosa *et al.* also achieved real time 3D interactive segmentation of US images [151]. In this work, user-defined points were used to steer the active contour. This was achieved by minimising an interactive energy term of the active contour. The interactive energy term was based on the parametric distance between the closest point of the active contour segmentation and the ‘user-defined’ point in non-Cartesian space [151].

2.2 Interactive segmentation methods in US images

Interactive segmentation methods have also used deep learning methods, which have been shown to achieve robust and accurate segmentations by relying on complex feature extraction with CNN training. User interactions have been supplied in several ways, such as bounding boxes [152], scribbles [153], extreme points [154] and clicks [155]. Chen *et al.* [156] utilised a trainable framework and an extreme points method, which extended the interactive segmentation method called Deep Extreme Cut (DEXTER) [154]. DEXTER uses extreme points directly as guiding signals during training by introducing an additional channel with the image as input to CNN. The new channel contains a Gaussian, which is centred on each of the extreme points. The CNN learns to transform this information into an ‘edited’ and, in theory, ‘improved’ segmentation that includes those extreme points. Chen *et al.* proposed a method that required four user clicks for accurate interactive segmentation of breast lesion in US images [156].

There are limited examples of other interactive deep learning-based methods for other US segmentation tasks. This may be due to the challenging nature of US images, and the noisy or missing boundaries due to noise, speckle and artifacts. Most interactive deep learning-based methods are evaluated on MRI and computed tomography (CT) image modalities. To better understand the ‘state-of-the-art’ deep-learning-based interactive segmentation methods, several of these examples will be discussed.

A two-stage interactive segmentation approach was developed by Wang *et al.*, where in the first stage a CNN inferred an initial segmentation, and in the second stage the user could refine the automatic segmentation with scribbles of background or foreground pixels [157]. In other work, Wang *et al.* proposed a method that utilised a bounding box alongside user-provided scribbles to perform image-specific fine-tuning of the CNN segmentation [158]. The initial segmentation was obtained within a user-defined bounding box, followed by image-specific fine-tuning with user-scribbles. This method was more robust than ‘state-of-the-art’ CNNs for segmentation of unseen objects, and more accurate with fewer user interactions than traditional interactive segmentation methods. Luo *et al.* integrated user-clicks with the input image and the initial segmentation using exponential geodesic distance for interactive segmentation [159]. An interactive segmentation framework, called uncertainty-guided efficient interactive refinement (UGIR), was proposed by Wang *et al.* [153]. The framework used scribbles in a level-set method for CNN segmentation refinement.

These deep learning-based interactive methods may have improved the performance of the ‘state-of-the-art’ automatic CNN segmentation methods. However, they lacked real-time feedback (i.e., user-based interactions did not edit the contours in real-time), which can be frustrating and less intuitive for the user. This limitation was addressed by Asad *et al.*, where real-time scribble-based feedback was achieved by using a lightweight CNN [160]. These deep learning-based interactive segmentation methods show the value of using expert knowledge to improve automatic segmentation. The main disadvantage of these deep-learning based methods is that they lack explicit control and manipulation of the 3D surface. This has been tackled

somewhat in 2D by Mortensen *et al.* [161], but although a 3D extension exists, it is difficult to integrate into a deep learning pipeline. In addition, these methods have not been evaluated on US data, and they may struggle due to the challenging nature (i.e., speckle, noise, low contrast and artifacts).

2.3 Segmentation of pelvic floor structures from US images

2D levator hiatus segmentation The task of levator hiatus segmentation is to delineate and extract the levator hiatus from US images. Our group introduced the first 2D semi-automatic segmentation algorithm for levator hiatus segmentation, which required the user to input vertical hiatal limits [100]. The two-step segmentation pipeline first created an initial outline by fitting a pre-defined template to an intensity invariant edge map of the US image. Secondly, the initial segmentation was refined using BEAS. This method was relatively fast compared to manual outlining and improved the inter-observer agreement. The downside of this method was that it still required user input and clinical time.

Following this work, our group [78] then developed the first automatic segmentation solution for 2D levator hiatus segmentation from a 2D US image. Bonmati *et al.* adapted U-Net into a self-normalising U-Net by including a scaled exponential linear unit function. As a result, batch-dependent normalisation was not needed, which meant there was no limit to the size of the mini-batch, and the network could achieve equivalent results to U-Net with reduced memory constraints. This method was fully automatic, and it achieved comparable accuracy to [100], without requiring operator input.

Subsequently, another group presented work on the automatic segmentation of the levator hiatus from a 2D US image [104]. The method also utilised a CNN, however, the architecture and implementation details were not reported. The method automatically measured the length, width and area of the levator hiatus, and achieved comparable accuracy to Sindhwani *et al.* [100] and Bonmati *et al.* [78] (note the test dataset of [104] did differ from [100, 78]). This work further highlights and emphasises the impact CNNs can have on automating the manual imaging analysis of transperineal ultrasound images.

From a clinical perspective, the main limitation of ‘state-of-the-art’ automatic levator hiatus segmentation is that it still requires the manual identification and extraction of the C-plane from the TPUS volume, which is a time-consuming and highly skilled task. Therefore, a pipeline that segmented the levator hiatus from a TPUS volume would have a significant clinical value.

Puborectalis muscle segmentation As well as 2D segmentation van den Noort *et al.* introduced the first 3D automatic segmentation of the puborectalis muscle (i.e., part of the levator ani muscle). [99] presented a 3D active appearance model used to segment the puborectalis muscle from a TPUS volume [99]. However, the authors noted it would fail in cases where stool was

present in the TPUS volume. This highlights the need for a CNN segmentation approach for 3D pelvic floor structure segmentation, as the CNN would be exposed to varied data and learn to accurately segment structures despite stool or urine presence in the TPUS volume. In addition, a CNN approach should be less sensitive and more robust to other factors such as imaging artifacts, acoustic shadowing and patient pathologies, than an active model approach.

2.4 Plane detection techniques in US images

In the previous section, all levator hiatus segmentation approaches required manual selection and extraction of a standard plane, referred to as the plane of minimal hiatal dimensions or as the ‘C-plane’. Therefore, this thesis focussed on the development of a plane detection technique that followed the clinical workflow. Several studies have investigated and developed algorithms for 2D standard plane extraction from an US volume.

Chykeyuk *et al.* [162] used a Regression Random Forest (RRF) to extract standard view planes from cardiac US volumes. This method integrated voxel class label information into RRF training to exclude irrelevant classes from contributing. This yielded a class specific RRF. The method was validated on a dataset of 25 3D cardiac US volumes, and had an improved performance compared to the classic RRF [162, 163]. Yaqub *et al.* [164] adopted an RRF and feature asymmetry to emphasise informative voxels during network training. This method was used to localise a 2D diagnostic plane from fetal brain US images. This method was shown to be more reproducible than manual plane finding. Li *et al.* [165] developed an iterative transformation network (ITN) for standard plane detection from 3D US volumes of the fetal brain. The ITN used a CNN to learn the relationship between images of 2D-extraction planes and the necessary transformation needed to reach the desirable ‘ground truth’ plane. This method was evaluated on 72 US volumes of the fetal brain and accurately extracted two standard diagnostic planes.

These methods used clinical knowledge, but they did not follow clinical guidelines directly. Therefore, they may not be interpretable to the clinician, and one can describe them as a ‘black box’ approach. This could reduce the clinical acceptance of these methods, and reduce the likelihood of clinical implementation. Zhu *et al.* [166] attempted to address the limitation of clinical acceptance by proposing a guideline-based learning approach. This method determined 2D planes with anatomical regularities according to clinical guidelines. The anatomical guidelines improved the results and running time on various images. This method was used to extract six standard planes, commonly used to analyse cardiac function, from 3D cardiac US volumes.

In summary, inspired by the ‘state-of-the-art’ and to increase the clinical acceptance of the work in this thesis, a guideline-based approach to plane detection will be used to detect the plane of minimal hiatal dimensions.

2.5 Conclusion and addressing gaps in the literature

This chapter highlights the ‘state-of-the-art’ literature and efforts in the automated analysis of pelvic floor images at the time this work was initiated. There are many gaps in the literature, which the work presented in the remaining chapters of this thesis seek to address.

At the start of the work, there was no literature based on the 3D CNN automatic segmentation of pelvic floor structures from TPUS volumes, despite the advances in 2D CNN levator hiatus segmentation [78, 104]. In addition, there was limited literature on the automatic segmentation of pelvic floor structures in general, only two cases had been presented, 2D levator hiatus segmentation [100] and 3D puborectalis muscle segmentation using active models.

The active model proposed by van den Noort *et al.* [99] for 3D puborectalis muscle segmentation failed in TPUS volumes with stool, suggesting a great sensitivity to the quality of the TPUS image required for accurate segmentation. In addition, the active model used for levator hiatus segmentation relied on manual initialisation of the active model. As a result, Chapter 3 presents exploratory work, based on the novel application of automatic CNN segmentation of the urethra, levator ani muscle and symphysis pubis. The CNNs used in Chapter 3 are motivated by ‘state-of-the-art’ work regarding automatic semantic segmentation (i.e., CNN segmentation).

Chapter 7 then expands on Chapter 3 to present an interactive 3D segmentation framework. It aims to address a gap identified in the literature, that is, interactive segmentation methods can lack the real-time and explicit control of the 3D surface of a segmentation. In addition, interactive segmentation methods to date, are rarely extensively evaluated on US data, potentially due to the challenging nature of these images. Chapter 7 provides an intuitive solution based on deep learning and an active model called BEAS, that allows the clinician to robustly and explicitly edit the automatic segmentation surface in real-time. Chapter 7 also aims to address a gap identified in the literature of BEAS, being that BEAS hyper-parameters require manual fine-tuning, by presenting an automatic hyper-parameter optimisation algorithm.

Prior to the work initiated in this thesis, the plane of minimal hiatal dimensions had to be manually identified and extracted from a TPUS volume. Chapter 4 aims to address this gap in the literature, by providing a novel algorithm based on the clinical workflow to automatically extract this diagnostic plane from a TPUS volume. The method aims to follow the clinical workflow closely and focus on key landmarks (unlike other plane detection algorithms explored in this ‘state-of-the-art’ review) to increase the chance of clinical implementation and acceptance. As mentioned previously, the plane of minimal hiatal dimensions is clinically relevant as it is used to measure the hiatal area, which can be used to identify pelvic organ prolapse.

Within literature, the levator hiatus had been segmented accurately from a pre-defined plane of minimal hiatal dimensions [78, 100]. However, the levator hiatus not been segmented automatically from a TPUS volume. Chapter 5 extends from Chapter 4 to address this gap in the

2.5 Conclusion and addressing gaps in the literature

literature. The proposed automatic framework first localises and extracts the diagnostic plane from the TPUS volume, and then subsequently automatically segments the levator hiatus.

As discussed in the ‘state-of-the-art’ review, automatic segmentation has its limitations, due to the domain shift between training and testing data. Therefore, we focused on addressing this limitation, by proposing a 2D interactive segmentation framework for the levator hiatus. As stated previously, there was limited literature focused on interactive segmentation methods for ultrasound images, and segmentation methods did not give the user explicit control of the 2D segmentation boundary. Hence, Chapter 6 aims to address these specific limitations by providing a 2D interactive segmentation framework that is intuitive to use. The framework aims to improve clinical implementation and acceptance of automated pelvic floor disorder analysis, specifically related to the automatic outlining of the levator hiatus.

Finally, despite the clinical importance of assessing the anal sphincter complex with TPUS imaging, there is currently no literature that presents an automated or automatic solution for anal sphincter assessment. Therefore, Chapter 8 aims to address this gap in the literature by automating a common ultrasound imaging analysis task (i.e., TUI extraction), used by clinicians to visually assess anal sphincter integrity and tearing. This work proposes an automatic solution for TUI extraction from a TPUS volume based on the clinically relevant segmentation of the external anal sphincter.

Chapter 3

Three dimensional segmentation of pelvic floor structures

In this work, we aim to automate several US analysis tasks used in PFD assessment by using AI and deep learning techniques.

Clinicians rely exclusively on visual cues of anatomical information to perform these medical imaging analysis tasks. Visual detection of important pelvic floor structures and their boundaries allows clinicians to manipulate the TPUS volume as desired for the specific task. Therefore, to automate these tasks, automatic segmentation and landmark detection are required.

Several CNN segmentation models were trained throughout this work to identify important pelvic floor landmarks and their boundaries. Localisation and/or segmentation of pelvic floor structures were used for the following medical imaging analysis tasks: segmentation (Chapters 5, 6, 7 and 8), plane detection (Chapters 4, 5 and 8), and biomarker extraction (Chapter 5). The following structures were segmented during the thesis: 2D levator hiatus (Chapters 5, 6), 3D urethra (Chapter 3), 3D levator ani muscle (Chapter 3), 3D symphysis pubis (Chapter 3) and 3D external anal sphincter (Chapters 7 and 8).

The next section will briefly describe some models and data augmentation techniques developed and used during this thesis. This section is exploratory work only and the 3D CNNs and implementation methodologies will be built upon and utilised in subsequent chapters, such as Chapters 4, 5, 7 and 8.

This chapter is based on the conference paper published in *2019 IEEE International Ultrasonics Symposium (IUS)*: **Williams, H.**, Cattani, L., Li, W., Tabassian, M., Vercauteren, T., Deprest, J., D'hooge, J., 3D Convolutional Neural Network for Segmentation of the Urethra in Volumetric Ultrasound of the Pelvic Floor, 2019 IEEE International Ultrasonics Symposium (IUS), 2019, pp. 1473-1476.

3.1 Single-class urethra segmentation

The first section is based on the conference paper presented at IUS IEEE 2019 [106]. This work focuses on single-class binary segmentation of the urethra from TPUS volumes. It was the first work to focus on 3D CNN segmentation of a pelvic floor structure from a TPUS volume.

The urethra is of clinical significance because it lies along the mid-line of the plane of minimal hiatal dimensions (as described in Chapter 1, Section 1.2.3) used to assess POP. 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 TPUS volume using a CNN. A dataset with 35 volumes from 20 patients during a pelvic floor examination (i.e., imaged at Valsalva, contraction and rest) was labelled by an expert and 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 (95th percentile), which was comparable to intra-observer results.

3.1.1 Introduction

The aim of this study was to segment the urethra from a 3D TPUS volume, to generate context for a C-plane detection algorithm. To our knowledge, this was the first work to segment the urethra from a TPUS volume. Segmentation of pelvic floor structures prior to this work was mainly based on the two dimensional structure, the levator hiatus and one paper focused on the segmentation of the puborectalis muscle [99], however, the active model approach was sensitive to stool presence and hence, one can expect TPUS image quality. The levator hiatus was automatically segmented from the C-plane using a CNN by Bonmati *et al.* [78] as discussed in Chapter 2, Section 2.3. Bonmati *et al.* used an adaptation of the CNN 2D U-Net and achieved a Dice score of 0.90.

Based on this and the literature review from Section 2, a 3D CNN was used in this study for automatic urethra segmentation. Unlike the segmentation of the levator hiatus, in this work, we have a highly unbalanced task, meaning the number of urethra foreground voxels is significantly lower than the number of background voxels within the US volume. Therefore, in this work, we use a label-based approach to weight sampling during training.

Due to the relatively low image quality of TPUS volumes, we also employed pre- and post-processing techniques, such as specific data augmentation and connected component analysis, to enhance results. We used a 5-fold cross validation to train and validate the performance of the CNN. A clinical expert segmented each TPUS volume, and the computer-observer results were compared to the intra-observer results.

3.1.2 Material and methods

Data

A dataset of 35 TPUS volumes from 20 patients was collected, with an average voxel resolution of 0.3 by 0.3 by 0.3 mm. TPUS volumes were acquired during a pelvic floor examination (i.e., when the patient was imaged at full Valsalva, rest and pelvic floor 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 at Valsalva and would make training difficult with a small dataset. Thus, we start our research by focusing on ‘healthy’ urethrae. All 35 TPUS volumes were manually segmented using GE Vocal (4D View, GE Healthcare; Zipf, Austria) software by one clinical expert with over four years’ experience in pelvic floor TPUS imaging.

Network architecture

The network architecture used in this work to segment the urethra was HighRes3DNet by Li *et al.* [94] and it was implemented using NiftyNet – a Tensorflow based package for medical imaging segmentation [167].

The network architecture, as shown in Fig. 3.1, uses multiple layers of dilated convolutions and residual connections to map the input volume to a voxel-level segmentation. The network architecture allows multi-scale visual feature extraction due to the dilation factor of dilated convolutions, which is increased as layer depth increases, leading to an increase in the receptive field size. Residual blocks with identity mapping are used to group every two convolutional layers. This design improves training speed and ensures that the propagation of information during training is smooth. Thus, the network has a relatively large effective receptive field; can

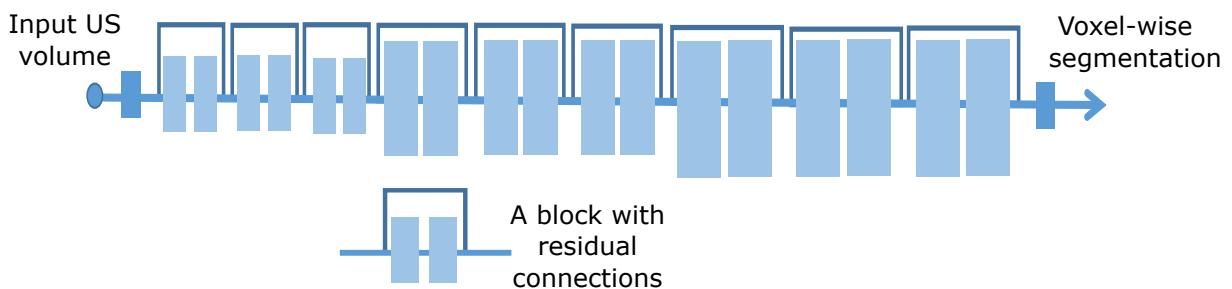


Fig. 3.1 The network architecture used for volumetric image segmentation in this paper. The network consists of dilated convolutions and residual connections. The architecture allows for multi-scale visual feature extraction as the dilation factor of dilated convolutions increases as layer depth increases. For more detail please refer to work by Li *et al.* [94].

generate high resolution feature maps; and can be trained efficiently. For more information on this architecture, please refer to the original paper [94].

Network 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 a data augmentation technique. 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 designed to facilitate patch-based medical image analysis [167]. Due to graphics processing unit (GPU) limitations, the original TPUS volume was too large to load into one sampling window without heavy down sampling. Therefore, a patch-based segmentation was used.

Due to the imbalanced nature of this segmentation task, the urethra was purposely sampled more than the background voxels during training, to ensure the network learned key features related to the urethra. This was achieved by weighted sampling, where the ground truth mask provided information for sampling. Therefore, patches supplied as input to the network during training were ‘more likely’ to be of the urethra ground truth position, helping to balance the weighting between the urethra and background voxels. Therefore, it is expected to improve network performance and efficiency [168–170].

The loss function used was the Dice similarity coefficient with the definition given by Milletari *et al.*, which had performed well for unbalanced segmentation tasks. For more information, please refer to [93]. During network training a learning rate of $5 \cdot 10^{-4}$, a batch size of 2, Rectified linear unit activation function (ReLU) and Adam optimiser were used.

Evaluation

The best performing model was found by selecting the training iteration with the highest validation Dice coefficient. Evaluation was performed in a 5-fold cross validation, in which the 5 models were obtained by training the network 5 times, with different volumes for training, validation and inference each time. After all 5 cross validation models had been trained, 35 automatic 3D urethra segmentations were obtained.

Post-processing

For each automatic segmentation obtained, segmentation post-processing morphological operators were applied. To fill holes, a flood-fill operation on background pixels enclosed by

foreground pixels was applied. In addition, unconnected collections of pixels were removed by selecting the largest component using connected component analysis, as shown in Fig. 3.2.

Evaluation metrics

Segmentation results were evaluated using a region-based metric (i.e., the Dice similarity coefficient) and a surface distance-based metric (i.e., the 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, \mathbf{A} , and the output 3D mask generated from the network, \mathbf{B} .

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

$$H(\mathbf{A}, \mathbf{B}) = \max(h(\mathbf{A}, \mathbf{B}), h(\mathbf{B}, \mathbf{A})) \quad (3.1)$$

where

$$h(\mathbf{A}, \mathbf{B}) = \max_{\forall p \in \mathbf{A}} \min_{\forall q \in \mathbf{B}} \| \mathbf{p} - \mathbf{q} \| \quad (3.2)$$

and $\| \cdot \|$ is an underlying norm on the points of \mathbf{A} and \mathbf{B} (e.g., the L_2 or Euclidean norm) [171].

Having a small Hausdorff distance indicates a good approximation as for a point in \mathbf{A} there is a point on \mathbf{B} within radius defined in (3.1) [172]. Unlike Hausdorff distance, Robust Hausdorff distance is less sensitive to outliers. ‘Robust’ means it uses the ‘ K^{th} percentile of the distances, and not the maximum distance [172]. It is defined as:

$$H_R(\mathbf{A}, \mathbf{B}) = K_{a \in \mathbf{A}}^{th} d_{\mathbf{B}}(a), \quad (3.3)$$

where $d_{\mathbf{B}}(a)$ denotes the minimum distance at position a to the position set \mathbf{B} , and $K_{a \in \mathbf{A}}^{th}$ represents the K^{th} ranked value of $d_{\mathbf{B}}(a)$ [173]. In our evaluation, we use the 95^{th} percentile.



Fig. 3.2 The post-processing effect on a network output, A) shows the initial urethra segmentation output of the network and B) shows the filtered result after connected region filtering and filling holes operations are applied.

Chapter 3. 3D segmentation of pelvic floor structures

Table 3.1 Average performance of HighResNet for urethra segmentation by employing comparison with the manual segmented labels for each TPUS volume. The standard deviation (SD) of each evaluation metric is given in square brackets.

95^{th} Robust Hausdorff distance [SD] (mm)	Hausdorff distance [SD] (mm)	Dice similarity coefficient [SD]
4.68[0.49]	7.56[1.65]	0.65[0.08]

In addition to the 95th Robust Hausdorff distance, the Dice similarity coefficient is used in evaluation. Dice expresses the overlap between the ground truth 3D binary mask, \mathbf{A} , and the network output binary mask, \mathbf{B} . This metric is on a voxel-wise basis and defined as:

$$D(\mathbf{A}, \mathbf{B}) = \frac{2|\mathbf{A} \cap \mathbf{B}|}{|\mathbf{A}| + |\mathbf{B}|}, \quad (3.4)$$

where $|\mathbf{A} \cap \mathbf{B}|$ is the overlap of pixels between the ground truth 3D mask, \mathbf{A} and network output 3D mask, \mathbf{B} and $|\mathbf{A}| + |\mathbf{B}|$ is the total sum of pixels from the ground truth 3D mask, \mathbf{A} and network output 3D mask, \mathbf{B} . Note $\frac{|\mathbf{A}|+|\mathbf{B}|}{2}$ is the mean volume of \mathbf{A} and \mathbf{B} .

Intra-observer analysis

Evaluation was recorded by measuring the difference between the computer and the 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 TPUS volumes.

3.1.3 Results

The results presented are averaged over all 5-folds trained, and can be seen in Table 3.1. Intra-observer variability using the same performance metrics is shown in Table 3.2. Fig. 3.3 shows an example of an automatic segmentation overlaid on the corresponding input TPUS volume, which was from the CNN model that achieved the smallest robust Hausdorff distance. Fig. 3.4 shows the corresponding overlap between the manually segmented urethrae and the automatically segmented urethrae for three different TPUS volumes from three different CNN models that achieved the lowest performance, average performance and highest performance based on the Robust Hausdorff distance, Hausdorff distance and Dice Similarity score.

3.1 Single-class urethra segmentation

Table 3.2 Average intra-observer variability between five manually segmented urethrae on a test-retest basis. Standard deviation (SD) of the evaluation metrics is given in square brackets.

95^{th} Robust Hausdorff distance $[SD]$ (mm)	Hausdorff distance $[SD]$ (mm)	Dice similarity coefficient $[SD]$
6.84[9.68]	9.25[10.65]	0.60[0.12]

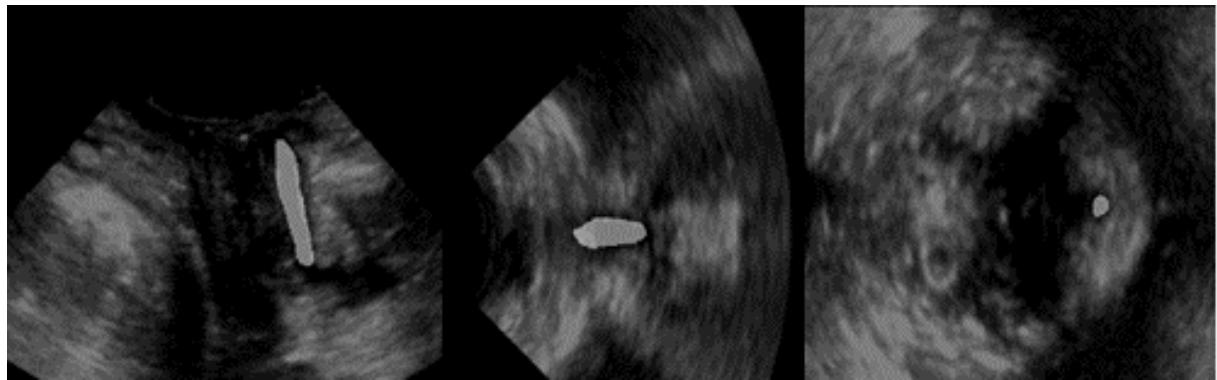


Fig. 3.3 Shows a deep learning network result (white) superimposed on top of the input TPUS volume. (Left) shows the mid-sagittal view, (middle) the coronal view and (right) the axial view.

3.1.4 Discussion

Manual segmentation of TPUS 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 TPUS 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 section, we used HighRes3DNet, a powerful network for semantic segmentation [94]. It is to our knowledge the first time this network has been used on TPUS volumetric data. The network training had to be adapted for this highly unbalanced problem (i.e., small foreground to background voxel ratio). To solve this, we sampled our network during training in the region of the ground truth foreground and used the Dice similarity coefficient as the loss function.

In this work, we compare our automatic segmentation approach against the observer's own variability, calculated by having the observer re-segment five random TPUS volumes. Comparing the computer-to-observer difference (Table 3.1) with intra-observer differences (Table 3.2), 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 CNN's standard deviation. When all re-tests from the observer were investigated, it was highlighted that this result was because of one volume having a particularly high Hausdorff

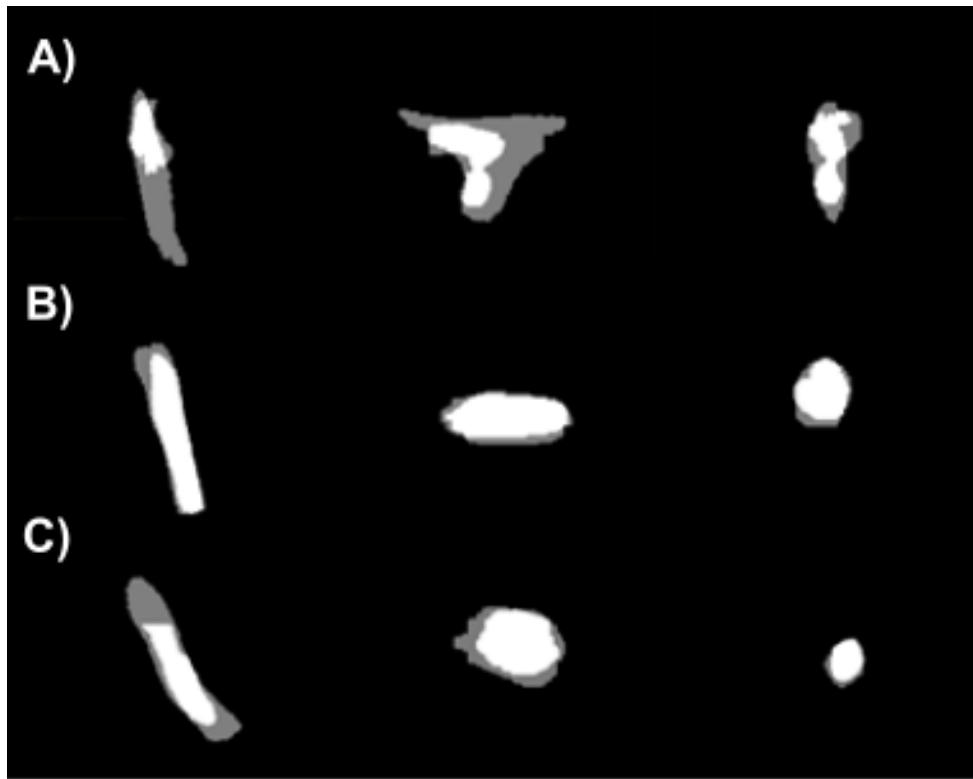


Fig. 3.4 Shows the deep learning network result (white) superimposed over the corresponding ground truth mask for three separate volumes. A) represents the lowest performing result (Hausdorff 95th percentile - 11.4mm and Dice loss - 0.40) B) represents the highest performing result (Hausdorff 95th percentile - 2.6mm and Dice loss - 0.78) C) represents an average performing result (Hausdorff 95th percentile - 5.09 mm and Dice loss - 0.62).

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 Hausdorff distance was 4.56 [1.91] mm, and the 95th percentile Robust Hausdorff distance was 2.86 [1.51] mm. The high Hausdorff distance was due to an error during manual segmentation of one volume using GE Vocal software (GE Healthcare; Zipf, Austria). 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 perform manual segmentation on a proportion of the data to generate inter-observer variability performance metrics.

If this work is to be directly used in the future, a larger training dataset should be acquired containing varied patient pathologies and imaging and acquisition qualities, to improve the segmentation performance. Further investigation of network architecture and implementation methods should also be explored. In addition, regarding 5-fold cross validation, future work could explore ensemble modelling, which aggregates the predictions from multiple CNN models either via averaging or weighted averaging.

3.1.5 Conclusion

In this research, we present an automatic approach using a convolutional neural network to segment the urethra from a 3D TPUS volume. The task was highly unbalanced, and to overcome this, we implemented several techniques to ensure that features from the urethra were learnt during training. The performance metrics show that the automatic approach is comparable to manual segmentation.

3.2 Multi-class segmentation

This section discusses multi-class segmentation and is based on an abstract submission to the IUS IEEE 2020 conference.

Multi-class segmentation is the task of segmenting several structures from an US image. Therefore, instead of a binary segmentation (i.e., the voxels of the volume are either 0 for background and 1 for foreground), the foreground values will be integers between and including the values 1 and n, with n being the number of structures to be segmented. The CNN learns features of all structures being segmented during the training phase, and the output of the trained CNN is a multi-class segmentation. To our knowledge it is the first work to present the simultaneous segmentation of several pelvic floor structures from a TPUS volume.

The performance of CNNs for segmentation has increased with data augmentation techniques, where realistic transformations are performed on available images and their annotations to create new data for network training. *Mixup* has recently been proposed as a successful data augmentation technique, whereby two training examples are combined linearly. In this work, traditional *mixup* was compared to an adapted version of *mixup*, where three training examples are combined linearly. In addition, these methods were compared to a CNN that did not use *mixup* as a data augmentation technique.

These methods were trained on a small dataset of 25 TPUS volumes as a multi-class segmentation task of the symphysis pubis, levator ani muscle and urethra. These landmarks are clinically relevant due to their location within the plane of minimal hiatal dimensions. Specifically, the symphysis pubis and levator ani muscle define the position of the plane of minimal hiatal dimensions, and the urethra lies within the mid-line of the plane of minimal hiatal dimensions. Automatic localisation of the symphysis pubis and levator ani muscle will be used in subsequent chapters, to automatically define the plane of minimal hiatal dimensions (as a first step towards automatic hiatal area segmentation from a TPUS volume).

The traditional *mixup* of two volumes improved the 95th percentile Robust Hausdorff distance from 9.71mm (without *mixup*) to 8.90mm, while the proposed *mixup* of three volumes improved the results further to 8.49mm. Suggesting that *mixup* may be a useful data augmentation technique for segmentation tasks with a small training dataset size.

3.2.1 Introduction

Data augmentation has been proposed, where realistic transformations of the available images and their annotations (e.g., rotation, scaling and elastic deformation) are considered new images to help train the CNN model. Recently, an alternative technique was proposed, called *mixup*, which creates new training examples by linearly combining two existing examples. This method was shown to improve the generalisation error of CNN models on ImageNet (a large visual database designed for visual object recognition), speech, and medical image segmentation tasks [174, 175].

The aim of this study was threefold: i) to validate this method in a multi-class landmark detection task in 3D transperineal pelvic floor US volumes, ii) to propose an adapted version of *mixup* to improve the performance further and iii) to generate context for a future diagnostic plane detection algorithm.

3.2.2 Material and methods

Data

A dataset containing 25 TPUS volumes of the pelvic floor from 20 patients was collected from UZ Leuven, Belgium. 18 TPUS volumes were used for training, 3 for validation and 4 for testing. The US volumes were acquired during a PFD examination (i.e., images were obtained at full Valsalva, rest and pelvic floor contraction). For each US volume, three expert-annotated landmarks were obtained. These landmarks were the symphysis pubis, the levator ani muscle and the urethra. These landmarks are clinically relevant due to their location within the plane of minimal hiatal dimensions. Specifically, the symphysis pubis and levator ani muscle, which define the position of the plane of minimal hiatal dimensions, and the urethra which lies within the mid-line of the plane of minimal hiatal dimensions.

The manual segmentations were curated using GE Vocal software (GE Healthcare; Zipf, Austria) by an expert with over four years of experience in TPUS imaging.

Network training

The network architecture used was 3D U-Net [92], as it has been proven to perform well in other 3D US semantic segmentation tasks. An Adam optimiser, ReLU activation function, learning rate of $1 \cdot 10^{-4}$, and batch size of 1 were used during training. Given the highly unbalanced nature of the data (i.e., the size of the landmarks is small compared to the size of the TPUS volume), an adapted version of cross-entropy was used during training as the loss function. This adapted version of cross-entropy was used to weight classes according to their volume to total-volume ratio. This loss function increases the exposure the CNN would have to the foreground voxels of the structures, despite their small volume compared to the total TPUS volume.

3.2 Multi-class segmentation

Validation of the network was performed every 100 epochs and the network was trained for 4000 epochs. Early stopping based on the lowest validation cross-entropy loss value was employed, to select the CNN model for each experiment.

During training, the following data augmentations were used for all methods: random rotation (with a range of -5° , $+5^\circ$) and random scaling (with a range of -10% , $+10\%$). The traditional *mixup* was used to augment the dataset in one method, and we used a modified version of *mixup* for the proposed method. Traditional *mixup* proposed by Zhang *et al.* constructs virtual training examples from the training dataset:

$$\tilde{x} = \lambda x_i + (1 - \lambda)x_j$$

$$\tilde{y} = \lambda y_i + (1 - \lambda)y_j$$

where x_i, x_j are TPUS volumes and y_i, y_j are the corresponding ground truth labels. (x_i, y_i) and (x_j, y_j) are two examples drawn at random from the training data, and $\lambda \in [0, 1]$. *Mixup* extends the training dataset by incorporating the prior knowledge that linear interpolations of features should lead to linear interpolations of the associated target features. A visual example of *mixup* is shown in Fig. 3.5. *Mixup* is favourable, as it can be implemented easily and has minimal computation overhead [174, 175].

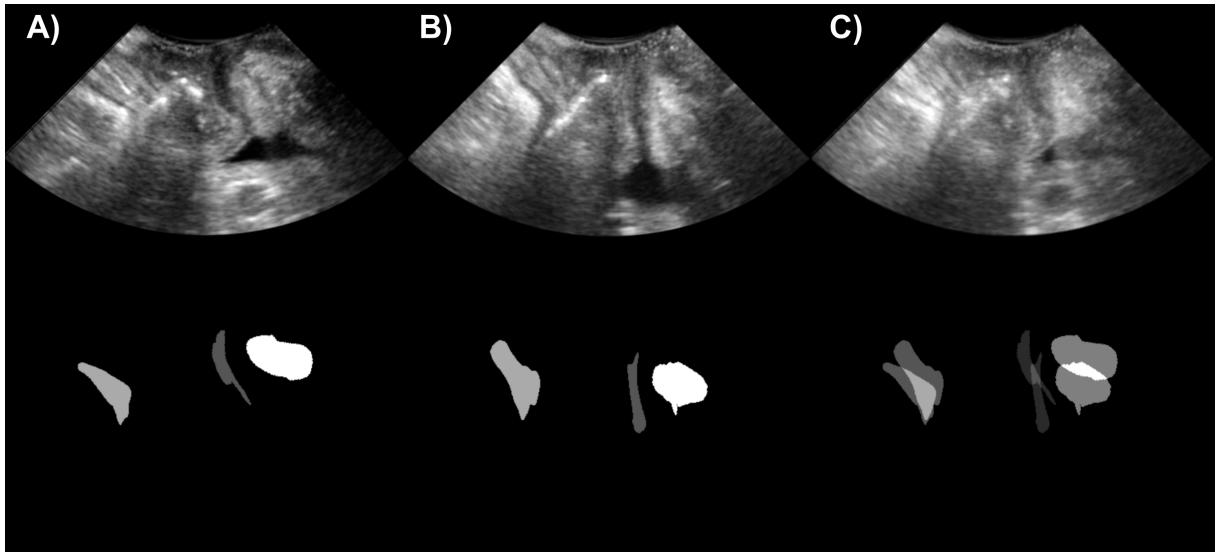


Fig. 3.5 Examples of TPUS mid-sagittal views of the levator ani muscle, symphysis pubis and urethra from a patient at rest, A), and at pelvic floor contraction, B), with accompanying segmentation masks below. C) shows the result of the traditional *mixup* data augmentation technique of the two TPUS images A and B with $\lambda = 0.5$.

The proposed modified version of *mixup* linearly combines three examples:

$$\tilde{x} = (0.7 - \lambda)x_i + \lambda x_j + 0.3x_k$$

$$\tilde{y} = (0.7 - \lambda)y_i + \lambda y_j + 0.3y_k$$

(x_i, y_i) , (x_j, y_j) and (x_k, y_k) are three examples drawn at random from the training dataset, and in this case $\lambda < 0.7$.

Evaluation

Evaluation was performed on a testing dataset of four TPUS volumes. To assess the true performance of the data augmentation techniques, no post-processing was applied to the output of the CNN, unlike traditional methods. To evaluate the accuracy of this multi-class segmentation model, the 95th percentile Robust Hausdorff distance between the CNN output and ground truth label was calculated for the three structures and averaged.

3.2.3 Results and discussion

Fig. 3.6 shows an example of the detected landmarks for standard training, training with data augmentation using traditional *mixup* and modified *mixup*. Quantitatively, the traditional *mixup* of two volumes improved the 95th percentile robust Hausdorff distance from 9.71mm to 8.90mm, while the proposed *mixup* of three volumes improved the results further to a 95th of 8.49mm. To note, the 95th percentile Robust Hausdorff distance is the average of three structures. Therefore, it is expected to be higher than the 95th percentile robust Hausdorff distance measured in the single-class segmentation task (Section 3.1).

Results showed that 3D U-Net could segment the symphysis pubis, urethra and levator ani muscle simultaneously on a small training dataset. In addition, our results suggest that *mixup*



Fig. 3.6 Shows the automatic segmentation result for the levator ani muscle (yellow), the urethra (green) and the symphysis pubis (red), compared to the expert-annotated ground truth segmentation (white overlay) on a TPUS volume. A) shows the CNN output trained with no *mixup* data augmentation, B) traditional *mixup* data augmentation of two images and C) modified *mixup* data augmentation of three images.

improves segmentation performance, especially when a small training dataset exists. Therefore, it may be a suitable and beneficial data augmentation technique for other small training dataset segmentation tasks, which is common in the research field of automatic segmentation of US images.

It is worth noting that there is limited methodological novelty in the proposed *mixup* algorithm presented in this work. Further validation is required, to determine the extent that *mixup* and the proposed *mixup* data augmentation techniques improve CNN performance on small training datasets. In addition, further investigation is required to determine whether the proposed *mixup* technique is significantly better than the original *mixup*. However, this preliminary study does suggest that such techniques could be beneficial for CNN segmentation training and *mixup* will be used in future studies regarding plane detection and hiatal area segmentation.

3.2.4 Conclusion

In this research, we adapted the data augmentation technique *mixup*, and segmented the symphysis pubis, urethra and levator ani muscle simultaneously from a TPUS using 3D U-Net. The adapted version of *mixup* reduced the 95th percentile Robust Hausdorff distance from 9.71 (no *mixup*) and 8.90mm (traditional *mixup*) to 8.49mm. Suggesting that *mixup* may be a useful data augmentation technique for segmentation tasks with a small training dataset size.

Chapter 4

Automatic extraction of the C-plane from a transperineal ultrasound volume

Clinical guidelines stipulate to make measurements of the levator hiatus, they must be made in an anatomically 2D plane within a 3D TPUS 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 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 post-processing step. When evaluated on 100 US volumes, our best performing method (multi-task regression with U-Net) achieved a mean error of 6.05mm 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.

4.1 Introduction

The plane of minimal hiatal dimensions (defined as the C-plane) as shown in Fig. 4.1, is defined as the angled axial plane of shortest distance between the symphysis pubis (SP) and the

This chapter is based on the peer-reviewed conference paper published in the book series, *Lecture Notes in Computer Science*: **Williams, H.**, Cattani, L., Yaqub, M., Van Schoubroeck, D., Sudre, C., Vercauteren, T., Deprest, J., D’hooge, J. (2020). Automatic C-Plane Detection in Pelvic Floor Transperineal Volumetric Ultrasound. Medical Ultrasound, and Preterm, Perinatal and Paediatric Image Analysis. ASMUS PIPPI 2020 2020. Lecture Notes in Computer Science(), vol 12437. Springer, Cham.

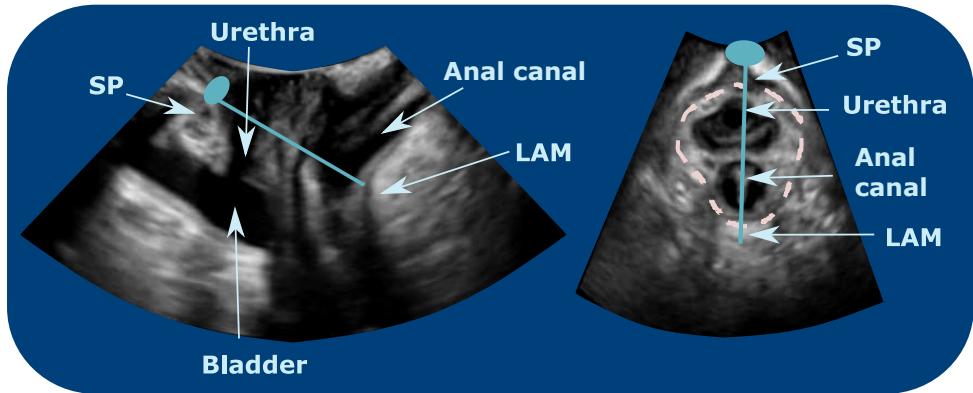


Fig. 4.1 US at rest, with identification of the C-plane in the mid-sagittal (MS) (left) and oblique axial (right) planes. The blue line shows the vector of minimal hiatal dimensions and the pink contour delineates the hiatal area.

pubovisceral muscle at the anorectal angle, referred to as the levator ani muscle (LAM) [176]. Assessments follow International Urogynecological Association (IUGA) and The American Institute of Ultrasound in Medicine (AIUM) standards [177]. Manual C-plane detection is prone to error, 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

As described in Section 2.4 several studies have successfully extracted 2D *standard* planes from 3D US volumes. Several studies investigated the extraction of 2D *standard* planes from 3D fetal and cardiac US volumes [178, 162, 164–166]. These methods utilised clinical knowledge but not necessarily clinical guidelines. Zhu *et al.* [166] 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 mid-point of the plane. Our contributions consist of: a) a new methodological approach to identify the C-plane; b) a novel solution to a clinically challenging problem using landmark regression; and c) a clinically viable blinded evaluation of C-plane detection.

4.2 Material and methods

4.2.1 Data and manual detection of the C-plane

A dataset containing a training subset of 25 TPUS volumes from 15 patients, and a validation subset of 100 TPUS 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. Training ground truth data consisted of TPUS volumes and corresponding heatmaps of the SP and LAM extreme (i.e., the plane-defining) coordinates. The heatmap is a data visualisation technique which encodes the probability of a landmark being located at a particular voxel position within the TPUS volume. The process used to obtain these ground truth heatmaps is shown in Fig. 4.2.

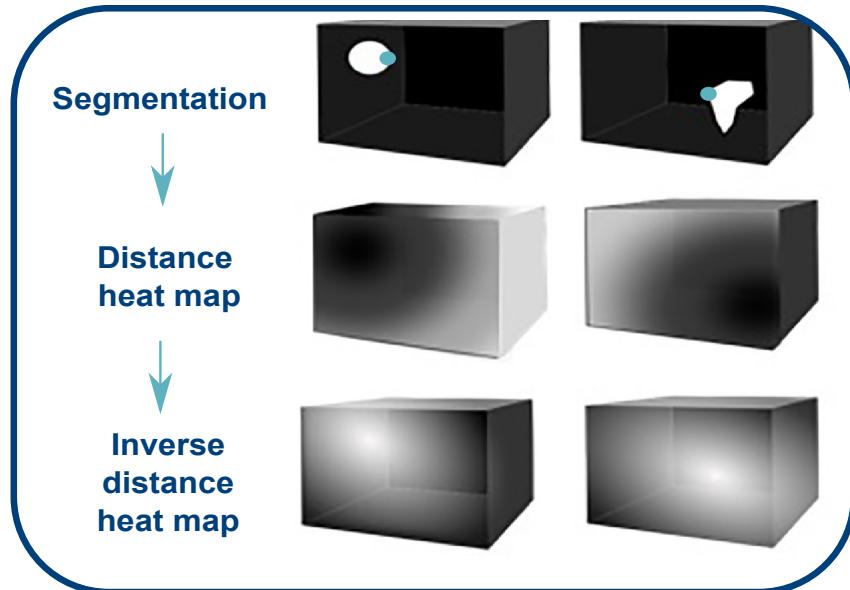


Fig. 4.2 Visualization of the steps to generate the ground-truth heatmaps used in this study. The desired heatmap of the extreme coordinates (*blue dots*) of the SP (left) and LAM (right) are identified from 3D segmentations manually delineated by experts. The first row shows the segmentations of the symphysis pubis and levator ani muscle; the second row shows the distance heatmap (i.e., extreme coordinate = 0 and the voxel value radially increases with distance); and the third row shows the smooth inverse distance heatmap used for CNN training (i.e., extreme coordinate is the maximum value and the voxel value radially decreases with distance).

The dataset was curated at UZ Leuven, Belgium, over the course of 6 months. It includes a 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 [177]. Evaluation data of the *desired* C-planes was curated following the clinical protocol. To acquire the C-plane manually, 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; Zipf, 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 an 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, patient pathology and image quality.

4.2.2 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. 4.3.

Fig. 4.4 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. The C-plane is defined in the IUGA guidelines as the plane orthogonal to the depth direction of the US volume (i.e., $\begin{vmatrix} 0 & 0 & 1 \end{vmatrix}$) and containing the vector \vec{AB} . The second orthogonal vector representing the plane can be defined as the cross product of the normal, and \vec{AB} , thus we define the final orthogonal vector, b_y , as $-AB_y\mathbf{i} + AB_x\mathbf{j} + 0\mathbf{k}$. \vec{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.

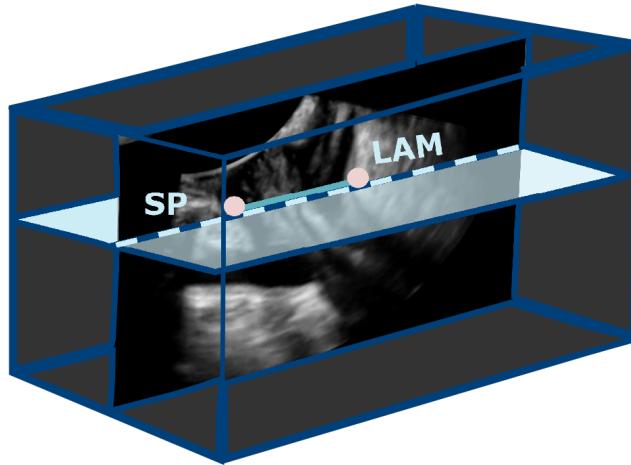


Fig. 4.3 Extreme coordinates shown as pink dots define the C-plane bisecting the MS plane. The vector of minimal hialt dimensions is shown as a cyan coloured line.

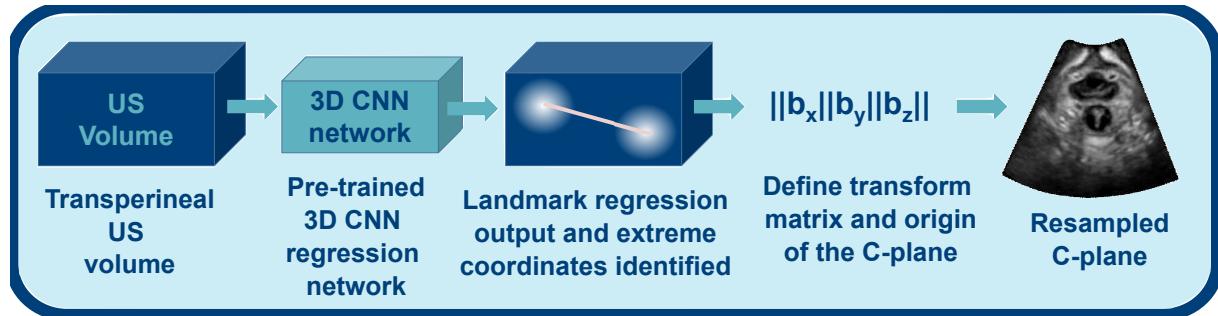


Fig. 4.4 Pipeline of the landmark regression approach to C-plane extraction.

Thus, we have defined the bases of the C-plane and can formulate the following direction matrix

$$\begin{vmatrix} b_x \\ b_y \\ b_z \end{vmatrix} = \begin{vmatrix} AB_x & -AB_y & 0 \\ AB_y & AB_x & 0 \\ 0 & 0 & 1 \end{vmatrix}. \quad (4.1)$$

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.

4.2.3 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. 4.4). The rationale for this approach is that regression of one coordinate from an average US volume of $450 \times 500 \times 450$ voxels is hard to train. Hence, to make the regression model robust, we propose to

regress a distance map of the SP and 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 outputs (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, \vec{AB} , and mid-point, \vec{O} , of the C-plane.

4.2.4 Evaluation methodology

Intra- and inter-observer variability

To evaluate the proposed method, we compare results against inter and intra-observer variability. The expert who determined the GS C-planes (Expert 1), determined the C-plane again a month later, on a subset of 37 volumes. We obtained intra-observer variability using the MP (same method as the GS) and the OV method. Inter-observer 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 performed a paired Wilcoxon test to compare the proposed method performance against the GS. Where a positive score indicates the computer-aided method performed visually better than the GS, a score of 0 indicates they performed similar to the GS, and a negative score indicates they performed worse than the GS.

Bio-marker analysis

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

4.2.5 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 [167] on a desktop with a 24GB NVIDIA Quadro P6000.

Two CNN architectures were compared. First an implementation of 3D U-Net [92] by Isensee *et al.* [96], almost identical to the original 3D U-Net, but due to high memory consumption of 3D convolutions with large patch sizes, the number of filters was reduced before up-sampling (i.e., [16,32,64,128] filters were used). This network was chosen due to its performance in levator hiatus segmentation [78]. The second network implemented was HighRes3DNet, proposed by Li *et al.* [94]. 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 [106]. HighRes3DNet, has less trainable parameters than U-Net, suggesting a faster inference time, which is beneficial for an *in-clinic* solution. It is important to remove the final SoftMax layer from the CNN architecture (if present) to ensure the output has a continuous range of values between zero and a maximum voxel value.

Both multi-task Regression with U-Net (mRegU) and multi-task Regression with HighRes3DNet (mRegHR) have a combined L2 loss of the SP and LAM distance maps. The networks had an initial learning rate of 10^{-4} , weighted decay factor of 10^{-5} and batch size of four. Histogram-based normalisation and whitening were used, to set the volume to have zero mean and unit variance. 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 (with a range of -10% , $+10\%$), and scaling (with a range of -10% , $+10\%$) were used.

4.3 Results and discussion

Fig. 4.5A 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

Chapter 4. Automatic extraction of the C-plane from a TPUS volume

smallest distance to the GS, highlighting that U-Net 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-observer 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. 4.5B, the angular error is presented, showing that mRegU’s interquartile range is smaller than the inter-observer 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 hiatus ballooning resulting in the levator hiatus 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. 4.6 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 levator hiatus); and all structures are visible. The results of the *Turing test*, presented in Table 4.1, 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 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 may obtain a better-quality image of the

Table 4.1 The Hausdorff distance and *visual Turing test* results for automatic C-plane extraction.

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-observer (MP)	8.30 ± 2.27	5.52 ± 2.99	-	-	120
Inter-observer (MP)	8.68 ± 3.83	5.99 ± 3.11	-	-	120

4.3 Results and discussion

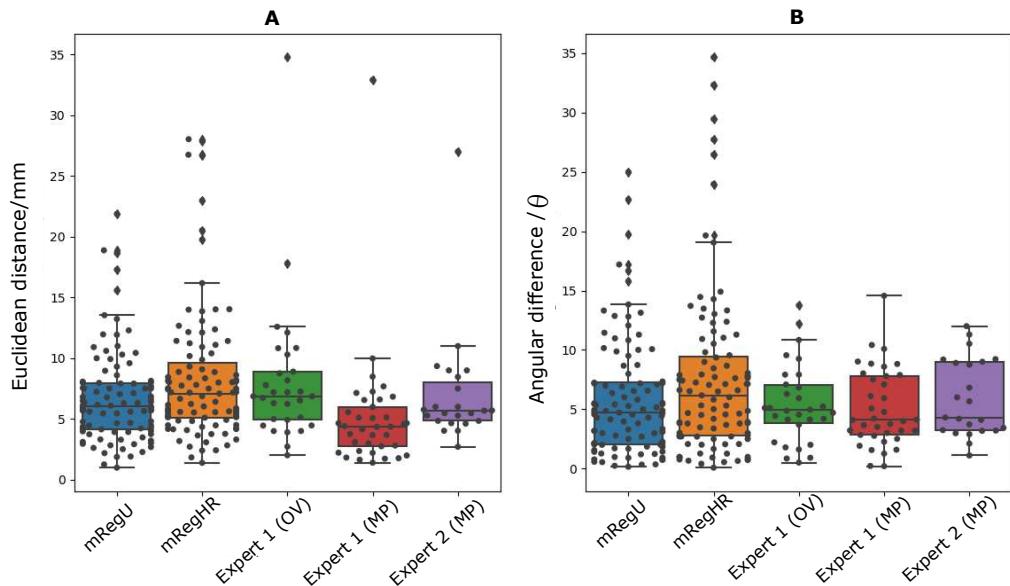


Fig. 4.5 A) Euclidean distance of the mid-point B) angular difference of the planes determined by the automatic method, Expert 1 (MP), Expert1 (OV), and Expert 2 (MP) against the GS (generated using MP).

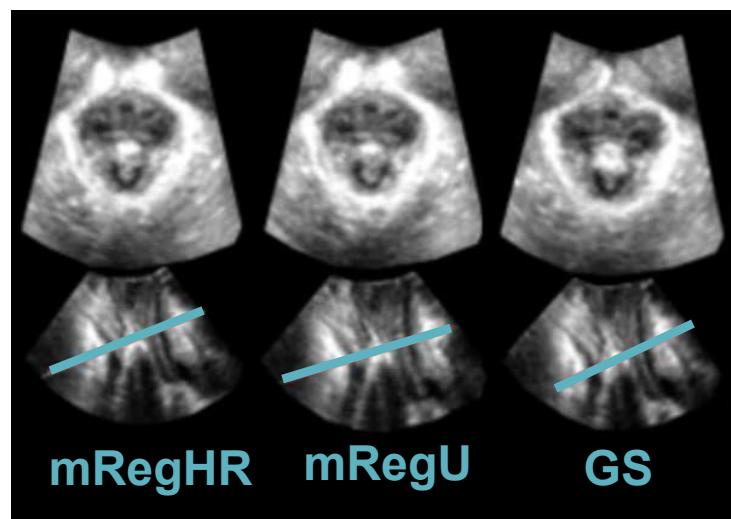


Fig. 4.6 Visualisation of the axial (upper) and MS (lower) plane of minimal hiatal dimensions for automatic methods and GS. The plane of minimal hiatal dimensions is shown as a blue line on the MS plane.

levator hiatus 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, which we assume helps to achieve relatively low errors with a small training dataset, as we utilise the geometry of the patient.

Results of Table 4.1 show the HD of levator hiatus segmentation compared to the GS, mRegU lies within inter-observer 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 optimisation. The time difference, is due to U-Net having more parameters than HighRes3DNet. Manually on average, Expert 1 and Expert 2 determine the C-plane in 2 minutes. However, this can vary greatly due to patient's exhibiting PFD, poor acquisition and limited clinical experience.

To 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 TPUS 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. This work is expanded by including 2D segmentation of the levator hiatus in Chapter 5, to produce the first, fully automatic PFD bio-marker analysis pipeline.

4.4 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 learns from a small set of examples. We identified that mRegU performed better than the mRegHR, suggesting the U-Net architecture is favourable. In Chapter 5, the 2D segmentation of the levator hiatus will be included [78] to the pipeline. This will produce the first fully automatic bio-marker analysis pipeline, that can detect the C-plane, segment the levator hiatus, and indicate PFD disorders from an US volume

4.4 Conclusion

[177, 179]. 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.

Chapter 5

Automatic segmentation of the hiatal area from a transperineal ultrasound volume

The aims of this chapter are to create a robust automatic software tool for measurement of the levator hiatal area on TPUS volumes and to measure the potential reduction in variability and time taken for analysis in a clinical setting. The proposed tool automatically detects the C-plane (i.e., the plane of minimal hiatal dimensions) from a 3D TPUS volume (based on Chapter 4) and subsequently uses the extracted plane to automatically segment the levator hiatus, using a CNN.

The automatic pipeline was tested using 73 representative TPUS volumes. Reference hiatal outlines were obtained manually by two experts and compared with the pipeline's automated outlines. The Hausdorff distance, hiatal area, a clinical quality score, C-plane angle and C-plane Euclidean distance were used to evaluate C-plane detection and quantify levator hiatus segmentation accuracy.

A visual Turing test was created to compare the performance of the software with that of the expert, based on the visual assessment of C-plane and hiatal segmentation quality. The overall time taken to extract the hiatal area with both measurement methods (i.e., manual and automatic) was measured. Each metric was calculated both for computer–observer differences and for inter-and intra-observer differences.

The automatic method gave results similar to those of the expert when determining the hiatal outline from a TPUS volume. Indeed, the hiatal area measured by the algorithm and by an expert were within the intra-observer variability. Similarly, the method identified the C-plane with an accuracy of $5.76 \pm 5.06^\circ$ and 6.46 ± 5.18 mm in comparison to the inter-observer variability of $9.39 \pm 6.21^\circ$ and 8.48 ± 6.62 mm.

This chapter is based on the journal paper published in *Ultrasound in Medicine & Biology*: **Williams, H.**, Cattani, L., Van Schoubroeck, D., Yaqub, M., Sudre, C., Vercauteren, T., D'hooge, J., Deprest, J. (2021). Automatic Extraction of Hiatal Dimensions in 3-D Transperineal Pelvic Ultrasound Recordings. *Ultrasound in Medicine & Biology*, Volume 47, Issue 12, 3470-3479.

The visual Turing test suggested that the automatic method identified the C-plane position within the TPUS volume visually as well as the expert. The average time taken to identify the C-plane and segment the hiatal area manually was 2 min and 35 ± 17 s compared with 35 ± 4 s for the automatic result.

This chapter presents a method for automatically measuring the levator hiatal area using AI-based methodologies whereby the C-plane within a TPUS volume is detected and subsequently traced for the levator hiatal outline. The proposed solution was determined to be accurate, relatively quick, robust and reliable and, importantly, to reduce time and expertise required for PFD assessment.

5.1 Introduction

Delivery-induced sarcomeric hyperelongation may cause substantial, irreversible ultrastructural trauma in the LAM [180, 181]. Irreversible overdistension of the levator hiatus ('microtrauma') has been described in postpartum women as a possible consequence of muscular atrophy and reduction in function and can alter pelvic floor distensibility after vaginal delivery [182].

Manual detection of the levator hiatus in a 3-D TPUS acquisition requires significant offline post-processing of the volumetric recordings by specifically trained sonographers. US manufacturers have implemented and previous works [100, 102, 104] have developed semi-automatic and automatic tools to aid pelvic floor ultrasound examination. For instance, real-time visualization of the desirable C-plane from a manually identified approximation of the C-plane was developed in Omniview-VCI (GE Healthcare; Zipf, Austria).

Clinicians assess the levator hiatus on the plane where the anteroposterior distance (between the dorsocaudal end of the symphysis pubis (SP) and the ventral end of the levator), is the smallest, and refer to this plane as the plane of minimal hiatal dimensions or C-plane. However, a fully automatic levator hiatus detection from a TPUS volume should obtain a more accurate representation of the anatomical findings, would be less operator dependent and may save clinicians time to allow more focus on patient care and counselling. Automation would lower the minimal threshold of expertise needed by clinicians to use TPUS.

This chapter aimed at building a fully automatic workflow that consists of C-plane detection (based on Chapter 4) followed by hiatal segmentation. A solution to this clinical problem [177] that ensures the trustworthiness and interpretability from experts while following the clinical guidelines is likely to have strong clinical value.

5.2 Material and methods

5.2.1 Manual C-plane detection

Three-dimensional TPUS acquisition is performed by orienting the 3D abdominal probe as on conventional transvaginal US images (cranioventral aspects to the left, dorsocaudal aspects to the right) [5]. In so-acquired 3D images of the pelvic floor, the midsagittal plane is in the top left corner (A), the coronal plane is in the top right corner (B) and the axial plane is in the bottom left corner (C) (Fig. 5.1). To visualise the C-plane on the axial plane of the US image, clinicians manually align the SP and LAM to a horizontal direction on the midsagittal plane. Eventually, the LAM lies on the coronal plane, as illustrated in Fig. 5.1. This makes the levator hiatus clearly visible on the axial plane as the pubic bones ventrally, and the LAM dorsally is hyperechogenic compared with the hypo-echogenic pelvic organs. Through analysis of the levator hiatus, levator avulsion and hiatal ballooning can be diagnosed [177].

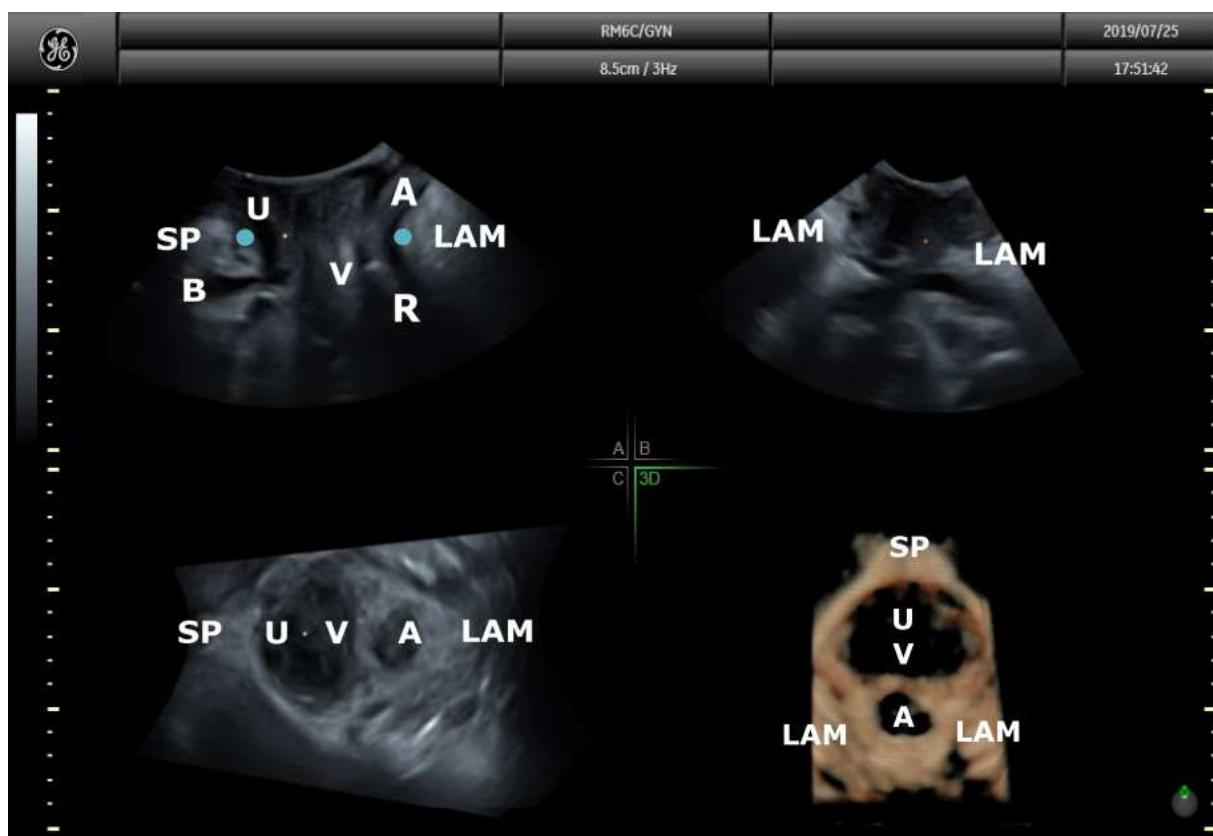


Fig. 5.1 The typical TPUS acquisition and evaluation screen on Voluson systems (GE Healthcare, Zipf, Austria) has three orthogonal planes: (top left/A) sagittal, (top right/B) coronal and (bottom left/C) axial; the bottom right image (3-D) is the axial plane rendered volume. This volume has been aligned as the desired minimal hiatal dimensions position, and the extreme coordinates are marked by blue dots. A = anal canal; B = bladder; LAM = levator ani muscle; R = rectum; SP = symphysis pubis; U = urethra and V = vagina.

5.2.2 Proposed biomarker extraction pipeline

The proposed automatic data analysis pipeline is composed of two sequential parts: a C-plane extractor and a levator hiatus outline extractor (Fig. 5.2). The C-plane extractor is based on Chapter 4. The proposed pipeline expands on this work to automatically outline the levator hiatus from the C-plane extractor's output. The proposed pipeline uses advances in CNNs, landmark detection (Chapter 4) and semantic segmentation (Chapter 3) and follows the IUGA/AIUM [177] clinical guidelines to ensure interpretability of the results. The solution requires no user input and is thus completely automatic. In brief, the pipeline starts by automatically detecting the SP and LAM extreme coordinates within a TPUS via CNN landmark regression. The extreme coordinates are defined as the voxel coordinates with the shortest Euclidean distance between the 3D segmentations of the SP and LAM within a midsagittal (MS) slice, as illustrated in Chapter 4 Fig. 4.2. Post-processing identifies the vector of the minimal hiatal dimensions, and a transformation matrix can be formed to resample the TPUS volume as the desired 2D C-plane. The extracted C-plane is then used as input to a pre-trained 2D semantic segmentation CNN model that segments the levator hiatus, defining the hiatal area.

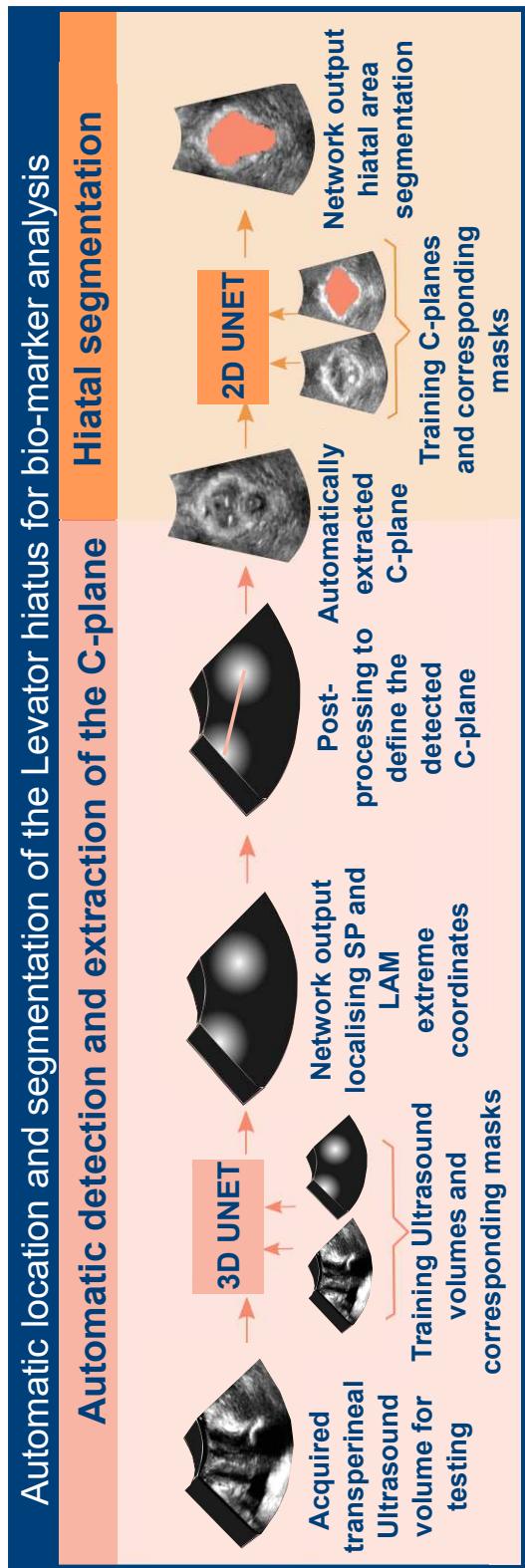


Fig. 5.2 Overall levator hiatus analysis pipeline split into two tasks by colour. The first section (*pink*) is the automatic detection and extraction of the C-plane; the second task (*orange*) being the automatic segmentation of the hiatal area within this C-plane.

5.2.3 Description of the biomarker pipeline

1. Three-dimensional landmark regression of the SP and LAM extreme coordinates. The first step of the C-plane extractor accepts a TPUS volume as input and results in a heatmap of the SP and LAM extreme coordinates within the TPUS volume. The heatmap is a data visualization technique that encodes the probability of a landmark being located at a certain voxel position within the TPUS volume. In this chapter, the heatmap voxels near the extreme coordinate have high values (with the highest at the extreme coordinate), and they smoothly and rapidly decrease with increasing distance from the extreme coordinate, as illustrated in Chapter 4 Fig. 4.2.

The rationale underlying this approach is that regressing one coordinate from a large volume can be difficult to train, and a heatmap is more robust (as described in Chapter 4) [107]. The CNN architecture used was an adaptation of U-Net [92], and the heatmaps were regressed in training. A multi-task approach was used to determine the distinct SP and LAM heatmaps simultaneously, using transfer learning between the two tasks. Finally, the SoftMax layer of U-Net was removed to generate a continuous output.

2. Post-processing to identify the minimal hiatal dimension. The second step identifies the extreme coordinates from the regression output. This was achieved with a computational post-processing step inspired by the IUGA clinical guidelines [177]. While our landmark regression was performed in 3D, clinicians normally identify the plane defining “extreme coordinates” within a single 2D MS plane (as described in Chapter 4, Section 4.2.3) [107]. Thus, to follow clinical guidelines, a 2D approach was also followed in our automatic pipeline, to create a workflow that was comparable to the clinical one. The combined voxel maxima of the SP and LAM heatmaps were determined within a small range of 2D MS planes to reduce computational load and running time. Thus, the SP and LAM combined overall voxel maxima, corresponding extreme coordinates and MS plane were identified.
3. Extraction of the C-plane. The final step of task 1 was to slice and resample the 3D TPUS as the automatically defined 2D C-plane. The C-plane was defined as the plane orthogonal to the depth direction of the TPUS volume at acquisition and thereby, containing the orthogonal vector, $\begin{vmatrix} 0 & 0 & 1 \end{vmatrix}$. The C-plane also contains the vector \vec{AB} that joins the extreme coordinates of SP and LAM identified in the previous step. The cross-product of these two orthogonal vectors defines the final orthogonal vector as $-AB_y\mathbf{i} + AB_x\mathbf{j} + 0\mathbf{k}$. Clinical guidelines suggest the vector \vec{AB} has a magnitude within the x- and y-directions only, as the extreme coordinates lie within the same MS plane (z-slice) [107], which was

determined in the preceding section. Therefore, the bases of the C-plane are defined as

$$\|b_x\| \|b_y\| \|b_z\| = \begin{vmatrix} AB_x & -AB_y & 0 \\ AB_y & AB_x & 0 \\ 0 & 0 & 1 \end{vmatrix}. \quad (5.1)$$

Once the TPUS volume was rotated, the C-plane was extracted at the midpoint between the SP and LAM extreme coordinates.

4. Levator hiatus segmentation. The second task of the proposed pipeline was to automatically define the hiatal area from the extracted 2D C-plane, elaborating on previous work [100, 78]. In this chapter, a 2D CNN accepts the extracted 2D C-plane from the previous task and automatically classifies the voxels as levator hiatus (1) or background (0). The network architecture used was an implementation of 2D U-Net [91]. Owing to the nature of US can make segmentation difficult because of noise, artifacts and blurring, advanced data augmentation was used, including elastic deformation and our own adaptation of the original *mixup* [175] (from Chapter 3, Section 3.2), where three instead of two images and their corresponding ground-truth labels were linearly combined. Post-processing morphological operators were applied to the CNN output, such as connected component analysis, fill-holes and Gaussian blur of σ value 0.5. This post-processing was used to ensure that the segmentation was complete (i.e., no holes) and that the boundary was smooth, which ensures the hiatal output was more realistic.

5.2.4 Implementation details

The CNN models were implemented using NiftyNet [167] on a desktop with a 24-GB NVIDIA Quadro P6000 (NVIDIA; Santa Clara, CA, USA).

3-D landmark regression

The network architecture of 3D U-Net [92] was adapted to have one input (i.e., TPUS volume) and two outputs (i.e., SP and LAM heatmaps) at testing, to ensure a multi-task approach to learning. The final SoftMax layer was removed to output a continuous value that ranges between zero and a maximum value. The loss function was a combined L2 loss of the SP and LAM heatmaps with an initial learning rate of 10^{-4} . An root mean squared propagation (RMSprop) optimizer, parametric rectified linear unit (ReLU) activation function, weighted decay factor of 10^{-5} and batch size of six were used. Histogram-based normalization and whitening were used; thus, the volume was set to have zero mean and unit variance. A combined smooth version of the heatmaps was used for weighted sampling during training. The following data augmentations were used: random scaling (with a range of $-10\%, +10\%$), random rotation of all axes (with a

range of $-10^\circ, +10^\circ$) and our own adaptation of *mixup* [175] (Chapter 3, Section 3.2). Methods were optimized until network convergence of a validation dataset (i.e., a subset of TPUS volumes from the training dataset).

C-Plane hiatal area segmentation

The network architecture used was an adaptation of 2D U-Net [91] as it has been proven to perform well in other 2D US semantic medical imaging tasks [78, 102]. An Adam optimizer, ReLU activation function, weighted decay factor of 10^{-5} and batch size of 32 were used. Whitening was applied to reduce the effects of noise; thus, the image was set to have zero mean and unit variance, and histogram normalization was performed [183]. A loss function of combined cross entropy and Dice score was used, with an initial learning rate of 10^{-3} . Balanced window sampling was used during training (i.e., regions of label and background were equally sampled). During training the following data augmentations were used: random rotation (with a range of $-5^\circ, +5^\circ$), elastic deformation (deformation $\sigma = 9$, number of control points = 4 and proportion to deform = 0.5), random scaling (range of $-20\%, +20\%$), vertical “flipping” and our implementation of *mixup* [175].

5.2.5 Data collection

Analysis of anonymized, archived, US images was retrospective; therefore at the time of this work, no ethics committee approval was required by KU Leuven, Belgium.

Training data: C-Plane detection

With respect to the 3D C-plane detection task, a training dataset of 25 3D TPUS volumes were used. This was the same dataset used in Chapter 4 [107]. The training dataset comprised 13 clinical cases with a range of pelvic floor dysfunctions, assessed at the pelvic floor clinic in UZ Leuven, Belgium. Multiple TPUS volumes were obtained from the 13 clinical cases (i.e., at rest, Valsalva and/or pelvic floor contraction). Three-dimensional segmentations of the SP and LAM were provided by an expert human annotator (referred to as expert 1) and used to generate the heatmaps via the process illustrated in Chapter 4, Fig. 4.2. Expert 1 was chosen for their experience in this domain and in annotating the 3D LAM and SP structures from a TPUS volume; expert 1 had 12 months of experience in annotating the SP and LAM in 3D TPUS volumes prior to data curation.

Training data: 2-D levator hiatus segmentation

Regarding levator hiatus segmentation, a training dataset of 256 2D C-planes and corresponding ground truth labels of the levator hiatus were used to train the CNN segmentation model. The

5.2 Material and methods

training dataset comprised of two sets of archived clinical images with expert annotations, acquired by several operators, which allows the CNN to learn a variety of acquisition parameters and image qualities. Within the training dataset, a subset of 91 2D C-planes with expert annotations were used in our previous studies [100, 78]; in this dataset the expert had > 4 y of experience in acquiring and analysing pelvic floor TPUS volumes.

Test data

The test data included a randomized selection of 73 anonymized TPUS volumes from 37 other symptomatic women assessed at the pelvic floor clinic, between February and June 2019. There was no patient overlap across training and testing sets. The test data was not used to train the CNN models; they were used purely for testing the proposed pipeline. Detailed patient information is included in Table 5.1.

Table 5.1 Characteristics of the study population. Data are presented as mean (standard deviation), as prevalence in % (ratio) or as median [IQR].

Demographic variables	Values
Age (years)	57.6 [14.3]
BMI (kg/m ²)	26.7 [3.8]
Obstetric variables	
Vaginally parous	75.7 % (28/37)
Only caesarian section	8.1 % (3/37)
Nulliparous	5.4 % (2/37)
Vaginal parity	2 [1.25]
Max birth weight in grams	3741 [439]
Symptoms of pelvic floor dysfunction	
<i>Urinary incontinence</i>	
- Stress urinary incontinence	48.7 % (18/37)
- Urge urinary incontinence	21.6 % (8/37)
<i>Pelvic organ prolapse</i>	81.1 % (30/37)
<i>Anal incontinence</i>	2.7 % (1/37)

Hiatal measurements were delineated by expert 1, resulting in gold standard (GS) C-plane orientations and levator hiatus segmentations used for validation. The GS C-plane orientations were extracted using GE 4DView software (GE Healthcare; Zipf, Austria), and the corresponding GS hiatal segmentations were delineated using 3D Slicer software, 2019 version [184]. Two operators (experts 1 and 2) participated in the inter-operator reliability studies. At the time of the analysis, both experts had > 4 y of experience in acquiring and analysing pelvic floor TPUS volumes. Both experts work as clinicians in the pelvic floor disorder clinic at UZ Leuven, Belgium, and were asked to identify the C-plane according to the IUGA guidelines [177]. The experts identified the C-plane using the multiplanar technique [107] on GE 4DView software

(GE Healthcare; Zipf, Austria). The experts performed manual hiatal outlining and C-plane detection on all 73 TPUS volumes.

5.2.6 Quantitative metrics for evaluation

Several metrics were used to describe the similarity of the manual C-plane detection and levator hiatus segmentation to the computer-generated output. As this was a “two-task” pipeline, both “tasks” were evaluated independently as well as jointly.

C-plane detection

Validation of the C-plane detection task was similar to that described in Chapter 4. To validate the accuracy of the plane detection task, the angular difference between the identified C-plane and the GS plane was measured, as was the Euclidean distance of the midpoints of the planes within the TPUS volume. The angular difference computed was the averaged x-axis and y-axis angular difference, as the z-axis was fixed per guidelines [177]. To evaluate clinical relevance, a visual Turing test was proposed and evaluated on 10 TPUS volumes. Hereto, expert 1 was asked to blindly rate a randomized selection of manually and automatically detected C and MS planes to give a Likert scale score from 0 – 5 (5 = excellent, 4 = above average, 3 = average, 2 = below average, 1 = poor and 0 = of no clinical use). Test 1 was based on the placement of the C-plane within the TPUS volume. Test 2 was based on the C-plane quality for clinical diagnosis. A paired Wilcoxon test was performed to compare the performance of the proposed method against expert 1’s GS recording; this generated an output score that was averaged per TPUS volume. The paired Wilcoxon test is calculated by deducting expert 1’s GS score from the method’s (i.e., algorithm, inter-observer or intra-observer) score. The score ranges from a negative value to a positive value depending on the performance of the detected C-plane against the GS. A positive score suggests the detected C-plane method performed “better” visually than the manual GS; a negative overall score suggests the detected method performed “worse” visually than the manual GS; and a score of zero means the methods performed the same.

Levator hiatus localisation and segmentation

The levator hiatus outline (i.e., hiatal area) identified in the C-plane is a biomarker used for the analysis of given pelvic floor disorders. To assess the quality of the extracted biomarker, the following metrics were computed: Hausdorff Distance (HD) and Robust 95th percentile HD of the levator hiatus segmentation, which were evaluated against the GS manual hiatal segmentation from the GS C-planes. The hiatal area is an important biomarker; thus, the area of the GS hiatal outline was compared with that of the hiatal outline of the automatically extracted C-plane. Moreover, the hiatal area difference and absolute hiatal area difference were calculated. To

evaluate the clinical acceptability of the segmentations, another visual Turing test (Turing test 3) was proposed and evaluated on 10 extracted C-planes and corresponding segmentations. The hiatal segmentations were rated a “clinical score” by expert 1, from 0 to 5 as described above, which were compared with the GS scores in a paired Wilcoxon test. Expert 1 performed the test 3 months after they annotated the GS hiatal segmentations to limit the impact of pre-learning bias. The average result per TPUS volume was presented, and the score ranges between +5 and -5.

5.2.7 Computer-observer, intra-observer and inter-observer differences

The computer–observer differences (CODs), intra-observer differences (IAODs) and inter-observer differences (IEODs) were evaluated. COD were evaluated by calculating all similarity metrics between automatic hiatal segmentations on automatic C-planes and expert 1 manual hiatal segmentations on GS C-planes. IAOD was evaluated by calculating similarity metrics between identified C-planes and hiatal outlines generated by expert 1 GS and a second analysis from expert 1 a month after the GS was generated. The second analysis was undertaken 2 months before the Turing test analysis, to reduce bias and the risk of experts recognising their analysis and thus rating it higher subconsciously. In addition, as the experts are active members of the clinical team at UZ Leuven, they analyse new TPUS volumes daily, and we assume bias is limited as this is a common and repetitive task. Finally, IEOD was evaluated by calculating similarity metrics between expert 2 and the first assessment from expert 1.

5.2.8 Statistical analysis

To evaluate the reliability of the automatic method, a paired F-test was used to test several null hypotheses. The first was that the automated method agreed with expert 1’s GS at least as well as expert 1 agreed with him- or herself (i.e., the variance of the differences between the automatic method and the GS was not larger than the variance in IAODs). The second null hypothesis tested was that the automated method agreed with expert 2 at least as well as expert 2 agreed with expert 1’s GS result (i.e., the variance of the differences between the automatic method and the GS was not larger than the variance in inter-observer differences). The final null hypothesis tested was that expert 2 agreed less with expert 1’s GS results than expert 1 agreed with him- or herself (i.e., the variance in inter-observer differences was statistically greater than the variance IAODs). Type one statistical errors (i.e., multiple testing) were accounted for using a Bonferroni correction; hence the p value obtained was reduced by a factor of 3. Therefore, a p value ≤ 0.017 was used as a cutoff to indicate statistical significance. To further evaluate the reliability of the automatic method, Bland–Altman limits of agreement were calculated for COD, IAOD and IEOD.

To evaluate the possibility of bias between the methods (i.e., automatic, expert 1 and expert 2) and expert 1's GS, several paired t-tests were used to test several null hypotheses. The null hypotheses were the same as above, however, based on the mean difference, (i.e., bias) rather than on the variance of the differences. As described above, type 1 statistical errors (i.e., multiple testing) were accounted for using a Bonferroni correction, and a p value ≤ 0.017 was used as a cutoff to indicate statistical significance.

5.3 Results

Fig. 5.3 shows examples of the C-plane position within the TPUS volume and the corresponding extracted C-planes and hiatal segmentations. The images represent the 0th, 25th, 50th, 75th and 100th percentiles, respectively, of the 95th Hausdorff distance metric (the corresponding 95th HD distances are included along with the hiatal areas). The *red lines* and *masks* represent the automatic method, and the *green lines* and *masks* represent the GS. Qualitatively, the computer-generated C-planes and levator hiatus segmentations matched well the GS C-planes and hiatal segmentations.

Table 5.2 Turing test score per transperineal ultrasound volume*.

	COD	IEOD	IAOD
Turing Test 1	0.00 ± 0.77	-1.00 ± 1.34	-0.20 ± 0.98
Turing Test 2	0.00 ± 1.07	-0.20 ± 0.77	0.30 ± 0.92
Turing Test 3	-1.50 ± 1.01	-0.80 ± 0.60	0.00 ± 0.44

COD = computer-observer difference, IEOD = inter-observer difference;
IAOD = intra-observer difference.

* The scores can range from -5 to $+5$ (dependent on the gold standard score and evaluated method score). A score of 0 means that the gold standard performed equally to the other method evaluated. A positive score would mean the method outperformed the gold standard, and a negative score implies that the gold standard performed better than the evaluated method.

Table 5.2 outlines the semiqualitative average results for COD, IEOD and IAOD from the *visual* Turing tests. The results indicate that the pipeline performs better than IEOD and IAOD in relation to the C-plane detection task within a TPUS volume, as COD scored 0.00 ± 0.77 for Turing test 1. The pipeline scored comparable to the GS with a score of 0.00 ± 1.07 for Turing Test 2 (C-plane quality), which was a lower error than IEOD scoring -0.20 ± 0.77 . The proposed pipeline scored -1.50 ± 1.01 for Turing test 3 (hiatal segmentation quality), whereas IEOD scored -0.80 ± 0.60 .

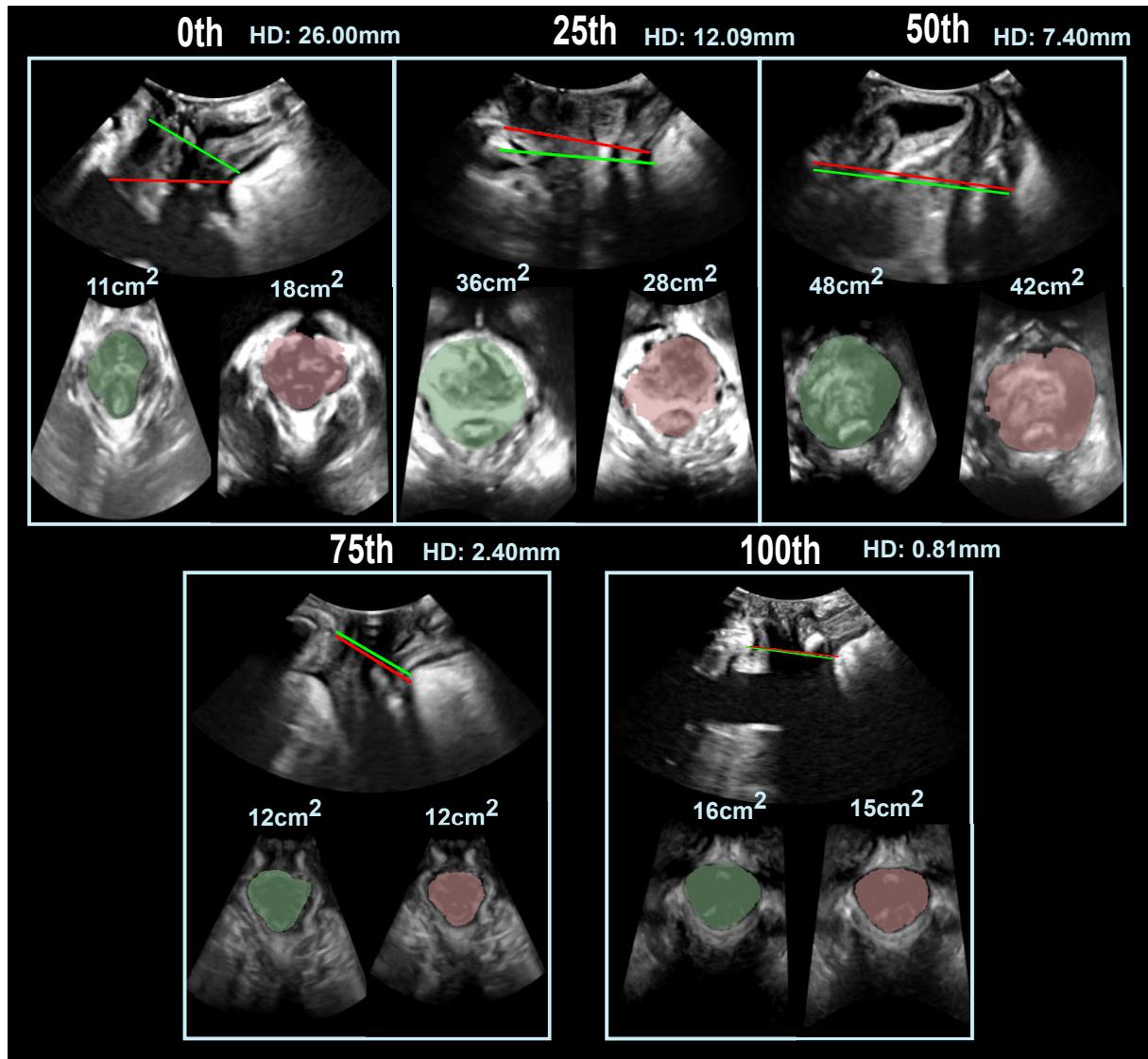


Fig. 5.3 The GS C-plane position is represented by a *green line*, and the computer-automated C-plane position is represented by a *red line*, for each corresponding TPUS. The corresponding GS manual segmentation of the hiatal area is the *green mask* and the automated segmentation of the hiatal area is the *red mask* under its corresponding TPUS image. TPUS images reveal an increasing computer-generated hiatal outline quality that represents the 0th, 25th, 50th, 75th and 100th percentiles, respectively, of the 95th Hausdorff distance. GS= gold standard; TPUS = transperineal ultrasound.

Chapter 5. Automatic segmentation of the hiatal area from a TPUS volume

Table 5.3 COD, IAOD and IEOD and standard deviations of C-plane detection metrics: Angular difference of the C-planes and Euclidean distances of the C-plane midpoints.

	COD	IAOD	IEOD
Angular difference (°)	$5.76 \pm 5.06^{\dagger}$	4.94 ± 4.24	$9.39 \pm 6.21^{*\ddagger}$
Euclidean distance (mm)	6.46 ± 5.18	5.80 ± 4.15	$8.48 \pm 6.62^{*\ddagger}$

COD = computer-observer difference, IEOD = inter-observer difference; IAOD = intra-observer difference.

[†] Mean is statistically significantly different from that for IEOD.

^{*} Mean is statistically significantly different from that for IAOD.

[‡] Variance is statistically significantly different from that for IAOD.

Table 5.4 COD, IAOD and IEOD and standard deviations of hiatal segmentation metrics: 95th Robust Hausdorff distance, Hausdorff distance, hiatal area difference and absolute hiatal area difference.

	COD	IAOD	IEOD
95th Robust Hausdorff distance (mm)	$7.30 \pm 4.99^{*\ddagger}$	5.10 ± 3.45	$8.48 \pm 6.13^{*\ddagger}$
Hausdorff distance (mm)	$11.26 \pm 5.95^{*\ddagger}$	7.62 ± 3.88	$11.52 \pm 6.60^{*\ddagger}$
Hiatal area difference (cm^2)	$0.98 \pm 3.74^{\ddagger}$	-0.52 ± 2.74	$2.05 \pm 2.86^*$
Absolute hiatal area (cm^2)	2.66 ± 2.78	1.81 ± 2.12	2.53 ± 2.34

COD = computer-observer difference, IEOD = inter-observer difference; IAOD = intra-observer difference.

^{*} Mean is statistically significantly different from that for IAOD.

[‡] Variance is statistically significantly different from that for IAOD.

The quantitative results from the C-plane detection task (Table 5.3) illustrate that COD's bias and variance are not significantly higher than those of IAOD and are significantly smaller than those of IEOD. Moreover, as expected, for all C-plane detection metrics, IEOD was statistically larger than IAOD for both bias and variance.

The quantitative results from the second task (i.e., levator hiatus segmentation) are given in Table 5.4 COD and IEOD both have statistically higher bias and variance than IAOD for the 95th robust HD and the HD. However, the variance and bias of COD did not statistically differ from those of IEOD for the 95th robust HD and HD.

Table 5.4 indicates that for the hiatal area difference, COD's bias and variance were not statistically higher than IEOD's bias and variance. COD's variance was statistically higher than the measured IAOD variance, and IEOD's bias was statistically higher than IAOD's bias. With respect to absolute hiatal area, COD's bias and variance did not statistically differ from those of IAOD or IEOD, and IEOD did not statistically differ from IAOD with respect to bias and variance.

In Table 5.5 are the limits of agreement of COD, IAOD and IEOD for all metrics evaluated in this study. The computer-automated C-plane detection and hiatal segmentation pipeline took

Table 5.5 COD, IAOD and IEOD limits of agreement of all pipeline metrics.

	COD	IAOD	IEOD
Angular difference (°)	{-4.15, 15.68}* {Lower limit, upper limit}.	{-3.37, 13.25}	{-2.78, 21.56}
Euclidean distance (mm)	{-3.69, 16.61}	{-2.33, 13.93}	{-4.50, 21.46}
95th Robust Hausdorff distance (mm)	{-2.48, 17.08}	{-1.66, 11.86}	{-3.53, 20.49}
Hausdorff distance (mm)	{-0.40, 22.92}	{0.02, 15.22}	{-1.42, 24.46}
Hiatal area difference (cm ²)	{-6.35, 8.31}	{-5.89, 4.85}	{-2.63, 6.55}
Absolute hiatal area difference (cm ²)	{2.79, 8.11}	{-2.35, 5.97}	{-2.06, 7.12}

* {Lower limit, upper limit}.

35 ± 4 s, and the manual process took expert 1 on average 2 min and 35 ± 17 s to identify the C-plane and segment the hiatal area on GE software.

5.4 Discussion

This chapter presents a fully automatic hiatal biomarker extraction pipeline from a TPUS volume. Previous studies reported promising results for automatic hiatal segmentation, but required manual determination of the 2-D C-plane [78, 102], which is time consuming and vulnerable to error.

Qualitatively in Fig. 5.3, a minimal difference is seen between the automatically extracted and GS C-planes. The 100th–25th percentile results reveal accurate SP and LAM positioning within the TPUS volume. However, the 0th percentile indicates an inaccurate SP position for the automated task. This particular case was a patient with severe hiatal ballooning. Ballooning may be so severe that the SP is not fully contained within the TPUS. In such circumstances, the operator will watch the SP move during Valsalva in real time and *estimate* the position. Unfortunately, this was not exploited by the proposed method and thus would not perform as well as an expert in these extreme cases.

Table 5.2 indicates that in the Turing test 1 (C-plane position quality), the automated method performed as well as the GS and better than IEOD and IAOD when visually assessed for the true C-plane position. The results from Turing tests 2 and 3 are based on C-plane and hiatal area segmentation quality, respectively. Even though COD achieved a high-quality C-plane, on average the segmentation quality score was noticeably worse. The lower accuracy may be caused by the variety of image qualities and pathologies within the testing dataset; hence, including more pathological training data may improve results. Nevertheless, the average score for the proposed method was > 3 (average) and hence still clinically acceptable.

Table 5.3 indicates that the pipeline performed with a lower bias and variability with respect to expert 1 than did expert 2, and performed similarly to expert 1 in the C-plane detection task. Table 5.5 indicates that the automated C-plane detection task performed within the limits of

agreement of the measured inter-observer difference, highlighting that the first part of the pipeline may reduce the observed bias and variability below the inter-observer variability measured in this study. This may be owing to subtle differences in the techniques used by the experts to identify the C-plane, although the experts were instructed to follow IUGA clinical guidelines using GE 4DView software (GE Healthcare; Zipf, Austria) and the multiplanar technique (described in Chapter 4, Section 4.2.1) [107].

To reduce bias, experts were not managed during the testing phase so that the real-world inter-observer variability of experts working in the same pelvic floor clinic at the same institute, and both with at least 4 y of experience, could be measured. The training data used for the C-plane detection task were generated by expert 1, who also identified the GS C-plane orientations. Thus, it may be assumed the network learned to identify the extreme coordinates more similarly to expert 1 or as the pipeline was based on the extreme coordinate position, expert 1 followed the IUGA guidelines more closely than expert 2, and the C-plane was positioned closer to the extreme coordinates.

This is a common trait for a majority of supervised learning tasks that use CNNs; the network is trained on data from a specific observer and hence will learn to identify features similarly. This trait may be seen as an advantage or disadvantage based on the application. For example, it can learn the behaviour of a specific expert or, in this case, a clinical guideline and can create a personalized automatic workflow that mirrors the expert with the lowest intra-observer variability and most experience, or it may mirror a standardized clinical guideline.

Nevertheless, for other applications (i.e., not guideline related), if desired it can be beneficial to expand the training dataset across several experts, to learn to identify features similarly to several experts rather than one in particular, which makes the CNN more generalisable. This approach was taken for the hiatal area segmentation task of this pipeline. However, a disadvantage of this approach, is that the accuracy can decrease if experts disagree or if one expert deviates with a large error. This approach could lead to no experts being satisfied with the algorithm's result. Therefore, quality control should be conducted to assess the training segmentation data prior to training; with respect to testing this is less important, and a variety of experts with adequate experience may be included to gauge the current clinical world inter-observer variability of a specific task.

Despite the C-plane detection algorithm achieving errors lower than the measured inter-observer variability. In the future, it may be beneficial to re-train the C-plane detection algorithm on a larger training dataset of varied patient pathologies, image qualities and acquisition qualities. To assess whether the performance of the algorithm improves, as one would expect, before clinical implementation.

The pipeline was able to extract the hiatal area to a high level of accuracy. In Table 5.4, the bias and variance of COD were not statistically higher than those of IEOD with respect to hiatal area error metrics (i.e., hiatal area difference and absolute hiatal area). This suggests that the

proposed method extracts hiatal biomarkers as well as experts and, thus, is clinically acceptable. IEOD's bias for hiatal area difference was statistically higher than that of IAOD, indicating that the proposed method may reduce the bias below the measured IEOD. COD's variance for hiatal area difference was statistically higher than IAOD's variance; however, as it was not statistically higher than IEOD's variance, it is still clinically acceptable.

In the literature reported hiatal area differences (bias) of 0.61 cm^2 [78] and 0.23 and 1.1 cm^2 (for U-Net and Dense U-Net respectively) [102]. This study recorded a bias of 0.91 cm^2 . The bias will be higher in this study as the levator hiatus is a 3D structure, and between C-plane positions, the area will differ, thus there is an accumulation of error and not direct comparability. Nevertheless, the bias may be higher because of the 2D levator hiatus segmentation training dataset used, which consists of contrasted post-processed C-planes, whereas the testing dataset is not post-processed. In addition, unlike in the literature, the training dataset was from a data center different from that of the testing dataset, hence, the differences in image quality. To improve results annotated non-post-processed C-planes may be used in training.

The method used in this study was tested on a clinical dataset of patients with a range of anatomical variability and pathological conditions, such as severe hiatal ballooning, levator avulsion and bladder neck hypermobility, as well as patients without pathology. The dataset was even more challenging as up to 81.1% of the patients had pelvic organ prolapse and, thus, a wide range of extreme coordinate movement.

The approach taken in this study uses information extracted from data, the geometry of the patient and clinical guidelines to drive a hybrid approach to extract hiatal dimensions. The proposed method had relatively low errors with a small training dataset, typically rare in deep learning applications. The proposed method performs faster than an expert, however not in real time. For real-time clinical implementation, the pipeline would have to be optimized. The proposed method performs within inter-observer and intra-observer error (for most evaluated metrics); thus, a high level of pelvic floor disorder analysis training may no longer be required for experts to extract high-quality hiatal biomarkers. Furthermore, the output is interpretable to clinicians as the extreme coordinates are well known and recognizable; thus, if the C-plane is incorrect, it is easy to identify the problem (i.e., misplacement of the SP as a result of shadowing).

Clinically, experts commonly acquire a 4D TPUS volume, referred to as a cine loop. Currently the volume of interest (i.e., volume of maximal contraction) is selected manually by the expert. In future work, the aim is to expand this method to localise the volume of interest from the cine loop. Finally, the proposed pipeline will be made interactive to allow operators to adapt the C-plane position and/or the 2D hiatal segmentation. Chapter 6 aims to address the limitations of automatic 2D hiatal segmentation highlighted in this chapter, by proposing an interactive segmentation tool for robust editing control of the levator hiatal boundary.

5.5 Conclusion

Our method was able to extract high-quality C-planes and hiatal area measurements from TPUS volumes without user input. The time taken for hiatal extraction decreased by 120s, saving clinicians time. Furthermore, the automated pipeline reduces error below the inter-observer variability for evaluated metrics within this study.

Chapter 6

Interactive two-dimensional segmentation of the levator hiatus

As described in Chapter 2, Section 2.1, automatic medical image segmentation via CNNs has shown promising results. However, as described in Section 2.2 they may not always be robust enough for clinical use. Sub-optimal segmentation would require clinician's to manually delineate the target object, causing frustration. To address this problem, a novel interactive CNN-based segmentation framework is proposed in this work. The aim is to represent the CNN segmentation contour as B-splines by utilising BEAS (discussed in Chapter 2, Section 2.1.2). The interactive element of the framework allows the user to precisely edit the contour in real-time, and by utilising BEAS it ensures the final contour is smooth and anatomically plausible. This framework was applied to the task of 2D segmentation of the levator hiatus from 2D US images, and compared to the current clinical tools used in pelvic floor disorder clinic (4DView, GE Healthcare; Zipf, Austria). Experimental results show that: 1) the proposed framework is more robust than current ‘state-of-the-art’ CNNs; 2) the perceived workload calculated via the NASA Task Load Index (NASA-TLX) was reduced more than half for the proposed approach in comparison to current clinical tools; and 3) the proposed tool requires at least 13 seconds less user time than the clinical tools, which was *significant* ($p=0.001$).

6.1 Introduction

Medical image segmentation of anatomical structures can be used for disease diagnosis [78]. Manual segmentation requires expertise, time and is prone to error, therefore, automatic methods

This chapter is based on the peer reviewed conference paper published in *Lecture Notes in Computer Science book series (LNIP, volume 12901)* **Williams, H.**, Pedrosa, J., Cattani, L., Housmans, S., Vercauteren, T., Deprest, J., D’hooge, J., (2021). Interactive Segmentation via Deep Learning and B-Spline Explicit Active Surfaces. Medical Image Computing and Computer Assisted Intervention – MICCAI 2021. MICCAI 2021. Lecture Notes in Computer Science(), vol 12901. Springer, Cham.

Chapter 6. Interactive 2D segmentation of the levator hiatus

are desirable. Deep learning-based solutions with CNNs have been extensively explored [185, 186]. However, their impressive average performance has not yet led to wide clinical adoption [187]. Medical images pose serious challenges to automatic methods, as they can be sensitive to small differences between training and testing data, due to such factors as image quality, imaging protocols (i.e., imaging acquisition discrepancies), pathology, and patient variation [188, 189, 158, 149].

Therefore, it is important for clinical impact and acceptance, to be able to recover from a poor result and address the limitations of automatic segmentation. As the clinician remains liable for the measurements obtained for diagnosis, if the automatic method is incorrect, it is the responsibility of the clinician to identify the problem and correct the segmentation. Interactive segmentation with an intuitive mechanism, for smart correction of poor segmentation, may solve these problems, and give liability to the clinician without them having to manually re-segment, which is not time efficient and may cause frustration. This work is motivated to combine ‘state-of-the-art’ CNN segmentation with a user interaction tool, which allows the clinician to view, correct (if needed) and save the desired segmentation.

An extensive range of CNN-based interactive methods have been proposed (as discussed in Chapter 2, Section 2.2) [190], exploiting bounding boxes [152], scribbles [191, 153], extreme points [154] or clicks [155]. These achieved higher accuracy and robustness than their automatic counterparts, however, they can require a high cognitive load and understanding. In addition, the user still relies on the CNN to segment correctly and is not always able to edit the contour precisely, in an adequate and time efficient manner.

In this chapter, an interactive segmentation tool for 2D semantic medical image segmentation is proposed. The tool is composed of three stages. In the first stage, a CNN automatically obtains an initial segmentation, this feeds as an initialisation to an active contour segmentation framework called BEAS [118], which smooths the contour to be more biologically plausible (acting as a post-processing step), and thirdly a novel algorithm which allows the user to interact with the contour in real-time was implemented.

The proposed approach is compared with manual tools used in clinic for 2D segmentation of the levator hiatus in pelvic floor disorder assessment: “Point” and “Trace” both available on the US software 4DView (GE Healthcare; Zipf, Austria), and compared with a ‘state-of-the-art’ scribble-based approach referred to as UGIR [153]. The segmentation methods are evaluated on 30 2D US images, the time taken to segment to a clinically acceptable level, and the perceived workload are measured and compared. The contributions of this work are four-fold: 1) A novel CNN-based interactive framework for 2D segmentation is proposed; 2) the interactive element works in real time and requires less user time and perceived workload than clinical methods and UGIR; 3) the method utilises the BEAS framework to ensure the final contour is more biologically plausible than the CNN segmentation, acting as a novel post-processing method;

and 4) a new energy term is introduced that is dependent on the probability map of the CNN output, which has not been utilised in BEAS before.

6.2 Material and methods

6.2.1 Proposed pipeline

The proposed pipeline is composed of three sequential parts: a 2D CNN which segments the target object (levator hiatus) from the US image (as described in Chapter 5); a BEAS-based post-processing method which smooths the CNN segmentation and represents the segmentation boundary as a B-spline explicit active surface; and a novel algorithm implemented in a graphical user interface (GUI) referred to as "Beyond", that allows the user to adapt the contour in real-time while benefiting from BEAS's active model properties. The framework is shown in Fig. 6.1. The first task of the pipeline, automatically defined the levator hiatus from the US image. This elaborates from Chapter 5 where 2D U-Net was used. The segmentation is fed as input to the following task.

BEAS-based smoothing

The second task utilises the BEAS framework [118]. A 2D version of BEAS is applied to the CNN segmentation after thresholding, to represent the CNN segmentation boundary as B-splines. The concept of BEAS, is to regard the boundary of a target object (i.e., the CNN segmentation) as an explicit function, where one of the coordinates of the boundary is given explicitly as a function of the remaining coordinates. As the contour is a closed 2D object the boundary can be represented in the polar domain, where the contour radius is represented as an explicit function of the polar angle, (i.e., $\rho = \psi(\theta)$). Inspired by Bernard *et al.* [192], the explicit function ψ can

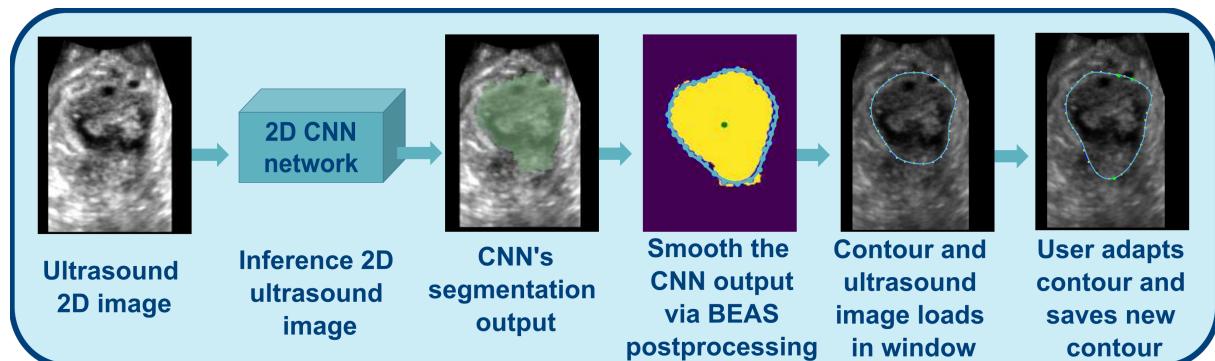


Fig. 6.1 The proposed 2D interactive segmentation pipeline that accepts a 2D US, segments the target object via CNN segmentation, smooths the segmentation and loads it in a window, to allow the user to adapt the contour via BEAS interaction.

be expressed as a linear combination of B-spline basis functions [193, 192, 118],

$$\rho = \sum_{\mathbf{k} \in \mathbb{Z}^{n-1}} c[\mathbf{k}] \beta^d \left(\frac{\theta}{h} - \mathbf{k} \right). \quad (6.1)$$

$\beta^d(\cdot)$ is the uniform symmetric $n-1$ -dimensional B-spline of degree d . $\psi(\theta)$ is separable and built as the product of $n-1$ one-dimensional (1D) B-splines. The knots of the B-splines are located on a regular grid defined on the polar coordinate system, with regular spacing given by h . The B-spline coefficients are gathered in $c[\mathbf{k}]$.

BEAS assumes that all coordinates of the boundary are visible from a fixed origin, which is a good approximation for this structure that tends to have a pear-like convex shape. Before refinement of the BEAS contour, the initial circular contour must be defined by parameters, such as the fixed origin and an initial radius. In this work, they are based on properties of the CNN segmentation output. The origin is defined as the center of mass of the CNN output, and the initial radius, r_c , is the average radius of the CNN segmentation output.

The initial contour can then refine and evolve towards the boundary of the CNN segmentation through the minimisation of a segmentation energy functional. To achieve this, a general localised region-based energy functional for level-set segmentation [120] was used. Barbosa *et al.* [118] adapted these localisation strategies for BEAS in terms of B-spline coefficients, and the expression of the energy gradient is given as,

$$\nabla_c E = \frac{\partial E}{\partial c[k]} = \int_{\Gamma} g(\theta) \beta^d \left(\frac{\theta - hk}{h} \right) d\theta. \quad (6.2)$$

The function $g(\theta)$ represents the features of the object to be segmented and is evaluated over the boundary Γ . In this work, the energy function used was the Localised Yezzi Energy, proposed by Lankton and Tannenbaum *et al.* [120, 194]. This energy depends on the average intensity of the CNN output inside and outside the evolving B-spline contour. The contour evolves to have the maximum separation between them. For Localised Yezzi the feature function is given as,

$$g(\theta) = \frac{(I(\psi(\theta), \theta) - u_{\theta})^2}{A_u} - \frac{(I(\psi(\theta), \theta) - v_{\theta})^2}{A_v}. \quad (6.3)$$

A_u and A_v represent the areas inside and outside of the contour, respectively; and u_{θ} and v_{θ} are the mean intensities inside and outside the evolving contour at the polar angle, θ , respectively. $I(\psi(\theta), \theta)$ corresponds to the image value (i.e., CNN output) at position (ρ, θ) . The Yezzi energy relies on the assumption that the interior and the exterior of the contour have the largest difference in average intensities. This is a good assumption for this work, as the goal is to represent the CNN segmentation output as a smooth B-spline contour. The final B-spline coefficients are saved and used in the following section.

Interaction framework

Finally, the contour formed from the previous step and the corresponding US image are loaded in a window, where the user can interact with the contour. A screenshot of the Beyond GUI can be seen in Fig. 6.2.

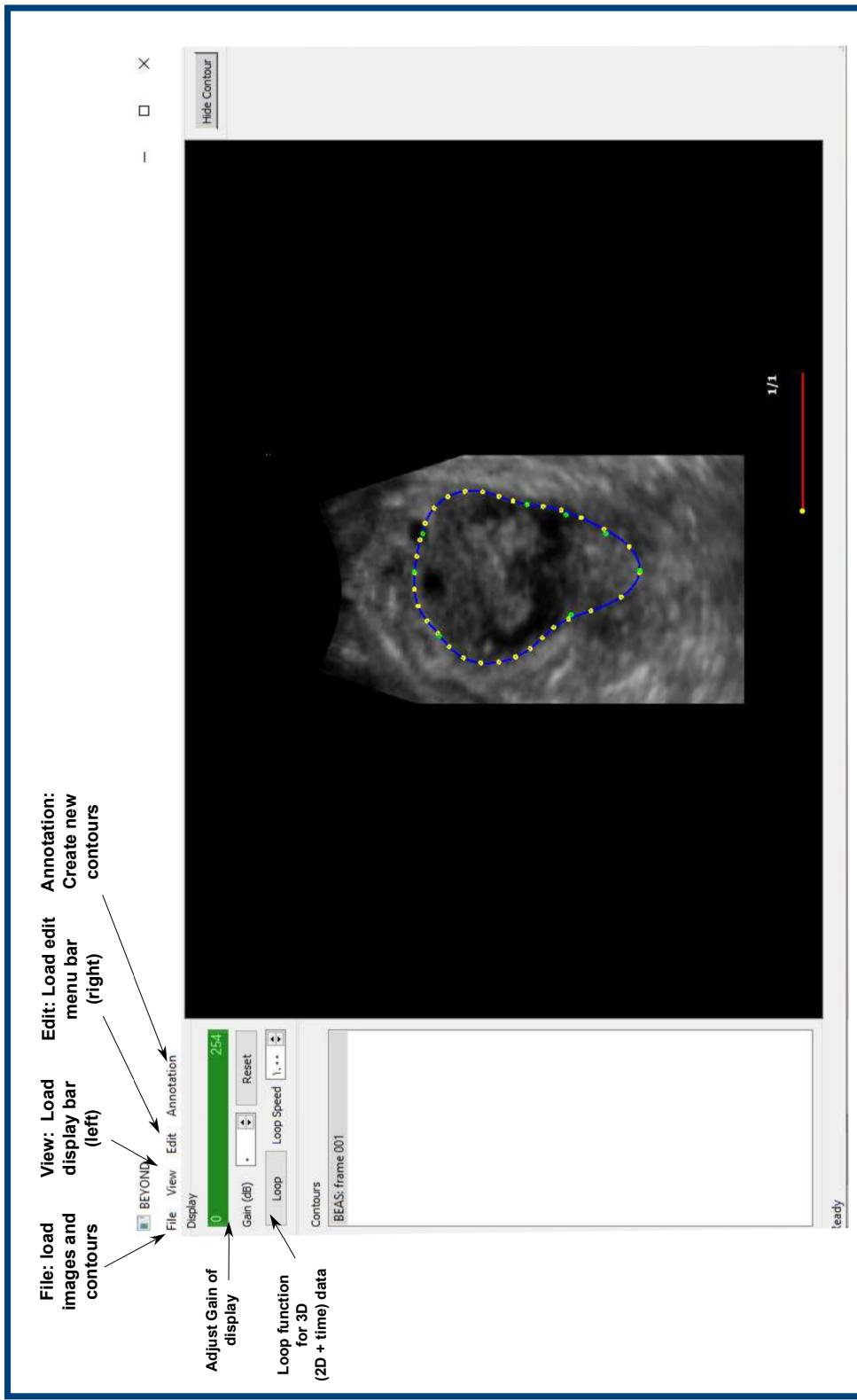


Fig. 6.2 An annotated screenshot of Beyond's GUI, the contour is shown as a blue line, the yellow dots show the knots that define the B-spline contour, and the green dots show the user-defined points.

In the interactive framework, the energy function driving BEAS, is compounded of three energy terms: Localised Yezzi of the US image, E_U , Localised Yezzi of the CNN probability map, E_{CNN} and an interactive energy function, E_i . The total energy is given as:

$$E_{total} = \alpha E_U + \beta E_{CNN} + \gamma E_i, \quad (6.4)$$

where α , β and γ are hyper-parameters. Here, the initialised contour is determined by the evolved B-spline coefficients from the previous section, and it will evolve with each user-interaction to minimise the energy function defined.

Finally E_i , is based on a 2D version of reported work [151], where user-defined coordinates interact with the B-splines. The user can create markers where they want the contour to pass through, these act as anchors attracting the contour. A point introduced by the user can be expressed as $p_{user} = (\rho_{user}, \theta_{user})$. The energy function penalises the parametric distance, D , between the current boundary position and p_{user} at each B-spline knot. Therefore, D is defined as $D = (\rho - \rho_{user})^2$. The energy term driving the contour towards the user-defined points was proposed in 3D by Barbosa *et al.* [151], where its minimisation with respect to B-spline coefficients, $c[\mathbf{k}]$ was demonstrated. E_{user} is defined as:

$$E_{user} = \int_{\Gamma} \delta(\theta - \theta_{user})(\psi(\theta) - \rho_{user})^2 d\theta, \quad (6.5)$$

where $\delta(\theta - \theta_{user})$ corresponds to the Dirac function which is non-zero only at the position $\theta = \theta_{user}$. When multiple user-defined points are present, the sum of the parametric distance between the current contour, Γ , and the user-defined points is used. This evolves the contour towards multiple user points, detailed information can be found in the paper by Barbosa *et al.* [151]. As the computational load is small, there is real-time feedback of the effect the modifications make.

6.2.2 Data collection

Analysis of anonymised, archived US images was retrospective, so no ethics committee approval was required by the institute, at the time of the study. The CNN was trained on a dataset of 444 2D US images from 213 patients, and corresponding ground truth labels of the levator hiatus. The training dataset comprised of two sets of archived clinical images with expert annotations, acquired by several operators. One dataset used for training, was a private dataset supplied by (GE Healthcare; Zipf, Austria) and the second dataset was a private dataset used in previous studies. 400 images were used for training and 44 were used for validation.

The test data included a randomised selection of 30 anonymised 2D US images from 10 symptomatic women assessed at the pelvic floor clinic between March and May, 2019 at UZ Leuven, Belgium. The US images were obtained from TPUS volumes acquired following the

clinical protocol defined by Dietz *et al.* [57] on the Voluson E10 US system (GE Healthcare; Zipf, Austria). The 2D planes that were used to segment and assess the levator hiatus were manually determined by an expert clinician, at rest, during the Valsalva manoeuvre and contraction.

6.2.3 Experimental details

Two clinical experts with over 4 years experience in pelvic floor US, participated in the experiment. They segmented the levator hiatus on 30 2D US images using 4DView Trace (GE Healthcare; Zipf, Austria), 4DView Point (GE Healthcare; Zipf, Austria) and the proposed tool. Prior to the experiment the experts were given a tutorial how to use the new tool. 4DView Trace and Point can be found on the ‘measure - generic area’ function of 4DView (GE Healthcare; Zipf, Austria). In 4DView Trace (GE Healthcare; Zipf, Austria) the contour starts once the user clicks the US image, and it will follow the user’s cursor around the levator hiatus until the user clicks on the US a second time. In 4DView Point (GE Healthcare; Zipf, Austria) the tool will trace the hiatus by the user defining multiple points around the levator hiatus with mouse clicks. The lines that connect the points are straight, therefore, the output segmentation is generalised and not anatomically accurate (i.e., sharp lines). 4DView Trace and 4DView Point (GE Healthcare; Zipf, Austria) may be referred to as Trace and Point respectively in this paper. Uncertainty-Guided Efficient Interactive Refinement (UGIR) utilises an interaction-based level set for fast refinement of segmentations [153], based on scribbles. The same CNN was used as the proposed model and scribbles were created in 3D Slicer [184].

The main aim was to compare the perceived subjective workload of the clinical tools and UGIR against the proposed tool. Therefore, half way through the experiment (after 15 segmentations) and at the end of the experiment, the perceived workload was subjectively evaluated by each expert and for each segmentation technique. To do this the National Aeronautics and Space Administration Task Load Index (NASA-TLX) was used [195].

The NASA-TLX assesses the perceived workload of a task across six workload factors: mental demand, physical demand, temporal demand, performance, effort, and frustration level. The clinician scores each workload factor between 1 to 20. The range of performance is from 1 (good) to 20 (poor), and for all other dimensions the range is from 1 (low) to 20 (high).

As not all factors may be of equal significance to the task assessed, a weighted score can be calculated. In this case, the clinician also provides information based on the subjective importance of each workload factor, referred to as weights. To obtain weights, the clinician conducts a pairwise comparison for every workload factor pair. In each comparison, the factor that subjectively contributes more towards the perceived workload is given a score of one, whereas the other is given a score of zero. Once all 15 pairwise comparisons are complete, the total score (weight) given to each factor ranges from zero to five [196].

The overall perceived workload score between 0 to 100 is calculated, which allows for direct comparison of tasks based on important factors that contribute to the subjective perceived workload. The limitation of the NASA-TLX survey is that currently two factors cannot be rated of equal significance in the pairwise comparisons.

Finally, the time taken for the expert to segment/edit the levator hiatus contour to a clinically accepted level was measured for each segmentation and compared.

6.2.4 Implementation details

The proposed tool was implemented on a Windows desktop with a 24GB NVIDIA Quadro P6000 (NVIDIA; California, United States). The CNN was implemented using NiftyNet [167], training and inference were ran on the GPU. The network architecture was an adaptation of 2D U-Net [91] with half the number of features [32, 64, 128, 256, 512]. An Adam optimiser, ReLU activation function, weighted decay factor of 10^{-5} and batch size of 64 were used. Whitening and histogram normalisation (i.e., when the image was set to have zero-mean and unit variance) were applied to reduce the effects of noise [183]. A Dice loss function was used with a learning rate of 10^{-5} . The data augmentation used were: elastic deformation (deformation sigma = 5, number of control points = 4), random scaling (-20% , $+20\%$), vertical ‘flipping’ and an implementation of *mixup* [175] (as discussed in Chapter 3, Section 3.2). Validation of the network training was performed every 250 epochs and the CNN trained for 12,000 epochs. The CNN model from epoch 10,000 was used at inference, as the validation loss function was lowest. The CNN hyper-parameters were determined based on literature [78] and the performance of the training dataset. BEAS optimisation was ran on the central processing unit (CPU).

In task 2, the size of the neighbourhood used to estimate the local intensity of the image was set to 100 pixels (i.e., $\approx 30mm$), allowing the contour to recover from a bad initialisation. For both tasks the BEAS contour was discretised into 32 points (i.e., knots) along the polar angle direction, causing the scale parameter, h , to be implicitly fixed to 1. The B-spline coefficients, $c(\mathbf{k})$, are gathered in a 1D index array, spanning the polar domain with 32 B-spline coefficients. For interactive BEAS, in (4) $\alpha = 0.5$, $\beta = 0.3$ and $\gamma = 3$, these hyper-parameters were determined by a grid search method and evaluated by assessing the performance on the training dataset. The sensitivity of these hyper-parameters and the ‘ease of interaction’ was low. The size of the neighbourhood used to estimate the local intensity was set to 10 pixels ($\approx 3mm$). This is low to avoid the contour evolving before user-interaction. Otherwise the contour may evolve towards bright regions of the US in order minimise the energy function. The hyper-parameters used for BEAS were determined by a grid search method where the range was guided by literature [118] and evaluated by assessing the performance of the training dataset.

6.3 Results

A video showing the tool being used by an expert can be found in Chapter 11 (as a QR code in Fig. 11.1 and url link). Fig. 6.3 shows examples of the segmentation obtained via the clinical tools and the proposed pipeline. The experts agreed that the proposed tool accomplished a clinical acceptable standard for hiatal diagnosis for almost all 2D US images (29 images), thus the proposed tool achieved a ‘clinical acceptability’ of 97%. However, only 2 and 1 CNN + BEAS post-processing segmentation’s required no editing from expert 1 and 2 respectively, equalling a ‘clinical acceptability’ of 5%. The ‘clinical acceptability’ of the CNN alone was 2%, and the ‘clinical acceptability’ of UGIR was 27%.

Fig. 6.4 shows the time taken for expert 1 and 2 to delineate a segmentation of the levator hiatus to a clinically acceptable standard for diagnosis, using the clinical tools and proposed method. The recorded time of the proposed pipeline does not include CNN inference time, to

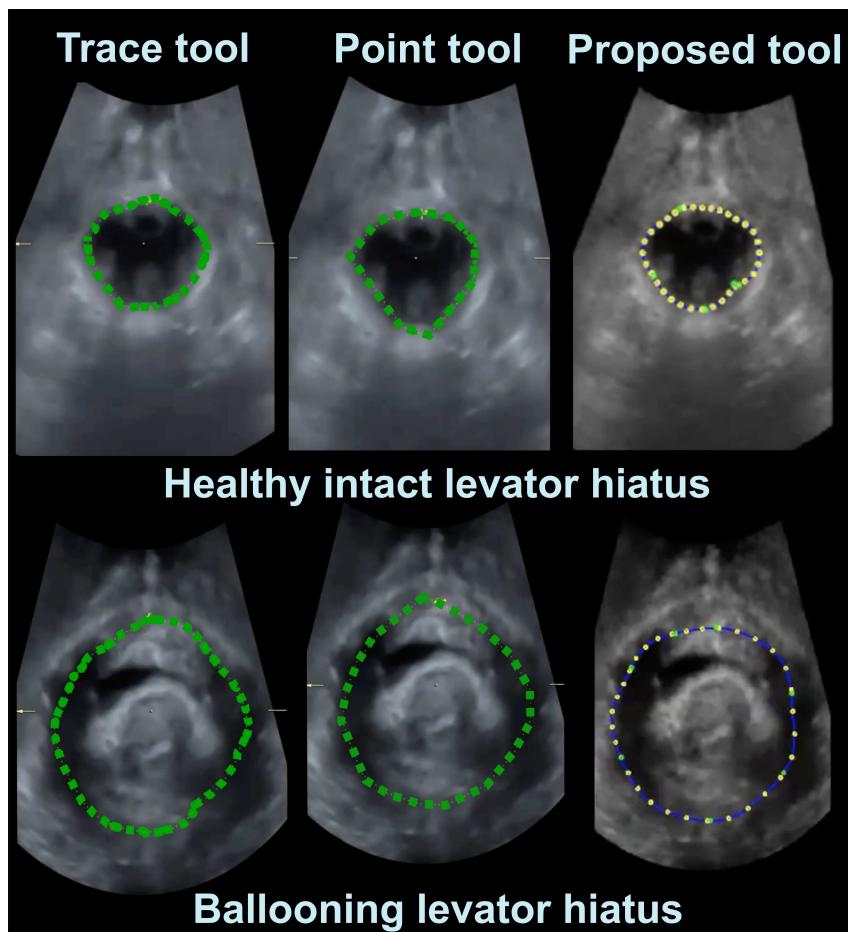


Fig. 6.3 Visual representation of the levator hiatus segmentation obtained with the clinical tools and proposed pipeline, in a healthy patient at contraction and a patient with ballooning hiatus at Valsalva.

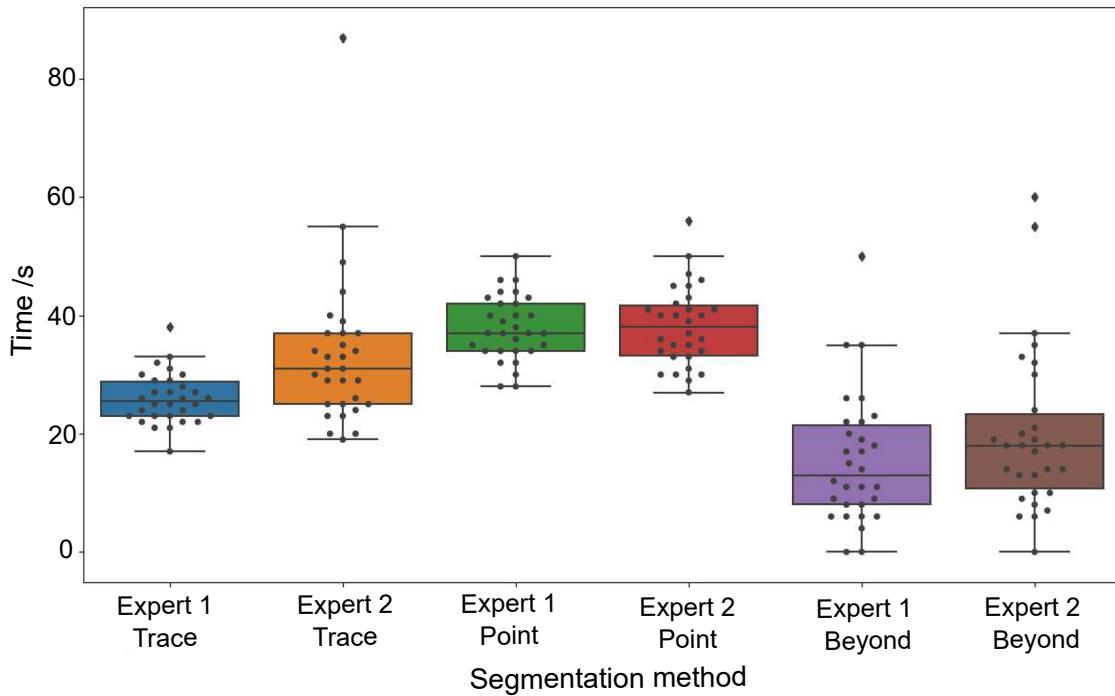


Fig. 6.4 Time taken to delineate the levator hiatus to a clinically acceptable level with 4DView Trace, Point (GE Healthcare; Zipf, Austria) and the proposed tool (Beyond).

be solely dependent on user interaction time and independent of the CNN's performance. The average CNN inference time was 4.90 ± 1.58 seconds. The time taken using the UGIR method was 63.77 ± 31.27 seconds. The time taken to edit the levator hiatus contour with the proposed tool was '*significantly lower*' (paired t-test, $p \leq 0.001$) than the clinical tools and state of the art method, UGIR. The time taken decreased over the experiment, as the combined mean of the first 15 segmentations was 19.33 ± 11.93 seconds and for the final 15 segmentations was 16.13 ± 11.40 seconds. Table 6.1 shows the mean NASA-TLX scores of both experts, and table 6.2 shows the NASA-TLX scores half way through the experiment (attempt 1), and at the end of the experiment (attempt 2). Within the tables the individual weighted sub-scales are reported, as well as the total weighted work-load. The perceived workload was '*significantly lower*' (paired t-test, $p=0.001$) in the proposed tool than the clinical tools perceived by both experts. It is worth noting a lower score for perceived performance in this table corresponds to a better perceived performance. Therefore, the experts found the proposed tool to perform better than the clinical tools and UGIR. Furthermore, the perceived mental workload of the proposed tool improved at the end of the experiment, showing improved performance after exposure to the tool.

Table 6.1 The perceived weighted workload score from the NASA-TLX questionnaire and individual sub-scale scores that contribute to it are shown. A low weighted workload score corresponds to a tool that requires less effort, less frustration, less mental, temporal and physical demands, and a tool that has a higher perceived performance.

NASA-TLX weighted scores	Average			
	Trace	Point	UGIR	Beyond
Effort	13.67	15.50	14.00	6.17
Frustration	12.00	11.00	21.34	2.17
Mental Demand	14.33	10.84	8.67	6.17
Performance	11.00	18.34	28.34	5.34
Physical Demand	6.67	4.67	4.00	3.67
Temporal Demand	0.84	0.00	0.00	0.34
Total workload	58.47	60.34	76.33	23.84

6.4 Discussion

Fig. 6.3 shows visually similar contours for all tools. The proposed method shows more anatomically plausible results than the Point tool. The proposed tool achieved visually ‘clinically acceptable’ results for almost all cases, however, only 2 and 1 CNN segmentations required no editing from expert 1 and 2 respectively. The sub-optimal segmentation was due to a poor US acquisition (that was noted as a clinically unacceptable US image), this reduced the visibility of the levator hiatal boundary, which severely impacted the CNN output. Retrospectively, the radius defining the initial BEAS contour in the second task was increased, and an optimal segmentation was obtained using the proposed tool. Thus, the tool is capable of a ‘clinical acceptability’ score of 100%, with further tuning.

Both experts achieved at least an average improvement of 13 seconds when using the proposed tool. The time measured did not include CNN inference time, to keep it independent to the experiment, and to allow for direct comparison with other segmentation tasks of different CNN architectures. It is assumed with optimisation the CNN inference time would reduce. The proposed tool nonetheless, is still quicker than the clinical tools and UGIR, and it may be assumed with further practice, the time taken would continue to decrease.

Following 30 levator hiatus segmentations, the NASA-TLX questionnaire demonstrated the tool improved perceived performance, reduced effort, frustration, mental and temporal demand. The proposed tool reduced the weighted perceived workload, by 36.50, 34.63 and 52.49 points on the NASA-TLX index scale, for Point, Trace and UGIR tools respectively. The performance improved by 13.00, 5.66, 23.00 points on the NASA-TLX weighted index scale compared to the Point, Trace and UGIR tools respectively. It may be assumed the performance of the Point tool was lower, due to the less anatomically accurate segmentation, shown in Fig. 6.3. Therefore, it may be assumed that this work could improve the clinical workflow. Table 6.2 showed that

Table 6.2 The perceived weighted workload score from the NASA-TLX questionnaire and individual sub-scale scores that contribute to it are shown from mid-experiment (attempt 1) and the end of the experiment (attempt 2). A low weighted workload score corresponds to a tool that requires less effort, less frustration, less mental, temporal and physical demands, and a tool that has a higher perceived performance.

NASA-TLX weighted scores	Attempt 1			Attempt 2		
	Trace	Point	Beyond	Trace	Point	Beyond
Effort	11.00	14.67	7.33	16.33	16.33	5.00
Frustration	12.00	11.00	3.33	12.00	11.00	1.00
Mental Demand	15.33	9.67	5.33	13.33	12.00	7.00
Performance	10.67	15.00	6.00	11.33	21.67	4.67
Physical Demand	7.00	4.00	3.33	6.33	5.33	4.00
Temporal Demand	1.67	0.00	0.00	0.00	0.00	0.67
Total workload	57.60	54.34	25.34	59.33	66.33	22.33

the perceived workload score was lower at the end of the experiment than at mid-experiment, highlighting that with increased exposure, the workload may continue to reduce.

The proposed tool is compounded of a post-processing filter and an interactive algorithm. It can be easily implemented on other 2D segmentation tasks, to improve the segmentation boundary and allow for easy editing of incorrect segmentation. This work is extended to a 3D interactive segmentation pipeline in Chapter 7. Currently, the hyper-parameters used for BEAS (i.e., number of B-splines) requires manual optimisation. In future work, it would be beneficial to automate hyper-parameter selection dependent on the initial 2D segmentation, and compare performance for several segmentation tasks. Automated hyper-parameter selection of certain BEAS hyper-parameters is addressed in Chapter 7.

6.5 Conclusion

To conclude, in this work, a novel CNN-based interactive 2D segmentation tool was proposed. The interactive element works in real-time and requires less user time and perceived workload than current clinical methods, suggesting the proposed work may improve the current clinical workflow. The method utilised the BEAS framework, which ensured the final contour was more biologically plausible than CNN segmentation outputs. This framework can easily be implemented for other 2D segmentation tasks, to make the results more robust while improving the clinical acceptability and giving liability to clinicians.

Chapter 7

Interactive three-dimensional segmentation of the external anal sphincter

This chapter extends work from Chapter 6 and presents a novel 3D interactive segmentation pipeline, to address the problems of automatic CNN segmentation that have been discussed (Chapters 2,6). The proposed novel interactive segmentation framework represents a 3D segmentation from a CNN as a BEAS. BEAS ensures segmentations are smooth in 3D and it increases anatomical plausibility while allowing the user to precisely edit the 3D surface.

We apply this framework to the task of 3D segmentation of the external anal sphincter from 3D TPUS volumes and compare against the current clinical tool used in pelvic floor disorder clinic (4D View VOCAL, GE Healthcare; Zipf, Austria).

Experimental results show that: 1) the proposed framework gives the user explicit control of the surface contour, unlike other ‘state-of-the-art’ interactive segmentation frameworks; 2) the perceived workload calculated via the NASA-TLX index was reduced by 12.67 points for the proposed approach compared to VOCAL; and 3) the proposed tool requires on average 170 seconds less user time than the clinical tools (statistically significant, $p < 0.00001$).

7.1 Introduction

Several interactive segmentation methods now allow the improvement of deep learning-based segmentation by allowing clinicians to refine automatic segmentation based on their anatomical knowledge, reducing workload as compared to manual methods [159, 157, 153] (as discussed in detail in Chapter 2, Section 2.2).

Deep learning-based interactive segmentation shows the value of using expert knowledge to improve automatic segmentation. The main disadvantage of these methods is that they lack

This chapter is based on the journal paper *in preparation*: **Williams, H.**, Pedrosa, J., Asad, M., Cattani, L., Vercauteren, T., Deprest, J., D’hooge, J., DeepBEAS3D: Deep learning and B-spline Explicit Active Surfaces.

Chapter 7. Interactive 3D segmentation of the external anal sphincter

explicit control and manipulation of the 3D surface contour. This is tackled in 2D in [161], however, although a 3D extension exists [197], it is difficult to embed in a deep learning pipeline. Additionally, deep learning-based interactive methods struggle with US due to the inherent speckle appearance of anatomical structures in US (as shown in Chapter 6).

In Chapter 6 [109], a 2D framework that utilised a CNN and BEAS [118] was shown to outperform a ‘state-of-the-art’ method called Uncertainty-Guided Efficient Interactive Refinement (UGIR) [153] for US segmentation. UGIR utilised user-scribbles in a level set method for CNN segmentation refinement. The method proposed in Chapter 6 gave the user real-time explicit control of the 2D contour. However, the method has not been expanded to 3D segmentation and required manually setting several hyper-parameters defining the 2D contour. Finally, the selection method of these hyper-parameters has not been explored before.

Motivated by these limitations, we propose a 3D interactive segmentation tool based on deep learning and BEAS referred to as DeepBEAS3D. Where the user has explicit control of the surface contour and can define specific boundary points by extending the existing 2D method from Chapter 6 [109].

Given a US volume, DeepBEAS3D uses the probability map from a pre-trained CNN to initialise BEAS (as shown in Fig. 7.1). This representation of a probability map as BEAS allows user interaction in real-time and BEAS’s inherit regularising effect eliminates the need for any post-processing steps that may otherwise be needed.

We compare DeepBEAS3D against Virtual Organ Computer-Aided Analysis (VOCAL) (GE Healthcare; Zipf, Austria), a clinical gold standard tool for manual US volumetric segmentation. Our experimental validation utilises 30 3D US volumes of the anal sphincter, where a comparison of the time and perceived workload required (measured using the NASA TLX index [195]) to segment the external anal sphincter to a clinically acceptable standard are compared.

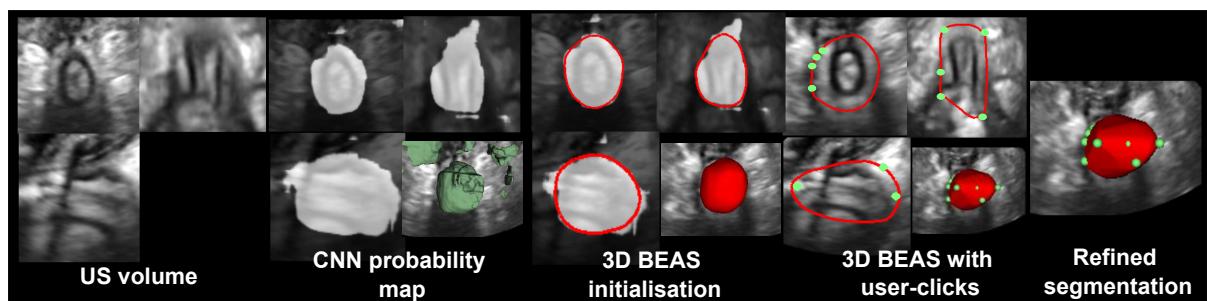


Fig. 7.1 Interactive segmentation using DeepBEAS3D shows input 3D US volume, initial probability map from CNN model (white) and the 3D rendered probability map (green surface), 3D BEAS initialisation to smooth CNN segmentation (red outline), 3D BEAS updates using user interaction (green points) and the BEAS surface (red surface). The framework is implemented in 3D Slicer, where the user can adapt the contour via the interactive energy term.

DeepBEAS3D is implemented as an interactive segmentation module in 3D Slicer which will be publicly made available. DeepBEAS3D addresses several of the limitations from Chapter 6 [109] to improve the performance and the applicability of BEAS.

The contributions of this work are three-fold: 1) A novel extension of a CNN and active learning-based interactive framework for 3D segmentation; 2) a novel application of a hyper-parameter optimisation algorithm for wider applicability; and 3) the work will be made publicly available and includes training documentation as a 3D Slicer module.

7.2 Material and methods

In the pipeline illustrated in Fig. 7.1, a pre-trained 3D U-Net CNN segments the external anal sphincter from a US volume; BEAS evolves to represent the output probability map as an active surface, and the interactive energy adapts the surface in real-time using user interaction clicks.

7.2.1 BEAS-based evolution

Here, we extend the work of Chapter 6 [109] to represent a 3D probability map as a BEAS surface. BEAS represents a coordinate explicitly as a function of the remaining coordinates. Here, the BEAS surface is a complete 3D object and we can represent it in the spherical domain with a defined origin. This is a good approximation for the external anal sphincter which has a convex shape. We represent the surface radius from the origin as an explicit function of both azimuthal and zenithal angles (i.e., $\rho = \psi(\theta, \varphi)$). In [118], the authors expressed ρ , in n -dimensional space, as a linear combination of $n - 1$ dimensional B-spline basis functions of degree, d , $\beta_h^d(\cdot)$ [193, 192]. The mesh points (i.e., knots) of the B-splines are located on a regular grid defined on the spherical coordinate system with regular spacing given by h , where $h = 2^s$ and s is the scale. The number of knots, N_k , depends on the azimuthal mesh size, $mesh_\theta$, and zenithal mesh size, $mesh_\varphi$ ($N_k = mesh_\theta * mesh_\varphi$). Here, we initialise BEAS as a sphere with a fixed origin and radius from the centre of mass and average radius of the probability map, respectively.

The minimisation of an energy criterion, E_{CNN} , defines the evolution of the surface boundary towards the CNN probability map boundary. We define this energy as a localised Yezzi energy [120, 194], which maximises the localised difference between the inside and outside of the boundary. Through minimisation of E_{CNN} , BEAS evolves towards the boundary of the external anal sphincter, as used in Chapter 6 [109]. For more detail, we refer the reader to [118].

7.2.2 Interactive framework

This work is based on the original work [151], where the interactive energy function, E_i , is driven by the parametric radial distance of the user-defined point and the BEAS surface position (i.e., $D = (\rho_u - \rho_m)^2$), where ρ_u is the radius of the user point and ρ_m is the radius of the BEAS surface at the user point. Here, we refer to the explicit function $\rho = \psi(\theta, \varphi) = \psi(\mathbf{x}^*)$ for brevity. E_i is minimised w.r.t. each B-spline coefficient $c[k_i]$ as:

$$\frac{\partial E_i}{\partial c[k_i]} = 2 \iint_{\Gamma} \delta(\mathbf{x}^* - \mathbf{x}_u^*) (\psi(\mathbf{x}^*) - \rho_u) \beta_h^d(\mathbf{x}^* - h\mathbf{k}_i) d\mathbf{x}^*, \quad (7.1)$$

where Γ is the closed surface, \mathbf{x}_u^* corresponds to the point introduced by the user in spherical coordinates. $\delta(\mathbf{x}^* - \mathbf{x}_u^*)$ corresponds to the Dirac delta function, which is non-zero at the position $\mathbf{x}^* = \mathbf{x}_u^* = (\theta_u, \varphi_u)$.

The total energy, E_{total} , driving BEAS on interaction is compounded of three energy terms: E_U (i.e., Localised Yezzi energy of the US volume), E_{CNN} from the previous section and E_i defined above:

$$E_{total} = \alpha E_U + \eta E_{CNN} + \gamma E_i, \quad (7.2)$$

where α , η and γ are hyper-parameters.

7.2.3 BEAS hyper-parameter tuning algorithm

In existing applications of BEAS for segmentation [118, 127], BEAS hyper-parameters have required manual fine-tuning, which limits its application to other structures. These include mesh sizes, N_θ and N_φ and scale, s , affecting the smoothness and the ease of updating from user interaction. To ensure wider applicability, we propose an automated hyper-parameter tuning method which defines an energy term to ensure that BEAS has suitable smoothing and is intuitive regarding user interaction updates (we define this as the *Goldilocks zone*). The ease of user interaction with BEAS is shown in Fig. 7.2. In this work, hyper-parameter tuning is first applied to one training label of the external anal sphincter in a brute force approach. Once complete, the hyper-parameter tuning method is applied to four other training labels in a refined search approach (i.e., with a set scale and smaller mesh size range). The refined search provides a range of *optimal* hyper-parameters and the smallest mesh size is used to reduce the complexity of the mesh representation and avoid over-fitting. We make the approximation that these hyper-parameters will be suitable for other external anal sphincter segmentations, as in literature they are not adjusted from image to image [118, 127].

As shown in Fig. 7.2, we want hyper-parameters (N_θ , N_φ and s) that lead to a BEAS surface that are within what we refer to as the optimal *Goldilocks zone*, where there is a trade-off between

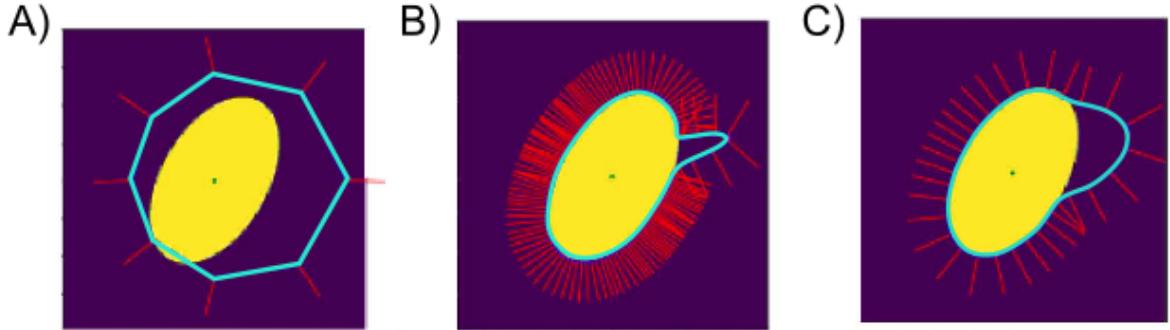


Fig. 7.2 Number of knots vs the ease of user interaction shows, A) too few knots showing large radial movement, B) too many knots showing *spiky* behaviour, C) Goldilocks zone.

neighbourhood (i.e., surface close to the point of interaction) curvature movement and remote region (i.e., surface far from the point of interaction) radial movement.

To measure the ease of updates from interaction, five randomly simulated user-defined points are added in spherical space within a radius of (10% – 20%) from the initial surface. After adding each interaction point, the radial change of the remote region surface and the local Gaussian curvature change of the neighbourhood surface are measured. An energy term, E , is defined, where the minimum value determines the BEAS mesh and scale hyper-parameters that correspond to the optimal *Goldilocks zone*:

$$E = \frac{1}{DSC(I_C, I_S)} + HD(I_C, I_S) + 2(K_S - K_I) + (r_S - r_I), \quad (7.3)$$

where DSC and HD are the Dice score and Hausdorff distance between the CNN segmentation C and the BEAS surface S ; and K and r are the local Gaussian curvature and radial distance [198], and I is the BEAS surface after an interaction point is added. The curvature has a weighting factor of 2, to balance the effect of DSC and HD where a larger mesh size results in a lower HD and higher DSC score, favouring higher mesh sizes.

In the brute force approach, we calculate E for all combinations of mesh size N_θ [6...24], N_ϕ [6...24] and scales [0, 1]. This optimisation takes 50 minutes (i.e., 15 seconds per mesh size and scale combination). We observe the global minimum energy at a mesh size of [12, 16] and scale of 0. Following this; we perform the refined search on four other training labels, with a refined range of mesh size N_θ [12...24], N_ϕ [12...24] and a scale of 0. The refined range is defined by the minimum and maximum mesh size that achieved an error less than 10% higher than the global minimum energy at [12, 16]. The refined search identifies several combinations of optimal hyper-parameters (i.e., [12, 16], [20, 16], [16, 20], [20, 16] and [16, 20]). [12, 16] is chosen as it has the smallest mesh size to reduce model complexity.

7.2.4 Data collection and experimental details

Data collection

Analysis of anonymised US volumes was retrospective, and we had ethical committee approval from the institute. The CNN was trained and tested on a dataset of 115 and 30 US volumes, respectively. The dataset was from 145 patients. External anal sphincter ground truth labels of the training dataset were acquired by an expert operator (with over five years' experience in anal sphincter US imaging and four years' experience in VOCAL). All volumes were acquired at UZ Leuven, Belgium, following the clinical protocol defined by Dietz *et al.* (i.e., transperineal transducer placed on the perineum in the coronal plane) [56] on a Voluson E10 US system (GE Healthcare; Zipf, Austria).

VOCAL protocol

VOCAL segments by rotating the structure around a fixed contour axis and 2D contours are manually delineated in each plane. We defined the axis along the external anal sphincter central axis. In VOCAL, a rotation angle of $[6^\circ, 9^\circ, 15^\circ, 30^\circ]$ could be used, which relates to the rotation steps required for segmentation [199]. 30° in a 180° range was used, meaning after 6 manual delineations a complete 3D segmentation was formed, as VOCAL assumed a symmetrical structure. We chose 30° as in preliminary studies it achieved a lower workload and the fastest time. After manual delineation, VOCAL allows the clinician to make manual contour corrections in other planes of the surface [200].

3D Slicer protocol

A video shows the 3D Slicer protocol and VOCAL protocol in Chapter 11, QR code can be found in Fig. 11.2. If the optimal BEAS hyper-parameters were not known for the structure, the 'Initialise BEAS' button was run on an external anal sphincter training label and the hyper-parameter tuning method described in Section 7.2.3 was used. Following initialisation, the user loaded a dataset of US volumes and CNN probability maps into the data loader. To begin segmentation, the clinician ran the 'Run BEAS' button and an initial BEAS surface was displayed over the US. They visually assessed the output and added user-defined points. The protocol used for the external anal sphincter was that the clinician would assess the axial and midsagittal planes first to ensure the total length and width of the external anal sphincter were included. After, the clinician would assess coronal plane slices and make corrections if necessary.

Implementation details

DeepBEAS3D was implemented on a 24GB NVIDIA Quadro P6000 on a Windows desktop. The 3D U-Net [92] was implemented using MONAI [201] (a PyTorch-based framework for

deep learning in healthcare imaging) and training was performed using Dice loss and a Nesterov optimiser (batch size = 3, momentum = 0.99, learning rate = 0.001, learning rate decay = 0.9 and weight decay = $3 \cdot 10^{-5}$) for 1000 epochs. 100, 15 and 30 US volumes were used for training, validation and testing respectively. Data augmentation included random scaling (-30% , $+40\%$), x-axis rotation (-45° , $+45^\circ$) and Gaussian noise. Validation of the network training was performed every 100 epochs and early stopping based on the highest Dice similarity coefficient at validation was employed, to select the CNN model from epoch 300 for inference. BEAS was run on an Intel(R) Core(TM) i9-7900X CPU @ 3.30GHz, the size of the neighbourhood used to estimate the initial probability map was 100 voxels (i.e., $\approx 30mm$), allowing recovery from a poor initialisation. Optimisation determined BEAS to be discretised into 12 knots along the azimuthal angle direction and 16 knots in the zenithal angle direction, with spacing, h , set to 1. For interaction, in (7.2) $\alpha = 1$, $\eta = 0.3$ and $\gamma = 1$ and the neighbourhood size was set to 10 voxels ($\approx 3mm$), based on Chapter 6 [109].

7.3 Results

Fig. 7.3 shows a segmentation using VOCAL and DeepBEAS3D. The clinician agreed DeepBEAS3D accomplished a clinically acceptable standard for all US volumes and achieved a ‘clinical acceptability’ of 100%. However, only 2 CNN and CNN + BEAS initialised (i.e., DeepBEAS3D without user interaction) segmentations required no editing and achieved a ‘clinical acceptability’ of 7%.

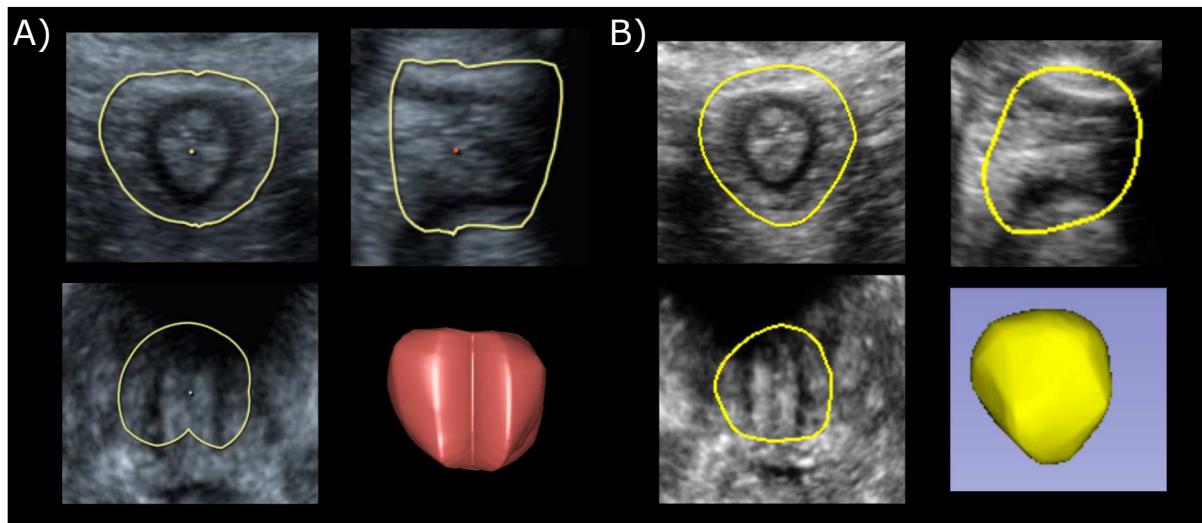


Fig. 7.3 Visualisation of segmentations obtained with A) VOCAL and B) DeepBEAS3D showing coronal (top left), midsagittal (top right) and axial (bottom left) planes, and 3D reconstruction (bottom right).

Chapter 7. Interactive 3D segmentation of the external anal sphincter

Table 7.1 The perceived weighted workload score from the NASA-TLX questionnaire and subscale scores for VOCAL and DeepBEAS3D. A low score corresponds to a tool that required less effort, less frustration, less mental, temporal and physical demands, and had a higher perceived performance.

NASA-TLX weighted scores	Average	
	VOCAL	DeepBEAS3D
Effort	16.00	10.00
Frustration	12.00	10.00
Mental Demand	6.67	6.67
Performance	6.67	3.33
Physical Demand	1.33	0.00
Temporal Demand	0.00	0.00
Total workload	42.67	30.00

Fig. 7.4 shows the time taken to delineate the structure. The time taken by the DeepBEAS3D

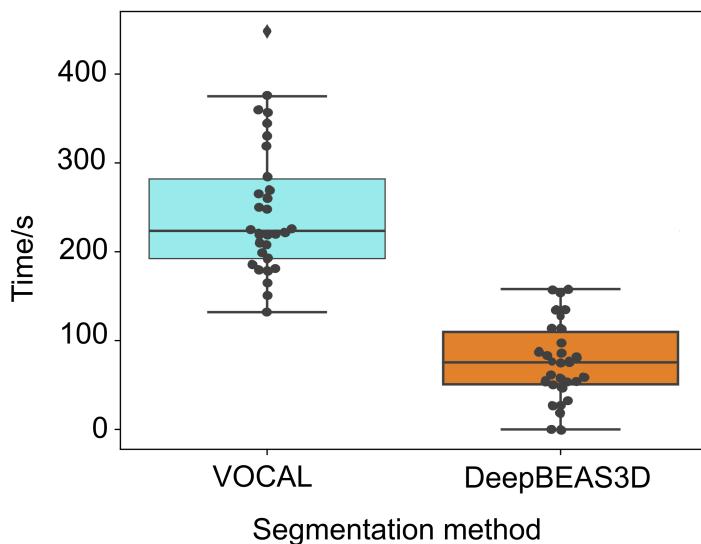


Fig. 7.4 Time taken to delineate the external anal sphincter to a clinically acceptable level with VOCAL and DeepBEAS3D.

does not include CNN inference time, to be independent of the CNN's performance. Here, the average CNN inference time was 12 seconds and the total time taken to delineate the external anal sphincter with DeepBEAS3D was '*statistically significantly lower*' (paired t-test, $p < 0.00001$) than VOCAL. Table 7.1 shows the mean NASA-TLX scores for segmentation.

7.4 Discussion

Fig. 7.3 shows visually similar delineations for both methods. The top right image shows a slight discrepancy between DeepBEAS3D and VOCAL, however, this was due to intra-observer variability, rather than poor performance of DeepBEAS3D. The other visual difference is caused by the smoothing nature of BEAS, which creates a more biologically plausible structure (i.e., DeepBEAS3D does not include the sharp feature shown in the bottom left image in Fig. 7.3 A). Sub-optimal CNN segmentations were due to poor US acquisition, external anal sphincter tearing, US shadowing (shown in Fig. 7.3), which reduced the visibility of the external anal sphincter. In one case, the BEAS initialised mesh reduced the quality of the segmentation, as it had over-smoothed the external anal sphincter and was subsequently corrected with interaction.

The proposed BEAS optimisation algorithm was applied to five labels of the external anal sphincter, which showed a range of hyper-parameters and we chose the smallest mesh size to reduce the complexity of the mesh representation. If the user required editing of finer details in the structure (i.e., with a larger mesh size), the hyper-parameters could be selected from the range identified in Section 7.2. In addition, if desired, the user could tailor the hyper-parameters to the specific CNN’s output shape, guided by the range in Section 7.2. During optimisation, we added only five interactive points to reduce the running time. However, the user could increase the number of points depending on their computer specification. DeepBEAS3D and BEAS improved the CNN segmentation by removing multiple components of smaller volume, smoothing the surface, and removing non-anatomical holes, solving common issues of voxel-wise CNN segmentation [202]. DeepBEAS3D achieved an average 170 seconds improvement compared to VOCAL and interaction time may further reduce with increased user exposure. As we have seen in our study, the time to delineate using DeepBEAS3D decreased from 94 seconds for the first 15 volumes to 60 seconds for the last 15.

Table 7.1 shows that the expert perceived DeepBEAS3D to have a lower workload as compared to VOCAL. In particular, DeepBEAS3D reduced the perceived effort, possibly because, in most cases, an adequate initial segmentation is presented. Therefore, only slight editing is required, whereas VOCAL requires manual 2D delineation and secondary manual editing of the delineated surface.

The main advantage of DeepBEAS3D to other interactive segmentation methods is that it gives the user explicit control of the surface contour, without relying on a CNN for refinement, meaning specific points can be defined as the surface boundary and updates are applied in an efficient CPU-based implementation. This could be utilised by other 3D segmentation tasks once the 3D Slicer module is publicly available. The hyper-parameter tuning algorithm allows for wider adoption, as it does not require manual optimisation. This is the first work to apply such an algorithm for tuning BEAS hyper-parameters for the ease of

interaction and segmentation accuracy. We believe it will increase the applicability and usability of BEAS and DeepBEAS3D to other medical imaging problems.

The selected mesh hyper-parameters favoured little remote region radial movement of BEAS while reducing spiky behaviour close to the point of interaction, which can be common when interacting with splines. DeepBEAS3D segmentations were more anatomically plausible as they inferred consistency across slices of the volume. In future work, it could be beneficial to show DeepBEAS3D working for several structures and modalities and use user editing in active learning. It would be beneficial to refine the proposed BEAS surface hyper-parameter tuning algorithm by calculating the combined average energy of several external anal sphincter training labels to avoid BEAS under or over-fitting. We also aim to use hyper-parameter optimisation to determine hyper-parameters for equation (7.2).

7.5 Conclusion

In this work, a novel 3D interactive segmentation method, DeepBEAS3D, was proposed. DeepBEAS3D guides the segmentation through the radial distance between the user-defined point and the mesh surface. This framework gives the user explicit control of the surface contour, without relying on a CNN for refinement. DeepBEAS3D requires less user-time and has a lower perceived workload than VOCAL, suggesting it could benefit the clinical workflow. BEAS ensured the segmentation was a single component of maximum volume, smooth and complete. Hence, it was more biologically plausible than the CNN segmentation alone. Finally, the 3D Slicer module and the hyper-parameter training algorithm ensure that other segmentation tasks can benefit from DeepBEAS3D.

Chapter 8

Automatic extraction of a tomographic ultrasound imaging sequence from a transperineal ultrasound volume for analysis of the anal sphincter complex

8.1 Introduction

To assess the integrity of the whole length of the anal sphincter from 3D TPUS data, sonographers first extract a TUI sequence from the TPUS recording as described in Chapter 1, Section 1.2.5. TUI sequences consist of eight equally spaced and properly oriented 2D coronal-view slices of the anal sphincter complex. TUI sequences are visually assessed by a sonographer to diagnose anal sphincter injury. Obtaining TUI sequences is performed manually in clinical practice, which is labour-intensive and requires expert knowledge of pelvic floor anatomy. To the best of our knowledge, this work is the first to report an automatic method to aid this medical imaging acquisition task. We propose a novel CNN approach for the automatic extraction of the TUI sequences from a TPUS. The method utilises a CNN to segment the external anal sphincter, and the desired TUI sequences are subsequently extracted after several automatic post-processing steps.

This chapter is based on the peer reviewed conference paper published in *Lecture Notes in Computer Science book series (LNIP, volume 12967)* **Williams, H.**, Cattani, L., Vercauteren, T., Deprest, J., D’hooge, J., (2021). Automatic Tomographic Ultrasound Imaging Sequence Extraction of the Anal Sphincter. In: Noble, J.A., Aylward, S., Grimwood, A., Min, Z., Lee, S.L., Hu, Y. (eds) Simplifying Medical Ultrasound. ASMUS 2021. Lecture Notes in Computer Science(), vol 12967. Springer, Cham.

The proposed method is evaluated on 30 TPUS recordings and compared against manually acquired gold standard TUI sequences. One expert evaluated the quality of the automatically detected TUI sequences in terms of their clinical acceptability for diagnosis. The automatic method performs with an overall clinical acceptability of 90.00%. The method reduces the time required to extract the anal sphincter complex TUI sequence of a TPUS by 52.36 seconds and may reduce the need for high-level expertise in anorectal dysfunction analysis.

Anal sphincter integrity (or injury) can be assessed with exo-anal (TPUS or introital) or with endo-anal US. Endo-anal is more intrusive, and TPUS showed a substantial correlation with exo-anal with high sensitivity for anal sphincter complex evaluation [203]. TPUS has shown to have similar image quality to introital with lower inter-observer variability [204, 203, 58]. Therefore, TPUS was used in this study, further details can be found in literature [203, 58, 56].

Within clinical assessment, sonographers use TUI sequences of the anal sphincter complex to visually assess the integrity of the entire anal sphincter [205, 56]. TUI sequences consist of eight equally spaced and properly oriented 2D coronal view slices of the anal sphincter complex. Manual extraction of TUI sequences from a TPUS recording is labour intensive and recognised as a highly skilled task, as the sonographer must manually manipulate a TPUS recording to locate predetermined locations, based on the cranial termination of the external anal sphincter and the caudal termination of the internal anal sphincter [203, 58, 56], as shown in Fig. 8.1. The quality of TUI extraction is heavily dependent on the sonographer's skill, and significant inter-observer variability may lead to, in extreme cases, misdiagnosis.

Therefore, we aim to automatically extract the TUI sequences from a TPUS recording, to address the limitations above. In this work, the sonographer would only need to acquire a TPUS recording following a standard acquisition, (i.e., the transperineal probe is placed at the opening to the vagina and perpendicularly to the anal canal- as shown in Chapter 1, Section 1.2.4) [203]. Our solution aims to speed up assessment for skilled sonographers, and potentially allow non-experts to perform these assessments.

We briefly describe our work in the context of related literature that has proposed automated image analysis of pelvic floor structures, such as the levator hiatus [78, 102, 100] and the puborectalis muscle [104]. Automatic assessment of the levator hiatus [78, 102] utilised CNNs and active shape models [100], and performed within inter-observer variability. In other work, an automatic clinical solution was presented for the extraction of a plane of interest used in PFD assessment (as described in Chapter 4.) [107]. The chapter utilised CNN landmark regression, and performed within inter-observer variability, while reducing the time required for assessment by 100 seconds.

We believe the work presented in this chapter is of clinical impact, due to the difficult nature of manipulating TPUS recordings of the anal sphincter, the lack of current automation of TUI extraction, and the expertise required by sonographers. In this chapter, we describe to the authors' knowledge the first automatic anal sphincter TUI sequence extraction solution. The proposed

8.2 Material and methods

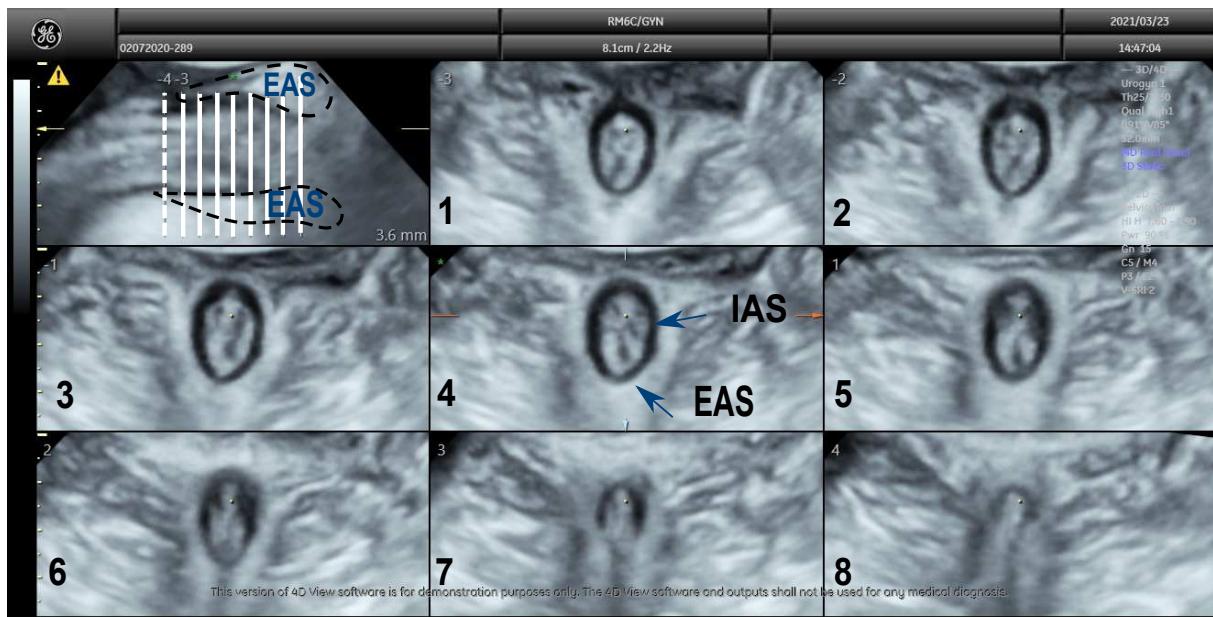


Fig. 8.1 TUI sequence of a normal anal sphincter. The top left image shows the midsagittal plane with the external anal sphincter annotated; the eight other images represent coronal slices through the anal canal. The locations of the slices are given by the vertical lines in the midsagittal plane. Slice 1 is the non-dashed vertical line on the left; slice 8 is at the right. The arrows show the location of the external anal sphincter (EAS) and internal anal sphincter (IAS) within a coronal view plane.

solution locates the external anal sphincter and extracts eight equi-distant 2D images of the anal sphincter in the coronal-view, comparable to manually acquired TUI sequences. This work utilises the advances of CNN segmentation and is evaluated on 30 TPUS recordings. The clinical acceptability and time taken are recorded and compared to an expert sonographer. We believe a fully automatic TUI extraction solution may save clinicians time to allow more focus on patient care and treatment planning.

8.2 Material and methods

During urogynaecological US examination, sonographers aim to evaluate sphincter integrity based on the sonographic appearance of the external anal sphincter and internal anal sphincter. The sonographer acquires a TPUS recording at approximately 60° aperture and 70° acquisition angle with a 3D convex transducer, when possible during pelvic floor muscle contraction. The TUI sequences are identified in post-processing steps. On the extracted TUI sequences, the sonographer assessed external anal sphincter and internal anal sphincter integrity, and if present measured the degree of tear in the external anal sphincter and in the internal anal sphincter which corresponds to the internationally accepted clinical classification [64]. Before describing the method in detail, we first describe the acquisition protocol.

8.2.1 Acquisition protocol

All data was acquired with a Voluson E10 BT16 ultrasound system (GE Healthcare; Zipf, Austria) equipped with a 3D 4-8 MHz convex probe placed transperineally with an average voxel resolution of 0.3 mm by 0.3 mm by 0.3 mm. For testing, a total 30 3D TPUS recordings were acquired. Volumes covering the entire length of the external anal sphincter were obtained and post-processed offline on a desktop computer using 4D View Software (GE Healthcare; Zipf, Austria) according to the international practice parameter [177].

8.2.2 The proposed pipeline

The proposed method is shown in Fig. 8.2. Firstly, the external anal sphincter was segmented from a TPUS recording, the centre of mass, X_{cm} , was determined and the corresponding mid-sagittal plane extracted. Four parallel planes were extracted and an averaged external anal sphincter segmentation was formed. The principal axes of rotation of the averaged segmentation was identified and a rotation matrix was formed. The TPUS was then rotated to ensure the anal sphincter was parallel to the coordinate axes, and eight equi-distant slices of the external anal sphincter in the coronal view were extracted.

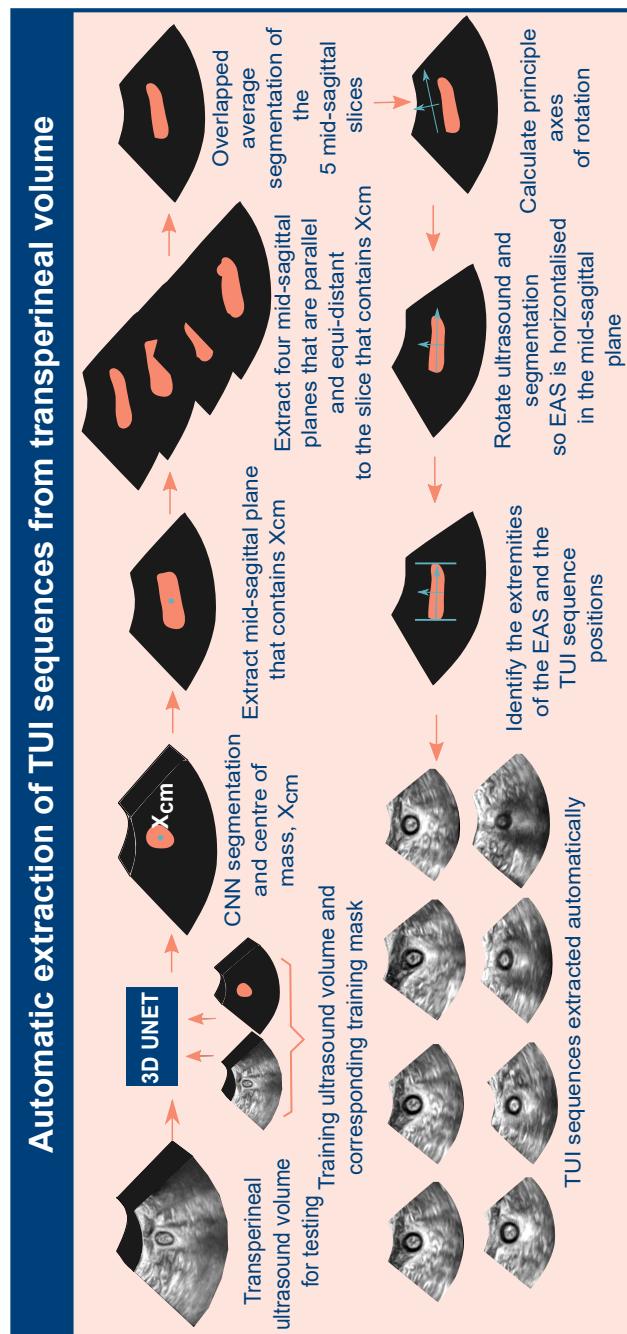


Fig. 8.2 Proposed pipeline of the automatic TUI extraction algorithm (EAS=external anal sphincter).

3D external anal sphincter segmentation

Firstly the external anal sphincter was automatically segmented. This was achieved by utilising a CNN, which accepted a TPUS as input and outputted a 3D voxel-wise segmentation of the external anal sphincter. The architecture used was 3D U-Net [92], and advanced data augmentation was used including an adaptation of the original *mixup* [175], where three images and their labels were linearly combined (as described in Chapter 3, Section 3.2).

Rotation of the TPUS recording

During manual acquisition the sonographer may need to rotate the TPUS to horizontalise the anal canal in the mid-sagittal plane. This ensures the axes of rotation of the sphincter lay along the coordinate axes. Here, we describe how the rotation matrix, R, was formed in order to automate this task. Firstly, X_{cm} of the 3D segmentation was identified, and the mid-sagittal plane of the segmentation which contained X_{cm} was extracted. The mid-sagittal plane is given by the x and y directions of the volume data, and is dependent on a standard acquisition protocol used within clinic (i.e., the probe placed at the entrance of the vagina perpendicular to the anal canal). Several equi-distant parallel planes to the mid-sagittal plane containing, X_{cm} were extracted and multiplied together to produce an averaged 2D external anal sphincter segmentation, based on the common overlap (i.e., common voxel values were equal to 1 and uncommon voxel values were equal to 0). The mid-sagittal planes used contained the coordinate X_{cm} , $X_{cm} \pm 1.5mm$ and $X_{cm} \pm 3mm$.

Principle component analysis (PCA) was used to identify the eigenvectors, \vec{v}_{av} , describing the principle axes of rotation of the averaged 2D external anal sphincter segmentation within the mid-sagittal view. PCA was only applied to the mid-sagittal view rather than the total 3D segmentation, to follow aspects of the clinical procedure. PCA was applied to the averaged 2D external anal sphincter segmentation, rather than the mid-sagittal plane containing X_{cm} to make the method more robust, and reduce the risk of incorrect rotation due to poor segmentation of the external anal sphincter within one mid-sagittal plane. To form the rotation matrix, R, the inverse of the averaged eigenvector, \vec{v}_{av}^{-1} was computed. The rotation matrix, R, was defined as:

$$R = \|R_x\| \|R_y\| \|R_z\| = \begin{vmatrix} 1 & 0 & 0 \\ 0 & \vec{v}_{av_{xx}}^{-1} & \vec{v}_{av_{yx}}^{-1} \\ 0 & \vec{v}_{av_{xy}}^{-1} & \vec{v}_{av_{yy}}^{-1} \end{vmatrix}. \quad (8.1)$$

Where $\vec{v}_{av_{xx}}^{-1}$ and $\vec{v}_{av_{xy}}^{-1}$ define the x and y component respectively of the eigenvector along the length of the anal canal, and $\vec{v}_{av_{yx}}^{-1}$ and $\vec{v}_{av_{yy}}^{-1}$ define the x and y component respectively of the eigenvector along the width of the anal canal. The TPUS and CNN segmentation were rotated in preparation for TUI extraction.

Unfortunately, occasionally the rotation angle determined as above may be too severe, due to a non cylindrical external anal sphincter segmentation. Therefore, before TUI extraction occurred an automated quality control process was performed. The ratio between the largest and smallest eigenvector component was calculated, and when the ratio was smaller than a pre-defined threshold, the rotation matrix was set to identity, and the TPUS and segmentation were not rotated. The pre-defined threshold was 2.11 and it was determined in preliminary studies, based on the relationship between the eigenvector component ratio and the rotational acceptability score. In detail, a sample of 10 incorrectly rotated TPUS recordings were used, the mean ratio and standard deviation were calculated, and the threshold was set to the upper bound of the 95% confidence limit.

Identification of extreme points

To extract TUI sequences, the extreme points as shown in Fig. 8.2 were identified. X_{cm} of the rotated CNN segmentation was calculated and the rotated mid-sagittal plane containing X_{cm} was extracted. After rotation the major axes of the external anal sphincter were parallel to the coordinate axes, and the first and last coordinate along the y axis of the external anal sphincter segmentation were extracted. The total length of the external anal sphincter was calculated and divided by 9 to determine the slice separation (i.e., distance between 2D slices in coronal-view of the anal sphincter complex), thus 8 slices were extracted excluding the first and last coordinate position of the external anal sphincter. This reduced the risk of selecting a plane too far from the optimal position due to poor segmentation.

During examination the spacing between TUI sequences should be larger than 2mm, thus a quality check was performed prior to extraction, and when the slice separation was smaller than 2mm the total length of the external anal sphincter was divided by 7 and the TUI sequences included the extremities of the external anal sphincter segmentation. If the slice separation was still smaller than 2mm the algorithm outputted the TUI sequences and a notification that the length of the detected external anal sphincter may be insufficient or abnormal.

8.2.3 Data collection

Analysis of anonymised, archived US images was retrospective, so ethics committee approval was not required by Belgian law at the time of this work. The TPUS recordings were acquired at the pelvic floor clinic at UZ Leuven, Belgium between February and November 2020. The data was separated into training and test sets such that each patient was in one set only. In total 148 3D TPUS recordings were used; 94 for training, 24 for validation and 30 for testing.

An expert sonographer with over four years' of experience in US PFD assessment, manually extracted TUI sequences of the anal sphincter for clinical diagnosis using 4D View software (GE Healthcare; Zipf, Austria). The same expert manually segmented the external anal sphincter

complex with rotations of 30° , using the volume analysis application VOCAL from 4D View Software (GE Healthcare; Zipf, Austria) from the 3D TPUS recordings, these were used as ground truth labels for training.

8.2.4 Evaluation methodology

The expert identified the TUI sequences in all TPUS recordings manually via the clinical protocol using 4D View software (GE Healthcare; Zipf, Austria). These TUI sequences are defined as gold standard and were visually compared to the automatically detected TUI sequences. To assess the performance of the proposed method, the expert was asked to rate the overall performance of the automatically detected TUI sequences for each TPUS volume: visually “clinically acceptable” or “unacceptable” for clinical diagnosis. They were also asked to rate the rotation of the anal canal within the mid-sagittal plane and the quality of each TUI slice as either “clinically acceptable” or “unacceptable”. The slice rating was dependent on the automatic slice being visually similar to the gold standard and of use for clinical diagnosis (i.e., showing the same pathology if present). The time taken for the automatic pipeline to identify the TUI sequences was compared to the time taken by the expert to manually extract the TUI sequence on a new subset of 19 TPUS recordings acquired within clinic, via the clinical protocol using 4D View software (GE Healthcare; Zipf, Austria). In addition the slice separation of TUI sequences determined automatically and manually were compared.

8.2.5 Implementation details

The proposed tool was implemented on a Windows desktop with a 24GB NVIDIA Quadro P6000 (NVIDIA; California, United States). The CNN was implemented using NiftyNet [167], training and inference were ran on the GPU. The CNN architecture was 3D U-Net [92], an Adam optimiser, ReLU activation function, weighted decay factor of 10^{-5} , Dice loss function with a learning rate of 0.0001 and batch size of 2 were used. The data augmentation used were: elastic deformation (deformation sigma = 5, number of control points = 4), random scaling (-20% , $+20\%$) and an implementation of *mixup* [175] (Chapter 3, Section 3.2). Validation of the CNN training was performed every 200 epochs and it trained for 6000 epochs. The model from epoch 3200 was used at inference, as the validation loss function was lowest.

8.3 Results

Qualitative results of the automatically extracted TUI sequences compared to the gold standard are shown in Fig. 8.3. Fig. 8.3C shows the worst performing result (based on overall, slice and rotational acceptability), Fig. 8.3B the average performing result, and Fig. 8.3A the best

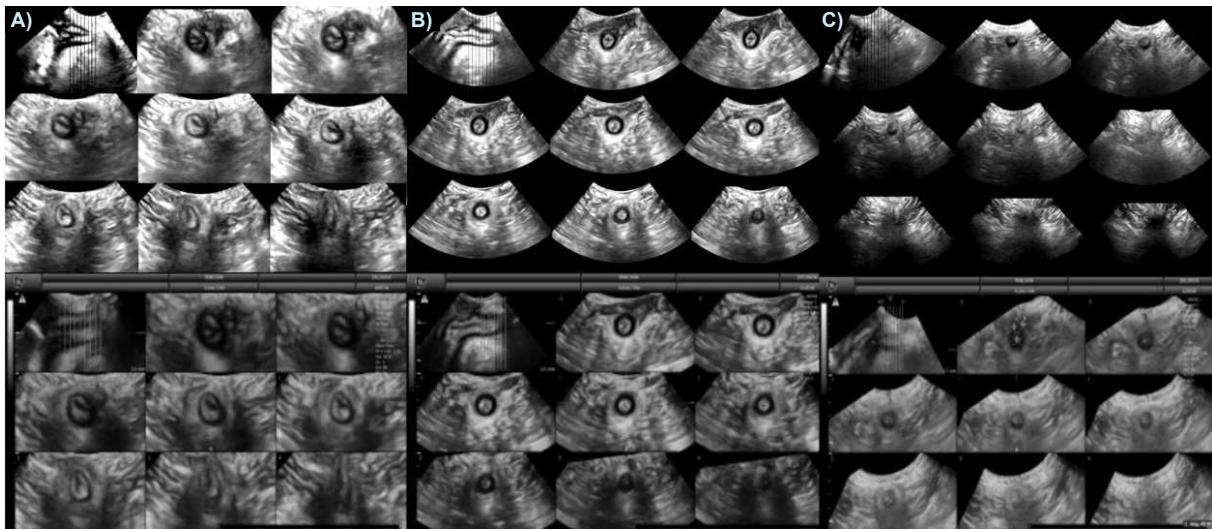


Fig. 8.3 TUI extraction results, A) the best performing result, B) the average performing result and C) the worst performing result (based on overall, slice and rotational acceptability). The corresponding gold-standard TUI sequence for each result are shown in the second row.

performing result. The overall clinical acceptability, the rotation performance, the time taken, and the average slice separation are shown in Table 8.1. Table 8.2 shows the clinical acceptability scores of the automatic method for each TUI slice.

Table 8.1 Overall and rotational clinical acceptability, time taken and slice separation.

Method	Overall clinical acceptability (%)	Rotation clinical acceptability (%)	Time (s)	Slice separation (mm)
Automatic method	90.00	93.33	8.64 ± 0.17	2.84 ± 0.50
Manual	100	100	61.00 ± 13.74	2.81 ± 0.37

Table 8.2 TUI sequence slice number and corresponding clinical acceptability.

Slice number	1	2	3	4	5	6	7	8
Clinical acceptability (%)	70.00	86.67	90.00	93.33	96.67	96.67	96.67	90.00

8.4 Discussion

The study presents to the-authors-knowledge the first automatic TUI sequence extraction pipeline from a TPUS recording. Qualitatively in Fig. 8.3 there is minimal visual difference between the automatically and manually extracted TUI sequences for the average and best performing result and they both show the same clinical diagnosis. The worst-performing result was clinically unacceptable for all slices, due to the incorrect rotation and location of the TUI sequence.

Incorrect rotation was due a non cylindrical external anal sphincter with a ratio of 3.45, which was larger than the pre-defined threshold. Incorrect rotation meant the TUI sequences did not intersect the anal canal perpendicularly, and incorrect location of the first TUI sequence resulted in sequences that were not clinically suitable.

The average-performing result had a rotation that was clinically acceptable, however, the position of the final slice was not optimal as it contained part of the internal anal sphincter unlike the gold standard. This does not impact the overall clinical acceptability as the same diagnosis was made. The best performing result, was rated as clinically acceptable for all slices and for the rotation of the TPUS volume. The visual difference between the automatic result and the gold standard is negligible and any differences are due to the post-processing of 4D View (GE Healthcare; Zipf, Austria).

The proposed method was 52.36 seconds faster than the clinical expert, which was significant ($p < 0.001$), and the variance of time taken decreased significantly ($p < 0.001$). The average slice separation of the proposed method was not statistically higher than the manually acquired slice separation ($p = 0.265$). The overall clinical acceptability of the proposed method was 90.00%, on average 7.16 TUI sequences out of 8 were marked as clinically acceptable and the rotation scored a clinical acceptability of 93.33%. 11 TPUS volumes were not rotated as the quality control process detected a ratio of eigenvector components smaller than or equal to 2.11. Slice 1 and 8 describe the extremities of the external anal sphincter, and are the most dependent on the segmentation. Table 8.2 shows slice 1 and 2 were the least clinically accurate, the location of the first slice may improve with a larger training dataset of external anal sphincter segmentations. The other slices performed similarly.

The strengths of this work are that it allows a non expert to extract the TUI sequences for diagnosis, and it saves a significant amount of time for all (expert and novice) sonographers. Automation may standardise the current procedure and reduce inter-observer variability, this will be studied in future work, on a larger dataset.

The main limitation, is the formation of the rotation matrix, as it is dependent on the external anal sphincter segmentation. Incorrect segmentation due to artifacts, may lead to a non-cylindrical shape, and the volume may not be rotated at all, or rotated too much. In some patients biologically the external anal sphincter may not be cylindrical during contraction, regardless of CNN performance. Thus, in the future, we aim to include segmentation information of the internal anal sphincter.

Another approach to be considered in future work for the formation of the rotation matrix would be to train the CNN to determine the rotation matrix directly from the TPUS volume via CNN regression, similar to work presented by Li *et al.* [165]. The slight disadvantage of such an approach is the *black box* nature of CNNs, i.e., the clinician may not understand why the AI-based algorithm fails to successfully rotate the TPUS volume. This may make the

implementation within US software difficult, as the clinician remains liable for the diagnosis of the patient, therefore, they must be able to interpret the output of AI-based software.

In future work, to follow clinical guidelines more closely, we also aim to ensure the anal canal is not only horizontally aligned in the mid-sagittal plane, but also that it is vertically aligned in the axial plane. This would improve results when the US is acquired sub-optimally (i.e., asymmetric), allowing less-skilled sonographers to perform TUI extraction. As the current method does not correct asymmetric US recordings in the axial plane, the TUI sequences may not intersect the anal canal perpendicularly, leading to sub-optimal TUI sequences.

Previous work highlighted that inter observer agreement for sphincter measurements was fair to excellent for transperineal acquisition [58], however, in future work the evaluation will be expanded to several clinical observers to calculate intra and inter observer variability to reduce bias. Furthermore, the pipeline will be extended to classify anal sphincter tears and disease if present.

We would like to emphasise that, the segmentation of the external anal sphincter is not a clinical goal in itself. From a clinical perspective, the relevant metrics are based on the quality of the re-orientation and re-slicing, which are measured in this work. The segmentation is only used as an intermediate step towards automatic TUI extraction. Although we recognise that the segmentation of the anal canal might be an interesting problem to solve, there is currently no clinical need for such a solution. As this work is driven by clinical relevance, we thus opted not to engage in solving problems that are technologically possible but clinically not required. That being said, we recognise that AI technologies could further support the diagnostic process for anal sphincter tear analysis. Either via the segmentation of relevant structures with subsequent (automatic) extraction of biomarkers or via direct image classification for disease. Both are currently beyond clinical practice but could make the clinical diagnostic process more robust; the added clinical value of such a quantitative process would yet need to be demonstrated, however.

8.5 Conclusion

To conclude, the proposed method achieved an overall clinical acceptability of 90.00%, despite the limitation of the rotation matrix and not rotating the axial plane as performed in clinic to improve asymmetric acquisitions. Thus, we believe with a more detailed pipeline which includes internal anal sphincter segmentation, the results will outperform this method, and may perform comparable to inter-observer variability. The proposed method was 52.36 seconds quicker than the clinical expert, which was significant. The proposed method allows non-expert sonographers to perform TUI sequence extraction for anal sphincter tear diagnosis. In future work, we will conduct an inter and intra observer variability study, and expand the evaluation dataset to 100 TPUS volumes.

Chapter 9

General discussion and future work

9.1 General discussion

This thesis addressed the research topic of automatic ultrasound imaging analysis to support the assessment of pelvic floor disorders. Several novel automated ultrasound imaging analysis frameworks have been developed to replicate common clinical manual tasks. These frameworks can be used either directly or indirectly by a clinician to assess pelvic floor dysfunction, in particular POP and anal sphincter tearing. The frameworks presented in this thesis were developed to follow the clinical workflow, to increase the clinical acceptance of this work, and increase the likelihood that these methods could be implemented in a clinical environment. As a next step, further support for computer-aided diagnostics could be given based on the automatic biomarker extraction to diagnose anal sphincter tearing and levator avulsion.

US imaging analysis is an important part of clinical practice that helps clinicians diagnose POP, levator avulsion, and anal sphincter tearing [55]. These tasks are typically performed manually in the clinic, so they are time consuming, repetitive, prone to inter-observer variability and human error, and require high levels of training and expertise.

Analysis of US images of the pelvic floor is a challenging task due to: acoustic shadowing (which can disrupt the view of important landmarks), the low-contrast difference between tissues (which makes it difficult to distinguish anatomical borders), US artifacts (which can hide or distort important landmarks or planes of view), inhomogeneous tissues causing speckle (which can make it difficult to determine borders between tissues and organs) and poor acquisition (which can cause important structures to not be fully visible within the US image).

In the near future, the burden on pelvic floor disorders clinics is expected to increase, as the number of women suffering and presenting pelvic floor complaints is expected to rise [29]. The *relatively recent* realised benefits of TPUS imaging require high levels of training, expertise and skill. Therefore, TPUS imaging is not a common practice in all pelvic floor disorder clinics. A solution to this ever-growing clinical problem is to automate the assessment of pelvic floor

Chapter 9. General discussion and future work

disorders, where AI-based technologies could be beneficial. AI algorithms have shown that they solve several tasks of US imaging analysis in other medical fields [68].

AI-based technologies can improve the pelvic floor clinical workflow by reducing user-time, standardising practice, reducing the level of expertise required, and reducing the inter-observer error. However, they have not been successfully used in the assessment of pelvic floor disorders. This may be due to the *recently realised* benefits of TPUS imaging, meaning the number of highly skilled clinicians may be limited. In addition, within TPUS imaging there are varied patient pathologies, and common TPUS imaging artifacts (as shown in Section 1.2.1, Fig. 1.9). Therefore, acquiring large, high-quality annotated datasets for deep learning-based techniques may be difficult.

Despite advances in research, the number of AI products for ultrasound imaging analysis approved by the FDA is significantly lower than the number of AI-based ultrasound imaging analysis algorithms in research. This highlights the huge disparity between research and clinical implementation, which may be due to the lack of clinical acceptance of automated ultrasound imaging analysis tools (in particular AI-based technologies) [12].

For clinical acceptance, the technology must perform robustly with a high degree of accuracy, on a challenging clinical dataset. Throughout this thesis, the pipelines were developed with clinical implementation in mind and evaluated on challenging clinical data with a diverse case mix of patient pathologies.

As US image analysis is a visual task that relies on the identification, localisation and extraction of pelvic floor structures, this thesis focuses heavily on automatic and interactive two and three-dimensional segmentation of specific anatomical landmarks and structures in the pelvis (i.e., symphysis pubis, levator ani muscle, urethra, external anal sphincter and levator hiatus) and landmark localisation via regression of the symphysis pubis and levator ani muscle from a TPUS volume. Within this thesis, the deep learning method CNNs and the mechanistic method (*traditional computer vision*) BEAS [118] were used. The combination of these methods can achieve similar results to expert manual segmentation, with no or a few user interactions and a shorter user-time [109].

The pipelines developed reduced user time and provided accurate location and extraction of the plane of minimal hiatal dimensions, accurate segmentation of the levator hiatus, and accurate extraction of the TUI sequence of the anal sphincter complex.

Chapter 1 reviews the most relevant topics for this thesis, starting with anatomical and functional considerations of the pelvic floor, and in particular the levator hiatus and anal sphincter complex. The focus was on US imaging, as it is a commonly used imaging modality to assess pelvic floor anatomy and function in real time. US has known advantages, such as accessibility, availability, relatively cheap compared to MRI imaging, non-ionising nature and three-dimensional real-time imaging capabilities.

In this work, 3D TPUS imaging was chosen, as it is widely available in clinics and most gynaecologists and urologists are familiar with it. It can be used for many important pelvic floor disorder assessment tasks, such as the diagnosis of levator hiatal ballooning, levator avulsion and anal sphincter tearing. 3D TPUS also gives clinicians real-time access to the 3D geometry of the pelvic floor structures, and allows clinicians to evaluate important diagnostic planes that are not accessible via 2D US.

Chapter 2 is a ‘state-of-the-art’ literature review, which focuses on automated US image analysis, automatic and interactive US segmentation, levator hiatus and puborectalis muscle segmentation and US plane detection. The review showed the advanced nature of automated US imaging analysis and segmentation [6]. However, there was limited literature regarding automated US imaging analysis of the pelvic floor when this work was initiated.

There was no literature based on CNN 3D segmentation of pelvic floor structures, the extraction of the plane of minimal hiatal dimensions, or the automatic extraction of a TUI sequence of the anal sphincter complex.

The only literature based on automated US imaging analysis of the pelvic floor were a semi-automatic [100] and automatic active model [99], and a 2D CNN [78] for the automatic segmentation of the levator hiatus from the 2D ultrasound plane of minimal hiatal dimensions, developed in and in collaboration with our team. Overall, the literature review highlighted the clinical and research need for this thesis, and for the development of automated US imaging analysis pipelines for the assessment of pelvic floor disorder.

Since starting the work for this thesis, several papers have proposed methods for 2D hiatal area segmentation from a pre-defined plane [103, 102, 101]. However, they still required *manual* extraction of the plane of minimal hiatal dimensions. These studies did not include inter and/or intra-observer studies, and the testing datasets were not publicly available. In addition, not all literature included important patient demographics of the patient case mix and pathology rates. Therefore, it is difficult to make direct comparisons between the results presented in this thesis and the literature.

Finally, at the end of 2021, van den Noort *et al.* presented a solution to extract the plane of minimal hiatal dimensions from a TPUS volume and subsequently segment the levator hiatus [105]. This method utilised two CNN networks, the first being a CNN that estimated the position of the plane of minimal hiatal dimensions, in a slice-by-slice approach. This network was only trained on the mid-sagittal slice, therefore, it requires an adequate US acquisition from the clinician. The output of the first CNN was then post-processed to fit a plane through the data points of the estimation, and an interpolated 2D slice was extracted. The second CNN was a 2D segmentation network that segmented the levator hiatus.

The results were promising, and the pipeline required a few seconds to process a single TPUS volume (although specific data of the time taken was not supplied). The evaluation presented by van den Noort *et al.* was not as comprehensive as in Chapter 5, and direct comparisons cannot

be made. This is because the performance accuracy of the entire pipeline was not evaluated, but the two CNN tasks (i.e., the plane of minimal hiatal dimension extraction and levator hiatus segmentation) were evaluated separately. Therefore, it is impractical to compare the hiatal area metrics measured in Chapter 5 with [105]. As there is a bias in [105] for the errors to be lower than the true total pipeline error, since manual and automatic levator hiatus segmentations were performed on the *same manually extracted* plane of minimal hiatal dimensions.

The literature review also highlighted key limitations in terms of interactive segmentation methods. The key limitations were that several ‘state-of-the-art’ approaches did not give the user real-time explicit control of the boundary of the segmentation. Therefore, the thesis tried to address this limitation in Chapter 6 and 7.

9.1.1 My technical contributions for automating pelvic floor disorder assessment

The pipelines developed in this thesis aimed to closely follow the clinical US imaging analysis tasks, while being intuitive, reducing user-time, and performing within or lower than the measured inter-observer error.

The first technical contribution of this thesis was the automatic segmentation of several pelvic floor organs from a TPUS volume. **Chapter 3** presented a novel application of automatic segmentation methods on TPUS volumes of the pelvic floor. First, an automatic single-class 3D CNN segmentation of the urethra, and secondly, an automatic multi-class 3D CNN segmentation of the urethra, symphyis pubis and levator ani muscle. Segmentation of these structures was investigated to understand the complexity of 3D segmentation from TPUS images, as there was no relevant literature available prior to this project of 3D CNN pelvic floor segmentation.

The method utilised CNNs, specifically the CNN architectures U-Net and HighRes3DNet. This section explored data augmentation techniques that generated (synthetic) data, increasing the size of the training dataset. Data augmentation improves segmentation accuracy, which is important when there is a limited dataset size. For all chapters, there were no public datasets available. Therefore, datasets of high-quality segmentations were acquired at the pelvic floor disorder clinic at UZ Leuven Hospital in Belgium. The images acquired were part of a clinical study from consenting patients. An expert with over four years of experience in pelvic floor ultrasound imaging curated the segmentations.

The CNN HighRes3DNet was used for single-class urethra segmentation, and experimental results showed that the proposed method achieved accurate segmentation. The automatic multi-class segmentation used a 3D U-Net architecture and several data augmentation techniques to improve the performance of a small training dataset. Small training datasets are common in US segmentation, because of the time-consuming nature of three-dimensional voxel-wise segmentation, and the high level of skill required to delineate anatomical boundaries in an

US image. Data augmentation is therefore necessary. A new version of a data augmentation technique called *mixup* was proposed, in which three US images and their corresponding labels were linearly combined, rather than the traditional two. The *mixup* version used in Chapter 3 improved the accuracy of the segmentation results and it was used in other chapters as a data augmentation technique (i.e., Chapters 5,6,8).

Within clinical practice, manual extraction of the plane of minimal hiatal dimensions requires the clinician to visually identify the location of the symphysis pubis and levator ani muscle by manipulating a three-dimensional volume. Chapter 3 showed promising results for the segmentation of pelvic floor organs from US images. However, when the CNN models for minimal hiatal dimension detection were evaluated on a newly acquired larger dataset, the results were sub-optimal.

The sub-optimal cases were on a variety of TPUS volumes, with noticeable ultrasound acoustic shadowing, severe pathology (i.e., hiatal ballooning) and poor image quality. To improve these results a larger, clinically varied training dataset could be used in the future. In addition, an in-depth study regarding network implementation and architecture could be performed. To note, sub-optimal 3D segmentation inspired the work and development of the 3D interactive segmentation framework in Chapter 7.

Sub-optimal automatic segmentation is a common observation, due to the difficult nature of automatic US segmentation [206], and due to the time consuming and highly skilled task of acquiring large training segmentation datasets. Therefore, another method (based on landmark regression) was proposed to identify these plane-defining landmarks in Chapter 4.

Chapter 4 investigated landmark regression of the plane defining coordinates (i.e., the symphysis pubis and levator ani muscle). Preliminary results showed that despite having a dataset of 25 TPUS volumes, accurate detection and localisation of the symphysis pubis and levator ani muscle was achieved on a clinical, newly acquired TPUS dataset. This led to the development of a novel, efficient framework that used landmark regression and 3D U-Net. 3D U-Net was trained to locate the heat-maps of the plane-defining coordinates of the levator ani muscle and symphysis pubis.

One can assume that the performance of landmark regression CNN was higher than the segmentation CNN, as more detailed voxel-wise information was provided during landmark regression training. In segmentation training, the foreground voxels of the levator ani muscle and the symphysis pubis were small compared to the background voxels. The segmentation task was therefore highly unbalanced, making it more difficult and better suited to landmark regression. Furthermore, the segmentation boundary in CNN is usually noisy due to the down-sampling of US volume and voxel-wise segmentation. Therefore, with a heatmap, the framework bypasses these limitations and the CNN learns to localise the exact extreme coordinate position.

The main strengths of this pipeline were that it closely followed the clinical workflow and did not require user interactions. It was the first method to propose a solution to this clinical

Chapter 9. General discussion and future work

problem, and it took, on average, 100 seconds less time than manual extraction. The pipeline was validated on a challenging clinical dataset of 100 TPUS volumes, and the accuracy of the pipeline was comparable in terms of sensitivity and specificity to a highly experienced clinician with over four years of experience.

This suggests the pipeline would benefit the pelvic floor disorder clinic by reducing the discrepancy between clinicians, standardising the assessment, saving time and allowing ‘novices’ to perform this US analysis task. This could increase the number of clinics that use US imaging to assess pelvic organ prolapse and hiatal ballooning.

In **Chapter 5**, the pipeline from Chapter 4 was extended to automatically segment the levator hiatus from the automatically extracted plane of minimal hiatal dimensions. This pipeline closely follows the clinical protocol of the calculation of the hiatal area. 2D U-Net was used to segment the levator hiatus from the automatically extracted 2D plane. This was the first work to extract the hiatal area from a TPUS volume. Since the publication of this work, another paper [105] has proposed a similar two-step pipeline. Similar to the method in Chapter 5, [105] utilises a CNN to estimate the position of the plane of minimal hiatal dimensions from a transperineal ultrasound volume. Based on this estimation a slice is extracted and fed into a second CNN, which segments the levator hiatus.

In Chapter 5, the computer-observer error was based on the difference between the automatic plane and segmentation against the clinicians manually detected plane and manual hiatal segmentation, to measure the true error of the entire pipeline, whereas in [105] the automatic segmentation of the clinician is performed on the manually extracted plane of the algorithm. Therefore, the error reported in [105] is biased to be lower, as the levator hiatus is a 3D structure, and different C-plane positions will have different hiatal areas. In addition, the Hausdorff distance between the two segmentations will be smaller when the hiatus is segmented in the same plane. Therefore, there is an accumulation of error in our reporting, and Chapter 5 reflects the true error of automating the entire clinical workflow. Whereas [105] reports the error of the two CNN tasks (i.e., the plane of minimal hiatal dimension extraction and levator hiatus segmentation) separately, which is less informative, and can not be directly compared to the inter or intra-observer variability of the entire clinical workflow.

The proposed pipeline did not require user input, and significantly reduced the time to extract the hiatal area from a TPUS volume by 120 seconds. In addition, the pipeline performed with a lower error than the measured inter-observer variability of evaluated metrics. Therefore, it is possible that this pipeline could standardise the extraction of hiatal areas from TPUS volumes if it were to be implemented in a clinical environment, enabling ‘novice’ sonographers to perform this challenging US imaging analysis task accurately.

Unfortunately, in several cases, the segmentation of the levator hiatus was sub-optimal (i.e., either under diagnosing or over diagnosing), due to several factors such as the difficult nature of US, the wide variety of patient pathologies and US image qualities, the limited access to large

US pelvic floor datasets, and the nature of CNNs. Despite the performance metrics of the levator hiatus segmentation in Chapter 5 showing the computer-observer error to be within or lower than the measured inter-observer error. In another study [109], on average the clinician wanted to edit (either minimally or more extensively) 98% (28.5 out of 30) of the automatic levator hiatus segmentations from the CNN. This is a common disadvantage of automatic segmentation, and it is unlikely that a CNN will achieve 100% accuracy, as it has not been exposed to all types of US images (i.e., wide case mix, US and acquisition quality, acoustic shadowing and US artifacts).

Since the levator hiatus is an important diagnostic biomarker for hiatal ballooning and POP, it is important that clinicians can easily correct the segmentation if necessary. Therefore a novel, interactive two-dimensional segmentation pipeline was developed that gives the clinician explicit and robust control over the segmentation boundary of the levator hiatus.

Chapter 6 presents the interactive segmentation pipeline that improves the accuracy of automatic segmentation, gives doctors liability, and reduces user time compared to manual segmentation. The interactive segmentation method represents an automatic CNN segmentation contour as a B-spline explicit active surface (BEAS), where the BEAS boundary can be manipulated and edited in real-time.

The interactive segmentation pipeline was compared to the current clinical tools used in the pelvic floor disorder clinic (4D ViewTM, GE Healthcare; Zipf, Austria). Extensive evaluation showed the proposed framework was more robust than the ‘state-of-the-art’ CNN and reduced the perceived workload (calculated by the NASA Task Load Index survey as described in Chapter 6, Section 6.2.3) by half compared to clinical tools. In addition, the proposed tool required at least 13 seconds less user time than clinical tools, which was significant.

Following the promising results of Chapter 6, the two-dimensional interactive segmentation method was extended to a novel three-dimensional interactive segmentation method called DeepBEAS3D, presented in **Chapter 7**. This interactive segmentation method benefits from the intrinsic nature of BEAS, which ensured smooth and complete 3D segmentation, increasing its anatomical plausibility. This advantage is not commonly acquired in other interactive deep-learning segmentation methods, due to the voxel-wise segmentation of CNN methods. DeepBEAS3D also enabled users to edit the 3D surface precisely in real time, which is unusual in many interactive segmentation frameworks.

DeepBEAS3D was applied to the task of 3D segmentation of the external anal sphincter from a TPUS volume, and compared to the clinical tool used in the pelvic floor disorder clinic (4D View VOCALTM, GE Healthcare; Zipf, Austria). A novel energy term was proposed that drove a hyper-parameter optimisation algorithm, to determine the optimal BEAS hyper-parameters, which could allow for a wider applicability of BEAS and DeepBEAS3D for other structures and modalities. The optimisation algorithm suggests optimal BEAS mesh size and scaling for accurate representation of the segmentation and intuitive user-interaction.

Chapter 9. General discussion and future work

A 3D Slicer module of DeepBEAS3D has been created and will be publicly available with training documentation for greater applicability. DeepBEAS3D achieved precise segmentations, significantly reduced perceived workload and user time compared to the clinical tool. Therefore, this work could be implemented in the clinical environment to increase the applicability of automatic segmentation methods for medical diagnosis.

Chapters 4 to 6 showed several pipelines that automated the clinical assessment of the levator hiatus. The plane of minimal hiatal dimensions was accurately extracted from a TPUS volume, and the levator hiatus was accurately segmented. In chapter 6 a novel interactive 2D segmentation pipeline was developed, to give clinicians explicit control over the levator's hiatus contour and the hiatal area calculation for the diagnosis of POP.

Following this work, this thesis focused on another pelvic floor disorder analysis task, that could aid clinicians in the analysis of anal sphincter tearing. This is to our knowledge the first work to investigate the automatic US analysis of the anal sphincter complex. **Chapter 8** presented a novel pipeline that automatically extracted the TUI sequence from a TPUS volume to analyse the anal sphincter complex.

The pipeline utilised the CNN, 3D U-Net, to segment the external anal sphincter from the TPUS volume. Principle component analysis (PCA) was used to align the segmentation, similar to how a clinician may orientate the US volume in manual analysis. Finally, the US volume was sliced into eight coronal view slices of the anal sphincter complex, similar to the manual TUI sequence. The pipeline is intuitive and follows the manual approach used in the clinic. The multi-step approach and work presented in Chapters 6 and 7 can be used to expand and adapt the pipeline to an interactive one. This would give the clinician explicit control over the correction of suboptimal results and increase the likelihood of clinical implementation and acceptance.

Experimental results showed the pipeline achieved accurate TUI sequence extraction, with significantly less user time than a highly trained expert. On average, the pipeline reduced the time required by 52.36 seconds. The visual acceptance rated by an experienced clinician was 90%. In the future, the pipeline could reduce the need for high-level expertise in anorectal dysfunction analysis and increase the number of clinics performing this difficult but clinically important analysis task.

9.2 Future work

There are many aspects in which the work in this thesis could or should be extended in the future. Five aspects that are most important are: automatic levator hiatus analysis, interactive segmentation, automatic anal sphincter analysis, clinical implementation and clinical deployment.

9.2.1 Automatic levator hiatus analysis and computer aided diagnostics

In this thesis, a pipeline was developed that extracted the hiatal area from a TPUS volume, and an interactive segmentation method was developed to allow for easy correction of sub-optimal hiatal segmentations. In the future, it would be worth exploring network architectures and methodologies to further improve the automatic segmentation result of the levator hiatus in Chapter 5. These methodologies could utilise shape prior information highlighted by Sindhwan *et al.* [100] within the automatic CNNs segmentation. Furthermore, it may be beneficial to increase the training dataset size of the C-plane detection algorithm in Chapter 5, to investigate whether the performance improves with a larger training dataset, as one would expect.

The pipelines presented in this thesis, did not aid the assessment and detection of levator *avulsion* (as described in Chapter 1, Section 1.2.5), an important US imaging analysis task performed to assess the integrity of the levator ani muscle and its attachment to the pubic rami.

To date, there is no literature or solution to automate the diagnosis of levator avulsion. Therefore, an obvious extension of this thesis would be the creation of a pipeline to support the diagnosis of avulsion (i.e., the classification of TPUS images as either intact, left avulsion, right avulsion or bilateral avulsion). To develop such a tool, experienced clinicians would first have to acquire ground truth data and diagnostic labels of ultrasound images. The dataset would have to be of high-quality representing the true clinical environment and several aspects such as: pathology case-mix, patient demographic, ultrasound imaging and acquisition quality need to be considered. As the development of an algorithm to classify levator avulsion has not been achieved in the literature, it is difficult to quantify the dataset size and characteristics required.

One can assume a similar dataset size as levator hiatus segmentation may be adequate for classification, however, it should be expected that more data will need to be acquired. Based on the literature, a dataset of 360 images would be needed, this number is based on the average training dataset size of several high performing CNNs [104, 101, 78]. Once the dataset is curated the performance will be measured and more data will be acquired if necessary.

However, it is worth noting that increasing the quantity of data will not always lead to a higher classification performance. Althnian *et al.* identified that the overall classification performance depends on the extent to which a dataset represents the clinical environment, rather than its size [77]. Therefore, the dataset characteristics will be based on literature that measures the clinical prevalence of levator ani avulsion. One paper [207] measured levator ani avulsion prevalence across multiple centres in a dataset of 300 TPUS volumes. Therefore, first a dataset size of 360 images based on a similar case-mix as reported in [207] will be curated. It is worth noting, that this is a simplified approach and more data may have to be acquired in this case, as the dataset must also account for a variety of imaging and acquisition qualities.

To diagnose levator avulsion, the 2D U-Net from Chapter 5 could be adapted to include a fully convolutional layer that computes a disease score learned during training. Preliminary ‘test’

experiments suggested the network can learn from this approach. However, the acceptance of an anomaly detection software must be high to be implemented in a clinical environment. Therefore, instead of a pure ‘black box’ approach, it would be beneficial to develop a pipeline that provides insights into why a particular disease classification was made.

To achieve this, one could use the automatic hiatal segmentation boundary and measure the symmetry of the automatic segmentation. The symmetry and area would create a risk score that could help doctors diagnose avulsion. This pipeline would have to be extensively validated on a large clinical, challenging dataset, with a fair representation of patients and a varied case mix of avulsion and ballooning.

It is difficult to quantify the dataset size and characteristics required to extensively validate this specific task. Within transperineal ultrasound imaging tasks, the average evaluation dataset size of ‘state-of-the-art’ levator hiatus segmentation [104, 101, 78] was 540. Therefore, this may guide an estimation of an adequate evaluation dataset size. Another approach is to estimate the dataset size based on ultrasound imaging analysis software that has already been approved by the Food and Drug Administration (FDA). One such FDA approved cardiac ultrasound imaging analysis tool [208] extensively validated their automated workflow on 600 cardiac ultrasound images. Therefore, an evaluation dataset of 600 ultrasound images from multiple centres, could showcase the real clinical world performance of this work.

Besides the ‘diagnosis’ aspect of demonstrating avulsion, it is common practice to evaluate levator avulsion in several slices (i.e., a TUI sequence of the levator hiatus). Therefore, to fully automate the diagnosis of levator avulsion from a TPUS volume, the slices fed to the classification algorithm should be automatically extracted from the TPUS volume. To achieve this, one could use the C-plane detection algorithm presented in Chapters 4 and 5 to select the mid-position of the TUI sequence. Once the mid-position is identified, three equi-distant parallel slices above and four equi-distant parallel slices below could be automatically extracted. This has to our knowledge not yet been automated in literature or within industry. However, there is a strong clinical need for such a tool, as it will save time, reduce error between clinicians, and standardise levator avulsion diagnosis.

In Chapters 4 and 5, an algorithm was developed to extract the plane of minimal hiatal dimensions from a 3D TPUS volume. However, clinicians often acquire a 4D (3D + time) clip of the pelvic floor during a pelvic floor contraction or Valsalva manoeuvre [5].

This 4D clip is referred to as a cine-loop. Currently, the clinician must correctly navigate through a sequence of 3D TPUS volumes to determine the volume of interest (i.e., the maximal Valsalva or maximal pelvic floor contraction), which is prone to error, highly repetitive and time-consuming. Thus, the automation of this US imaging analysis task would be beneficial to the current clinical workflow.

To achieve this, the 4D cine-loops could be pre-processed into separate 3D US volumes, which are fed into the regression network from Chapter 4 and Chapter 5. This algorithm would

locate the symphysis pubis and levator ani muscle, and the direction of travel of these structures would be determined (i.e., if the structures move away from each other, it is Valsalva, if they move towards each other, it is pelvic floor contraction, and if there is no overall direction of motion, it is at rest). The algorithm would then extract the TPUS volume of maximal Valsalva or pelvic floor contraction and calculate the hiatal area from this volume. This algorithm should be evaluated on a large dataset of clinical TPUS volumes containing a varied case mix of pathological cases. Following the logic described earlier, a dataset of 600 ultrasound images from multiple-centres should be sufficient in determining the real-world clinical performance.

In addition, in this work the segmentation of pelvic floor structures was performed on 3D US volumes that were commonly selected from 4D Cine-loop US volumes. The temporal information and subsequent US volumes within the Cine-loops could have been used as a data augmentation technique to improve the segmentation performance. This may be investigated and explored in the future.

Finally, this thesis presented several 3D segmentation tasks (i.e., of the urethra, symphysis pubis, levator ani muscle and external anal sphincter). However, the levator hiatus was segmented in two dimensions. It could be useful to segment this structure in three dimensions, and to investigate whether there is a clinical advantage of analysing the levator hiatus in three dimensions. If segmented in three dimensions, the maximum area in one slice could calculate the true maximum hiatal area to diagnose hiatal ballooning.

9.2.2 Interactive segmentation

In Chapter 7, an optimisation algorithm was proposed to increase the accessibility of DeepBEAS3D and BEAS for those who are not experienced in selecting hyper-parameters for contour initialisation and interaction. The optimisation algorithm was based on the lowest mesh size for five randomly selected labels. In the future, it would be interesting to compare this approach to the averaged combined energy of several training labels, to avoid BEAS from under or over fitting. In addition, the optimisation algorithm could be used to finely tune other hyper-parameters of the interactive energy equation of the active contour.

In future work, the DeepBEAS3D framework could be used for CNN active learning. Specifically, the user-defined edits and the ‘edited’ and improved BEAS-based segmentation would be used to compute an error heat-map. This heat-map would penalise the loss function in certain (miss-classification) regions in an active learning process and improve the performance of the original CNN over time.

Finally, the Slicer module developed in Chapter 7 could be expanded to curate segmentations using the BEAS framework (i.e., its original use) [118]. Again, this could increase the use and accessibility of BEAS for generating automatic segmentations for US imaging analysis (i.e., volume calculation of lesions, treatment assessment and surgical planning). The software

could include active learning, automatic segmentation, and interactive segmentation (as proposed in Chapter 7), to bring clinical-based and research-based benefits to US imaging analysis and automatic segmentation of medical images.

9.2.3 Automatic anal sphincter analysis

The evaluation performed in the work described in Chapter 8 was not as detailed as Chapters 4 and 5, so it would be beneficial to compare the automatic pipeline with inter- and intra-observer studies in the future. In chapter 8, the external anal sphincter was used as a reference structure to rotate the volume and extract the TUI sequence. However, this is not always cylindrical for all patients (i.e., the external anal sphincter may appear globular due to age and pathology). Therefore, principal component analysis (PCA) is not suitable for computing accurate rotation angles for TUI extraction in all cases.

Preliminary ‘experiments’ suggested that segmentation of the anal canal would be more accurate in defining the rotation angle required to horizontalise the anal sphincter complex, than the external anal sphincter. Therefore, to improve the work in Chapter 8, an improved pipeline could be developed. This pipeline would automatically segment the anal canal and perform weighted PCA on the segmentation. Weighted PCA would put a higher weight and emphasis on the upper part of the anal canal segmentation. Which is important, as it is where the anal canal border is more defined within the US image, and clinicians use it as a visual reference to determine the rotation angle. Preliminary studies suggest this method would improve the accuracy of TUI extraction, and therefore, it will be developed and extensively evaluated in future work.

There is a strong clinical need yet a lack of automated tools for anal sphincter US imaging analysis. Following the work of Chapter 8 and the improved TUI extraction version mentioned above, it would be beneficial to develop a tool to assist the diagnosis of anal sphincter tearing. This pipeline could help grade anal sphincter damage and tearing [56, 209], which is a difficult task and prone to human error. To achieve this, the algorithm would identify and measure the angle of discontinuity in the anal sphincter border, as is manually done in the protocol outlined by Dietz *et al.* [56]. The measured angles can then be used to grade the anal sphincter damage of a patient, as defined by Sultan [209].

Currently, the methods focused in this thesis and future work for anal sphincter analysis are from TPUS images from GE Healthcare (Zipf, Austria) ultrasound machines. In urology many centres use Bröel & Kjaer (BK) medical (Naerum, Denmark) ultrasound machines, which use endo-anal ultrasound probes to assess anal sphincter tearing, following the protocol defined by Sultan *et al.* [209]. Therefore, it would be beneficial for the pelvic floor disorders community, to develop a pipeline that automatically assess anal sphincter tearing of endo-anal ultrasound images.

9.2.4 Clinical implementation and improvement

Throughout this thesis, several pipelines have been developed to solve real-world, clinical problems in the assessment of pelvic floor disorders. To implement these pipelines in a clinical environment, it is important to implement these algorithms in real-time US software. Currently, these pipelines require a high degree of computational power and take several to 30 seconds (dependent on the task). Therefore, it would be useful to adapt several pipelines to have a lower computational load and to run in real time on ultrasound software. To achieve this, the number of parameters in the CNN could be reduced (as long as performance was not affected), reducing the complexity of the network and the time required to perform the segmentation or regression task. In addition, several of the 3D CNN segmentation or regression tasks could be converted into 2.5D CNN networks (i.e., several 2D slices of the ultrasound volume in a specific plane, either mid-sagittal, coronal or axial) [105]. Since the mid-sagittal plane is used in Chapters 4, 5 and 8, it may be possible to use a subset of these slices at the centre of the US volume, rather than the whole US volume. This would reduce the computational workload and decrease the time required.

The disadvantage, however, is that the clinician must acquire US volumes with adequate quality, as the structures of interest would have to be centralised in the US volumes. Technically, the current pipelines do not struggle with poor acquisition, but a 2.5D approach would only work if the slices extracted from the US contained relevant diagnostic information.

A potential solution to poor acquisition, which is not uncommon in clinical practice, would be to develop a tool that would help improve the *acquisition* of TPUS volumes. Wu *et al.* developed an image quality assessment tool for fetal ultrasound images [210], that localised the region of interest of the fetal abdominal region in the ultrasound image using a CNN. Pasdeloup *et al.* developed a deep learning-based tool that provides real-time feedback to the clinician on how to move the transducer in order to obtain an optimal standard plane view for assessment of cardiac function from 2D ultrasound images. They expect the method to aid less-experienced users to acquire better quality cardiac images for diagnosis, and improve standardisation between more experienced users. This will be investigated in their future studies [211].

The proposed solution could utilise the segmentation CNNs trained in this thesis, to detect whether the structure of interest is within the centre of the ultrasound volume. There are currently no tools for guiding TPUS acquisition. Fig. 9.1, demonstrates the effect of poor ultrasound acquisition. Fig. 9.1 A) shows a high-quality acquisition with the anal sphincter complex centralised and within the US frame. Fig. 9.1 B) shows a lower quality acquisition with the anal sphincter complex off-centre and almost out of frame. This makes manual manipulation, and the analysis of US images difficult, as key areas may not be imaged. In addition, it can reduce the performance of automated US analysis pipelines developed, which is frustrating for the user.

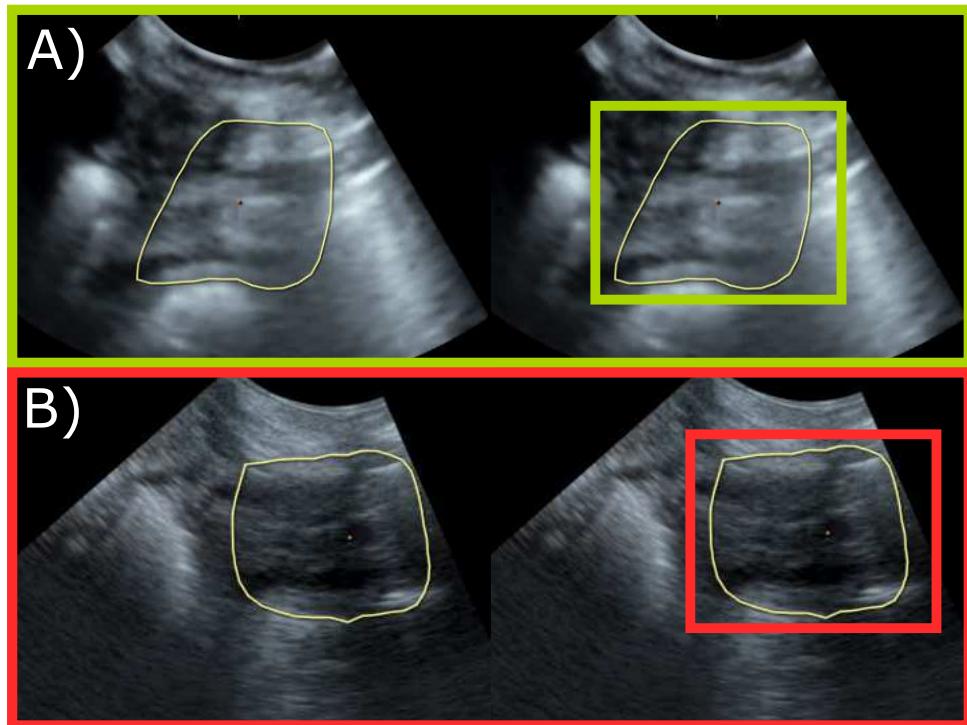


Fig. 9.1 A) shows a high-quality acquisition of the external anal sphincter (outlined in yellow), that is centralised within the mid-sagittal plane. A) (left) no bounding box and (right) shows the region of interest (defined by a green bounding box) is successfully imaged within the US frame, B) shows a sub-optimal quality acquisition of the external anal sphincter (outlined in yellow), that is off-centre and almost out of frame. B) (left) no bounding box and (right) shows the region of interested (defined by a red bounding box) is out of frame.

Therefore, providing a real-time tool that localises the structure of interest and highlights to the clinician when the US volume is of high-quality should improve the quality of the acquisition and performance of the algorithm. It can be suggested that such a tool would standardise the assessment of pelvic floor disorders and help a ‘novice’ acquire high-quality US images that contain relevant and useful diagnostic information [212, 211]. This has been demonstrated by Cheema *et al.* where an FDA approved deep learning-based technology (Caption Health; San Francisco, California) was developed that guides novice clinicians to acquire high-quality cardiac ultrasound images [212].

In combination with the pipelines developed in this thesis, these tools could be used to standardise the whole clinical workflow (i.e., from US acquisition to US imaging analysis for diagnosis) for assessing pelvic floor disorders in patients. Therefore, in the future, this could increase the use of TPUS imaging in pelvic floor clinics, reduce the level of skill required to perform pelvic floor disorder assessment and increase the quality and accuracy of diagnosis.

9.2.5 Clinical deployment and product development

The pipelines developed in this thesis automate the decision-making process of several clinical workflows, in the analysis of pelvic floor disorders from transperineal ultrasound images. The ultimate goal is to implement these AI-based pipelines as software in the clinical setting.

Automated ultrasound imaging analysis software (i.e., AI-based software) can help ease the pressure on the health service (due to an aging population, lack of health personnel and financial constraints) by reducing the clinical time required to perform these tasks. Software can also reduce healthcare inequities among low and middle socio-economic countries, by giving access to automated analysis of ultrasound images without the necessary skill and training required by clinicians, and reduce inefficiency within clinical practice, by reducing the inter-operator variability and clinical time required [213].

Risks of deploying AI-based software in healthcare

In the deployment of AI software in a clinical environment, there are several risks attached. Therefore, before clinical deployment and commercialisation of such algorithms, it is important to outline how to mitigate these risks. Some of the risks of AI-based software (or automated tools in general) in a clinical environment are patient harm due to performance error, the misuse of the tool by clinicians, the risk of bias and inequity, lack of transparency, and patient privacy and cyber-security issues [213] and more may become apparent in the future.

Performance error First, AI-based software could cause patient harm by missing a diagnosis due to false negatives, and causing unnecessary treatment and patient stress due to false positives. The domain shift between training and testing data can reduce the accuracy of the algorithm [214], which may be due to noise, artifacts, and acquisition quality of ultrasound images. This may also be due to differences in patient demographics, ultrasound hardware, acquisition settings, and acquisition protocols. Yan *et al.* [215] reported the lack of AI generalisation, where there was a significant difference in segmentation performance. Testing data from a Philips machine achieved a Dice score of 0.892, and testing data from a Siemens machine achieved a Dice score of 0.474. Therefore, the pipelines should be validated on a multi-centre dataset that represents as much variation as observed within the real-world. So that, failure cases (i.e., incorrect diagnosis or segmentation) can be identified before being released on the market.

Currently, the Medical Device Regulatory (MDR) requests software to be ‘extensively validated’ on ‘high-quality’ data. However, there is currently no literature or specific regulation that guides what this means specifically. The Coordinating Research and Evidence for Medical Devices (CORE-MD) consortium which is sponsored by the EU (H2020), aims to address this challenge by proposing guidelines on how to validate AI-based software for medical use [216].

Chapter 9. General discussion and future work

This shows that this is unclear territory as regulations and guidelines are still being made within the area of AI-based medical devices.

The term ‘high-quality’ is subjective and task dependent. Regarding this thesis and taking the analysis of the levator hiatus as an example, one may assume that a dataset of 500 patients should cover all pathologies sufficiently, as several pathologies exist (e.g., left/right/bi-lateral avulsion, hiatal ballooning and intact). In an over-simplified estimation, one may assume that 100 patients per pathology could be sufficient. However, to measure the AI performance on different ultrasound machines, imaging and acquisition qualities, medical centres and patient demographics, one can assume a substantially larger dataset may be required.

In future work, it would be beneficial to curate a dataset to determine the quantity and quality required to best represent the clinical environment. This could be achieved by observing the patients imaged across a year, and categorising the case-mix, patient demographics, and varying imaging and acquisition qualities observed within the clinic. One can use this dataset to determine the minimum dataset size, case-mix ratio, and patient demographic ratio, required to achieve the same clinical variation observed. This study could be expanded across several institutes to increase the generalisability of the dataset across Europe (to follow GDPR regulations).

Sometimes research groups can rely on public datasets to externally validate their algorithm [81], and accurately compare their method to other ‘state-of-the-art’ methods in the literature. However, there is no public dataset of TPUS images. The development of such a dataset would improve the standard of evaluation between institutes. Therefore, in future work, it would be beneficial to collaborate with several pelvic floor disorders clinics globally, to curate such an annotated dataset. However, current data protection rules (e.g., GDPR), and the variability of their interpretation at national or institutional level, do not make such initiatives easy.

It is important that these solutions include an element of interaction, to ensure the clinician is part of the clinical workflow. An interactive segmentation pipeline was developed for this reason, as shown in Chapters 6 and 7. Several other aspects could still be made interactive. For example, in Chapter 5 an interface that allows the user to adapt the position of the symphysis pubis and levator ani muscle could mitigate inaccurate localisation of the plane of minimal hiatal dimensions. This would save clinical time (when correction is required), increase the trust between the clinician and the AI algorithm, and mitigate AI error [213].

Finally, the pipelines developed could be adapted to allow for a continuous learning approach to mitigate error and improve the performance of the pipeline. Therefore, the performance can improve over time by learning and adapting to new data it is exposed to. However, a weakness of continuous learning is that the AI-algorithm could be negatively affected by incorrect diagnosis within the clinical setting. Therefore, the algorithm may reduce in accuracy as it conforms to the average performance of several clinicians (with different levels of experience). To mitigate this limitation, precautions should be made to ensure only highly skilled clinicians can impact the AI-algorithm, or the corrections are quality-controlled. Overall, the development of such

interfaces and tools would improve the performance and increase clinicians trust in the AI-based software [217].

Human error There is a risk of human error when using AI-based software, which could negatively impact the patient. Incorrect use of AI-based software may be more when a clinician was not involved in development. Therefore, it is important to involve the clinician in the development of the user interface and pipeline, and regarding the training and testing data selection [213, 218]. This will ensure that clinicians views, preferences and clinical knowledge are integrated into the tool [213]. This can increase the trust and interpretability of the software. In addition, by including a clinician in the development process, AI-engineers can ensure desired features are integrated in the software for clinical use, and biologically and physiologically acceptable solutions are created.

In addition, a lack of education and understanding of AI among healthcare personnel may decrease the performance of the software. For correct use, it is important that clinicians understand the advantages and limitations of AI, how the AI was developed (i.e., training data) and validated (i.e., testing data). Therefore, clinicians may more easily foresee and detect bias and fail cases. Therefore, the development of AI-based lectures and courses for medical students and personnel may be beneficial at UZ Leuven, or in relevant societies. The work in this thesis could form an example case study of AI-based software in healthcare [213].

Biased data If this work is to be implemented in the clinical setting, it is important that one addresses current inequalities and inequities that may be present in the AI algorithm and clinic (e.g., sex or gender, age, ethnicity, income, etc). These inequalities may be present in medical datasets collected from institutes used to train AI algorithms.

Unfortunately, as AI learns from data it is exposed to, it could amplify biases that will contribute to further inequality in healthcare [213]. Examples of bias and unbalanced datasets (i.e., under represented patient populations) were discussed in Section 1.3. Fortunately, the data used in this thesis are based on segmentations of anatomical structures, therefore, it is unlikely to be biased. However, a quality check could be performed on a selection of data to assess the segmentations, without the clinician having access to the patients personal information. It is worth noting, that it is difficult to eliminate and avoid bias completely, therefore, methods should focus on bias minimisation.

In the future, it is important to identify if bias is present prior to training, especially if the algorithm is for automatic diagnosis (i.e., anal sphincter tear detection). It is important that future AI-based technologies developed from this thesis are developed with bias and patient representation in mind.

AI transparency Transparency of AI-based software is a requirement of many EU regulations (e.g., MDR) [213, 217]. Transparency is achieved by continuously monitoring the performance and fail cases of the AI-based software. This ensures that patient safety is maintained. To increase AI transparency, it is beneficial to develop pipelines with multi-steps, similar to the pipelines proposed in this thesis. A multi-step pipeline allows the clinician to check the output of each step independently, which not only may increase the accuracy, but also the confidence of clinicians using such tools. In addition, it is important that the training and testing data are technically described. Therefore, for clinical implementation extensive documentation of the AI-based software would have to be created.

Data privacy and cyber security For clinical implementation, it is necessary to be aware of data privacy and cyber security. Globally, there is a risk that patients data may be shared and used without informed consent of the patient [213]. To mitigate security breaches and to address the problem of data governance, institutes can utilise decentralised federated learning approaches to AI [213]. Federated learning allows institutes to benefit from big data from multiple clinical centres, without ever needing to transfer patient's data. Instead, the AI model training process occurs locally, and only AI model characteristics and hyper-parameters are transferred between institutes [219, 218]. To protect patient's data the CNN needs to be trained behind a firewall of the institute and in a 'trusted environment' within the hospital information system. The performance of federated learning is comparable to centralised multi-centre AI learning [219]. Federated learning could enable big data to be used in the EU, while respecting patient privacy and GDPR regulations [219, 218]. Despite the advantages of federated learning, it is worthwhile to note the current challenges. Some challenges are communication between institutes can be expensive (due to the size of the CNN model and amount of hyper-parameters being shared) [220], system heterogeneity (the institutes systems may differ in CPU, memory, network connectivity) [220], therefore, the network size and CNN used should be able to accommodate all institutes hardware specifications and sharing model updates (e.g., gradient information) may still reveal sensitive information (albeit it is safer than sharing raw patient data) [220].

Accountability and liability To increase the adoption of our software in the healthcare setting, it is important to increase trust between clinicians, AI technology and AI developers. Current regulations (that will be discussed in Section 9.2.5) mean that the clinician is accountable and responsible for errors caused by software used in clinic [213]. Without clear documentation and understanding, it can be difficult for the clinician to identify the cause of a fail case (e.g., poor ultrasound image quality), as well as detect a fail case. This may reduce the acceptance and trust between the clinicians and our software, which inevitably may lead to a lower adoption of AI-based software in clinics. For that reason, it may be important to develop systems to alert the clinician when the AI algorithm is uncertain in its prediction [213]. It would also be beneficial,

to develop multi-step pipelines, that allow the clinician to check the output of the pipeline in a order similar to how they perform the manual analysis. In addition, it would be beneficial to also develop research in ‘explainable’ automatic ultrasound imaging analysis [213], which could provide insights into AI uncertainty (e.g., low image quality, incomplete scan, poor acquisition).

Regulatory framework for AI-based tools

AI-based software implemented in the clinical setting is labelled as a medical device. Therefore, for any software developed to be safely used within the clinic, it needs to be verified and validated in terms of reliability, accuracy, and cost-utility [217].

Medical Devices Regulation In the EU, the regulations for medical AI tools are the 2017/745 Medical Devices Regulation (MDR) and the 2017/746 In Vitro Diagnostic Medical Devices Regulation (IVDR). These regulations were passed in 2017, so they are not specific enough for AI-based tools [213, 217]. Current regulations do not account for continuous learning or biases within healthcare data. CORE-MD aim to address limitations of the MDR by bringing together medical associations, EU regulators, academic and public health institutes and industrial associations [216].

Within the MDR framework, all AI-based software is classified as Class II (medium risk) or Class III (high risk). Class II software can fall into the Class IIa or Class IIb category. The classification depends on whether the software can cause harm to the patient [217], as shown in table 9.1. Current regulation requirements from MDR include pre-market control, clinical investigation, post-market surveillance and transparency by the European database of medical devices [217].

The draft AI Act In 2021, the EU Commission published a proposal to further regulate AI technology in Europe, the draft AI Act [213]. It aims to address safety and human rights and classifies all AI-based medical devices in the high-risk category [213]. Therefore, there would be additional requirements and obligations for medical AI-based devices, such as: technical documentation, logging capabilities, transparency, detailed user information on limitations and capabilities, ensuring human oversight, robustness, accuracy and cyber security [213, 222].

The draft AI Act does not specifically consider the healthcare domain, so that, like MDR, continuous learning and healthcare biases are not addressed [213]. Continuous learning not being addressed in regulatory boards is problematic, as a change in the AI algorithm’s performance, which is not pre-determined, would require re-approval from the MDR and the draft AI Act [222, 218].

Chapter 9. General discussion and future work

		Significance of information provided by the medical device software to a healthcare situation related to diagnosis or therapy		
State of patient condition		High Treat or diagnose	Medium Drive clinical management	Low Informs clinical management
Critical		Class III Type IV.i	Class IIb Type III.i	Class IIa Type II.i
Serious		Class IIb Type III.ii	Class IIa Type II.ii	Class IIa Type I.ii
Non-serious		Class IIa Type II.iii	Class IIa Type I.iii	Class IIa Type I.i

Table 9.1 Classification Rule 11a of the MDR [221]. The classification rule is based on the risk framework that considers the significance of information from a medical device and the criticality of the patient's disease or condition.

If the draft AI Act was to come into place, future AI-based software would have to meet the demands and requirements of MDR, GDPR and the AI Act, doubling a lot of the workload with questionable gain [213]. This could lead to slower technological progress in Europe [222, 213].

In addition, there is little guidance on how AI-based software can meet the requirements of these regulatory bodies [222, 213, 218]. Therefore, it would be beneficial to improve the definitions of these regulations in the future, and for the medical and regulatory community to outline specific guidance requirements [218]. For example, the MDR and AI Act both ask for 'high-quality data that represents the clinical environment'. However, there is no guidance on the case-mix prevalence, desired quantities, and patient demographics. Therefore, each AI-based tool must define their own standard of 'high-quality'. To mitigate the different interpretations of these definitions, the regulatory bodies could outline these definitions in detail, and provide access to several successful case studies for standardisation.

Self-assessment questionnaires In addition, it may be worthwhile for medical institutes, like UZ Leuven, to develop self-assessment check-lists and questionnaires for AI-based tools in the specific medical field. These check-lists can be based on fairness, traceability, usability, robustness and explainability, in the interest of the end-user (i.e., clinicians) and the patient's standard of care [213].

Check-lists may standardise the evaluation of AI-based software in specific fields of medicine, and gather detailed information on the strengths and limitations of the software. An example questionnaire for AI-based software developed from work in this thesis is [213]:

1. Were clinicians involved in the design, development and evaluation of the AI-based software?
2. Was the data balanced regarding patient demographics and bias minimised?
3. Was the dataset annotated in a reproducible, standardised way?
4. Were reproducible performance metrics used to assess the AI algorithm?
5. Was the model evaluated on a high-quality dataset representative of the real-world setting (i.e., case mix, imaging qualities, imaging machines, acquisition qualities)?
6. Is the AI-based software well documented?
7. Is the pre-processing pipeline of input data well defined and reproducible?
8. Was the AI-based software output clinically accurate?
9. Did the tool interfere with the patient's assessment?
10. Do you trust the output of the AI-based software?
11. Does the AI-based software take too much time to run?
12. Would you use the AI-based software for all patients? If not, suggest which patients are not suitable?
13. Does the tool perform with the same accuracy for all patient demographics?
14. Did the AI-based software fail? If so, did you understand the potential reason for the fail case (i.e., imaging quality)?
15. Is the user interface easy to use?

These questionnaires would provide more detailed feedback for the improvement and development of AI-based software within healthcare and standardise evaluation.

Data governance and big data

To deploy AI-based software for healthcare in Europe, it must follow the General Data Protection Regulation (EU) 2016/679 (GDPR). GDPR requires the safeguarding of personal data, and requires data minimisation and limitation of re-purposing personal data [217, 218].

GDPR requires medical devices to be transparent with the specific purpose of the use of patient data, patient rights, and grants a patient full control over their personal data (i.e., patients may withdraw consent of a device using their data) [218]. This can be problematic when

Chapter 9. General discussion and future work

regulatory bodies require AI-based software deployed in clinic to perform regular surveillance of software's performance in clinic [218, 217, 213].

The development, deployment and performance of AI-based software depends on large, high-quality training and testing datasets. Current GDPR may hinder the development of such datasets in the EU due to strict regulations regarding patient privacy, and the absence of standardised databases within the EU [218].

The lack of public-available high-quality datasets in the EU and in certain fields of medicine (including pelvic floor disorders) may limit EU healthcare development [218], and for the deployment of AI-based software inspired by the work in this thesis.

The regulations in the EU could affect the innovation and creation of EU-based healthcare companies, as medical devices must comply with several strict regulatory bodies [217, 218, 213]. In addition, with larger interoperable data and less data privacy restrictions, the US and China may have an advantage over Europe, over the development of datasets and AI-based technology [217]. This is deplorable, however, the EU and institutes like UZ KU Leuven, can instead focus on collecting smaller quantities of high quality, trustworthy, and standardised data across a wide range of patient demographics and collaborating with other institutes in the EU, following GDPR regulations. [217].

These high-quality, unbiased datasets can ensure that current healthcare inequalities and inequities (i.e., between patient demographics and EU member states) are not emphasised and it may standardise care [217]. In addition, large datasets of TPUS images could be developed using data synthesis methodologies (i.e., general adversarial networks- GANS) to overcome GDPR restrictions and the lack of available data [217].

Finally, it is also worth noting that the EU regulations can be seen as a strength of the medical institutes within Europe. The goal of MDR is to avoid un-validated and un-reliable healthcare devices (i.e., software) to enter the market, with the aim to improve the standard of healthcare for patients. GDPR ensures that patients' privacy is respected [217], and this could improve the relationship between patients and AI-based technology. Which, in turn, could increase the acceptance, adoption and popularity of AI-based software within healthcare [217].

9.3 Final remarks

In this thesis, several pipelines have been developed and extensively evaluated on challenging clinical data with a diverse case mix of pathologies. The pipelines were developed to closely follow the manual clinical workflow to increase clinical acceptance of AI-based automated pipelines. In particular, the pipelines aided the assessment of hiatal ballooning, an indicator of POP, and aided towards the clinical assessment of anal sphincter tearing.

The first pelvic floor specific pipeline extracted the plane of minimal hiatal dimensions from a TPUS volume, the second extracted the hiatal area from a TPUS volume, and the third extracted the TUI sequence of the anal sphincter complex from a TPUS volume. These tasks before this thesis were not previously reported to have been automated in literature, and despite clinical need, there was limited work on the automation of TPUS analysis. However, since the work on this thesis began, another group has presented work on extracting the plane of minimal hiatal dimensions and the hiatal area from a TPUS volume [105].

For clinical implementation, it is important to develop clinical pipelines that are clinically acceptable. To increase clinical acceptance of this work, two interactive segmentation frameworks for 2D and 3D segmentation were developed. These interactive segmentation frameworks gave the clinician explicit control over the segmentation boundary, and corrections were performed in real time, which is uncommon in many other interactive 3D segmentation frameworks. The interactive segmentation frameworks reduced the perceived workload for the clinician, reduced user time, and improved the performance of the ‘state-of-the-art’ automatic segmentation method.

Automated pipelines that follow the clinical workflow are accurate (i.e., perform within or lower than inter-observer variability), robust (i.e., perform with a high level of accuracy on a wide range of TPUS images including varied image and acquisition qualities, patient pathologies and levels of noise and ultrasound artifacts) and semi-interactive (i.e., allow clinicians to intuitively edit and correct a sub-optimal result), should reduce the level of expertise required for ultrasound-based assessment of pelvic floor disorders. It is therefore logical to suggest that these technologies could further increase the use of TPUS imaging in pelvic floor clinics in the future. Automation can allow ‘novices’ to perform highly skilled tasks, and it can save clinical time both for ‘novices’ and ‘experts’.

A study suggests that the prevalence of pelvic floor disorders will increase in the coming years [32]. Therefore, AI-based technologies that are clinically acceptable could reduce the burden within healthcare hence also pelvic floor clinics. The pipelines in this thesis saved significant clinical time, and the hiatal ballooning assessment pipeline even performed better than the measured inter-observer error.

In the future, there are several new tasks that could be automated, such as levator avulsion detection, TUI extraction of the levator hiatus, maximal Valsalva detection from a cine loop, and anal sphincter tear calculation. In addition, to be implemented within the clinic and on real-time US software, the pipelines could be adapted to have a smaller computational load.

Finally, the acquisition of US volumes is another highly skilled task that has not been addressed in this work or literature. The accuracy of the pipelines and manual analysis depends somewhat on the quality of US acquisition. To increase the use of TPUS imaging within the pelvic floor clinic and to improve the performance of the pipelines in this thesis, there is a clinical need to develop US acquisition assisted tools. These tools could allow ‘novices’ to acquire high-quality TPUS volumes of the pelvic floor for accurate pelvic floor disorder diagnosis.

Chapter 10

Scientific acknowledgements, personal contribution and conflicts of interest

10.1 Scientific acknowledgements

Scientific acknowledgements are stated in each chapter. During the time-line of this thesis, Helena Williams was affiliated to King's College London and collaborated with the Contextual Artificial Intelligence for Computer Assisted Interventions (CAI4CAI) research group from the School of Biomedical Engineering & Imaging Sciences.

Specifically regarding TPUS data curation, framework validation and clinical knowledge:

- Datasets were curated by Dr. Laura Cattani;
- Clinical evaluation of frameworks (in Chapters 3-8) and intra-observer studies (in Chapters 3-5) were performed by Dr. Laura Cattani;
- Dr. Dominique Van Schoubroeck contributed in the clinical validation efforts of Chapters 4 and 5;
- Dr. Susanne Housmans contributed in the clinical validation efforts of Chapter 6;
- Clinical expertise which helped develop and validate 'interpretable' automatic frameworks (i.e., frameworks that accurately followed the clinical workflow) was provided by Dr. Laura Cattani and supervisor Prof. Jan Deprest.

Specifically regarding conception of frameworks, CNN methodologies and evaluation techniques:

- Technical expertise and guidance was provided by supervisors Prof. Tom Vercauteren and Prof. Jan D'hooge;

Chapter 10. Scientific acknowledgements, personal contribution and conflicts of interest

- Dr. João Pedrosa provided support and technical expertise in the implementation and debugging of the BEAS framework in Chapters 6 and 7;
- Dr. João Pedrosa provided support and technical expertise in the development and debugging of the GUI ‘Beyond’ in Chapter 6.
- Dr. Muhammad Asad provided support and technical expertise in the development and debugging of the 3D Slicer module in Chapter 7.

10.2 Personal contribution

I hereby assert that the personal contributions to this thesis included but were not limited to:

- Patient selection for all Chapters guided by the clinical team;
- Conception, TPUS data pre-processing, data augmentation, implementation and training of CNNs, data post-processing and validation of the frameworks in all Chapters 3-8;
- Implementation of the frameworks detailed in Chapters 3-5 and Chapter 8;
- Majority of the implementation ($\approx 60\%$) of the interactive segmentation pipelines in Chapters 6-7;
- Writing of the papers included in this thesis and the remaining chapters.

10.3 Conflicts of interest

Helena Williams was awarded a GPU from the NVIDIA Corporation (Santa Clara, CA, USA) as part of the NVIDIA Academic Hardware Grant Program. In the first year of the PhD Helena Williams was awarded a Travel Grant from ‘Fonds voor Wetenschappelijk Onderzoek’ (FWO) for a short stay abroad (K206619N) at King’s College London, UK. Finally, this project has received funding from GE Healthcare Women’s Health Ultrasound (Zipf, Austria). There are no other conflicts of interest.

Chapter 11

Video links

2D interactive segmentation of the levator hiatus The link from Chapter 6: 2D interactive segmentation of the levator hiatus: <https://www.youtube.com/watch?v=SJy6DmCWBZw>



Fig. 11.1 QR code for a video of the 2D interactive segmentation framework of the levator hiatus.

Chapter 11. Video links

DeepBEAS3D The link from Chapter 7: DeepBEAS3D 3D interactive segmentation:
<https://www.youtube.com/watch?v=-H0xLr2K-f8>



Fig. 11.2 QR code for a video of the 3D interactive segmentation framework of the external anal sphincter.

References

- [1] Allie Garcia-Serra. Principles of Radiation Therapy in Gynecologic Cancer. In *Breast Cancer and Gynecologic Cancer Rehabilitation*, pages 161–170. Elsevier, 2021. doi: 10.1016/B978-0-323-72166-0.00017-7.
- [2] Susan L. Gearhart. Disorders of Defecation in Women. In *Principles of Gender-Specific Medicine*, pages 318–325. Elsevier, 2010. doi: 10.1016/B978-0-12-374271-1.00029-0.
- [3] Jean M. Lawrence, Emily S. Lukacz, Charles W. Nager, Jin-Wen Y. Hsu, and Karl M. Luber. Prevalence and Co-Occurrence of Pelvic Floor Disorders in Community-Dwelling Women. *Obstetrics & Gynecology*, 111:678–685, 3 2008. ISSN 0029-7844. doi: 10.1097/AOG.0b013e3181660c1b.
- [4] Courtney A. Woodfield, Saravanan Krishnamoorthy, Brittany S. Hampton, and Jeffrey M. Brody. Imaging Pelvic Floor Disorders: Trend Toward Comprehensive MRI. *American Journal of Roentgenology*, 194:1640–1649, 6 2010. ISSN 0361-803X. doi: 10.2214/AJR.09.3670.
- [5] Hans P. Dietz. Pelvic floor ultrasound: a review. *American Journal of Obstetrics & Gynecology*, 202:321–334, 2010. ISSN 0002-9378. doi: <https://doi.org/10.1016/j.ajog.2009.08.018>. URL <https://www.sciencedirect.com/science/article/pii/S0002937809009430>.
- [6] Yu Wang, Xink Ge, He Ma, Shouliang Qi, Guanjing Zhang, and Yudong Yao. Deep Learning in Medical Ultrasound Image Analysis: A Review. *IEEE Access*, 9:54310–54324, 2021. ISSN 2169-3536. doi: 10.1109/ACCESS.2021.3071301.
- [7] Matthew Sinclair, Christian F. Baumgartner, Jacqueline Matthew, Wenjia Bai, Juan Cerrolaza Martinez, Yuanwei Li, Sandra Smith, Caroline L. Knight, Bernhard Kainz, Jo Hajnal, Andrew P. King, and Daniel Rueckert. Human-level Performance On Automatic Head Biometrics In Fetal Ultrasound Using Fully Convolutional Neural Networks. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 714–717. IEEE, 7 2018. ISBN 978-1-5386-3646-6. doi: 10.1109/EMBC.2018.8512278.
- [8] Eyad Elyan, Pattaramon Vuttipittayamongkol, Pamela Johnston, Kyle Martin, Kyle McPherson, Carlos F. Moreno-García, Chrisina Jayne, and Md. Mostafa K. Sarker. Computer vision and machine learning for medical image analysis: recent advances, challenges, and way forward. *Artificial Intelligence Surgery*, 2022. ISSN 27710408. doi: 10.20517/ais.2021.15.
- [9] Ohad Oren, Bernard J. Gersh, and Deepak L. Bhatt. Artificial intelligence in medical imaging: switching from radiographic pathological data to clinically meaningful endpoints. *The Lancet Digital Health*, 2:e486–e488, 9 2020. ISSN 25897500. doi: 10.1016/S2589-7500(20)30160-6.
- [10] Maxwell J. Mehlman. Medical practice guidelines as malpractice safe harbors: Illusion or deceit? *Journal of Law, Medicine & Ethics*, 40:286–300, 1 2012. ISSN 1073-1105. doi: 10.1111/j.1748-720X.2012.00664.x.
- [11] George Maliha, Sara Gerke, I. Glenn Cohen, and Ravi B. Parikh. Artificial Intelligence and Liability in Medicine: Balancing Safety and Innovation. *The Milbank Quarterly*, 99: 629–647, 9 2021. ISSN 0887-378X. doi: 10.1111/1468-0009.12504.

Chapter 11. Video links

- [12] Julia Amann, Alessandro Blasimme, Effy Vayena, Dietmar Frey, and Vince I. Madai. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Medical Informatics and Decision Making*, 20:310, 12 2020. ISSN 1472-6947. doi: 10.1186/s12911-020-01332-6.
- [13] G. Willy Davila. Concept of the Pelvic Floor as a Unit. In *Pelvic Floor Dysfunction*, pages 3–6. Springer-Verlag, 2006. ISBN 1-85233-730-3. doi: 10.1007/1-84628-010-9_1.
- [14] G. Willy Davila, Gamal M. Ghoniem, and Steven D. Wexner, editors. *Pelvic Floor Dysfunction*. Springer-Verlag, 2006. ISBN 1-85233-730-3. doi: 10.1007/b136174.
- [15] Jaap Stoker, Steve Halligan, and Clive I. Bartram. Pelvic Floor Imaging. *Radiology*, 218: 621–641, 3 2001. ISSN 0033-8419. doi: 10.1148/radiology.218.3.r01mr26621.
- [16] Jaap Stoker and Christian Wallner. The anatomy of the pelvic floor and sphincters. In *Imaging Pelvic Floor Disorders*, pages 1–29. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008. ISBN 978-3-540-71968-7. doi: 10.1007/978-3-540-71968-7_1. URL https://doi.org/10.1007/978-3-540-71968-7_1.
- [17] Supreeth N. Gowda and Bruno Bordoni. *Anatomy, Abdomen and Pelvis, Levator Ani Muscle*. StatPearls Publishing, 2022.
- [18] Adil E. Bharucha. Pelvic floor: anatomy and function. *Neurogastroenterology and Motility*, 18:507–519, 7 2006. ISSN 1350-1925. doi: 10.1111/j.1365-2982.2006.00803.x.
- [19] Qing-Kai Wu, Xiao-Yuan Mao, Lai-Min Luo, Tao Ying, Qin Li, and Yin-Cheng Teng. Characteristics of pelvic diaphragm hiatus in pregnant women with stress urinary incontinence detected by transperineal three-dimensional ultrasound. *Zhonghua fu chan ke za zhi*, 45:326–30, 5 2010. ISSN 0529-567X.
- [20] Anna Maria Lasak, Marjorie Jean-Michel, Phuong Uyen Le, Roshni Durgam, and Jessica Harroche. The Role of Pelvic Floor Muscle Training in the Conservative and Surgical Management of Female Stress Urinary Incontinence: Does the Strength of the Pelvic Floor Muscles Matter? *PM&R*, 10(11):1198–1210, 2018. doi: <https://doi.org/10.1016/j.pmrj.2018.03.023>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1016/j.pmrj.2018.03.023>.
- [21] Victoria L. Handa, Geoffrey Cundiff, Howard H. Chang, and Kathy J. Helzlsouer. Female sexual function and pelvic floor disorders. *Obstetrics & Gynecology*, 111:1045–1052, 5 2008. ISSN 0029-7844. doi: 10.1097/AOG.0b013e31816bbe85.
- [22] Meadow Maze Good and Ellen R. Solomon. Pelvic Floor Disorders. *Obstetrics and Gynecology Clinics of North America*, 46:527–540, 9 2019. ISSN 08898545. doi: 10.1016/j.ogc.2019.04.010.
- [23] Ingrid Nygaard. Prevalence of Symptomatic Pelvic Floor Disorders in US Women. *JAMA*, 300:1311, 9 2008. ISSN 0098-7484. doi: 10.1001/jama.300.11.1311.
- [24] John O.L. DeLancey, Lisa Kane Low, Janis M. Miller, Divya A. Patel, and Julie A. Tumbarello. Graphic integration of causal factors of pelvic floor disorders: an integrated life span model. *American Journal of Obstetrics and Gynecology*, 199, 2008. ISSN 00029378. doi: 10.1016/j.ajog.2008.04.001.

- [25] Barbara Bodner-Adler, Oliver Kimberger, Thomas Laml, Ksenia Halpern, Clara Beitl, Wolfgang Umek, and Klaus Bodner. Prevalence and risk factors for pelvic floor disorders during early and late pregnancy in a cohort of austrian women. *Archives of Gynecology and Obstetrics*, 300:1325–1330, 11 2019. ISSN 0932-0067. doi: 10.1007/s00404-019-05311-9.
- [26] Cristian Persu, Christopher R. Chapple, Victor Cauni, Stefan Gutue, and Petrisor Geavlete. Pelvic Organ Prolapse Quantification System (POP-Q) - a new era in pelvic prolapse staging. *Journal of medicine and life*, 4, 2011. ISSN 1844122X.
- [27] Michelle Verbeek and Lynsey Hayward. Pelvic Floor Dysfunction And Its Effect On Quality Of Sexual Life. *Sexual Medicine Reviews*, 7:559–564, 10 2019. ISSN 20500521. doi: 10.1016/j.sxmr.2019.05.007.
- [28] Ilknur Kepenekci, Betul Keskinkilic, Filiz Akinsu, Petek Cakir, Atilla H. Elhan, Ayhan B. Erkek, and Mehmet A. Kuzu. Prevalence of Pelvic Floor Disorders in the Female Population and the Impact of Age, Mode of Delivery, and Parity. *Diseases of the Colon & Rectum*, 54:85–94, 1 2011. ISSN 0012-3706. doi: 10.1007/DCR.0b013e3181fd2356.
- [29] Jennifer M. Wu, Camille P. Vaughan, Patricia S. Goode, David T. Redden, Kathryn L. Burgio, Holly E. Richter, and Alayne D. Markland. Prevalence and Trends of Symptomatic Pelvic Floor Disorders in U.S. Women. *Obstetrics & Gynecology*, 123:141–148, 1 2014. ISSN 0029-7844. doi: 10.1097/AOG.0000000000000057.
- [30] Minda Neimark. Genital Prolapse, Urogenital Atrophy, and Sexual Dysfunction. In *Pelvic Floor Dysfunction: A Multidisciplinary Approach*, pages 19–23. Springer London, London, 2009. ISBN 978-1-84800-348-4. doi: 10.1007/978-1-84800-348-4_2. URL https://doi.org/10.1007/978-1-84800-348-4_2.
- [31] Ambre Olsen, Virginia Smith, John Bergtorm, Joyce Colling, and Amanda Clark. Epidemiology of surgically managed pelvic organ prolapse and urinary incontinence. *Obstetrics & Gynecology*, 89:501–506, 4 1997. ISSN 00297844. doi: 10.1016/S0029-7844(97)00058-6.
- [32] Jennifer M. Wu, Andrew F. Hundley, Rebekah G. Fulton, and Evan R. Myers. Forecasting the Prevalence of Pelvic Floor Disorders in U.S. Women. *Obstetrics & Gynecology*, 114: 1278–1283, 12 2009. ISSN 0029-7844. doi: 10.1097/AOG.0b013e3181c2ce96.
- [33] G. Callewaert, M. Albersen, K. Janssen, M. S. Damaser, T. Van Mieghem, C. H. Van Der Vaart, and J. Deprest. The impact of vaginal delivery on pelvic floor function - delivery as a time point for secondary prevention. *BJOG: An International Journal of Obstetrics and Gynaecology*, 123, 2016. ISSN 14710528. doi: 10.1111/1471-0528.13505.
- [34] Hans P. Dietz and Valeria Lanzarone. Levator Trauma After Vaginal Delivery. *Obstetrics & Gynecology*, 106:707–712, 10 2005. ISSN 0029-7844. doi: 10.1097/01.AOG.0000178779.62181.01.
- [35] Lennox Hoyte, Margot S. Damaser, Simon K. Warfield, Giridhar Chukkapalli, Amitava Majumdar, Dong Ju Choi, Abhishek Trivedi, and Petr Krysl. Quantity and distribution of levator ani stretch during simulated vaginal childbirth. *American Journal of Obstetrics & Gynecology*, 199:198.e1–198.e5, 8 2008. ISSN 00029378. doi: 10.1016/j.ajog.2008.04.027.

Chapter 11. Video links

- [36] Kamil Svabík, Ka Lai Shek, and Hans P. Dietz. How much does the levator hiatus have to stretch during childbirth? *BJOG: An International Journal of Obstetrics and Gynaecology*, 116, 2009. ISSN 14700328. doi: 10.1111/j.1471-0528.2009.02321.x.
- [37] Jan A. Deprest, Rufus Cartwright, Hans P. Dietz, Luiz Gustavo Oliveira Brito, Marianne Koch, Kristina Allen-Brady, Jittima Manonai, Adi Y. Weintraub, John W. F. Chua, Romana Cuffolo, Felice Sorrentino, Laura Cattani, Judith Decoene, Anne-Sophie Page, Natalie Weeg, Glauzia M. Varella Pereira, Marina Gabriela M. C. Mori da Cunha de Carvalho, Katerina Mackova, Lucie Hajkova Hympanova, Pamela Moalli, Oksana Shynlova, Marianna Alperin, and Maria Augusta T. Bortolini. International Urogynecological Consultation (IUC): pathophysiology of pelvic organ prolapse (POP). *International Urogynecology Journal*, 33:1699–1710, 7 2022. ISSN 0937-3462. doi: 10.1007/s00192-022-05081-0.
- [38] Abdul H. Sultan, Ranee Thakar, and Dee E. Fenner, editors. *Perineal and Anal Sphincter Trauma*. Springer London, 2007. ISBN 978-1-85233-926-5. doi: 10.1007/978-1-84628-503-5.
- [39] Andrzej Pomian, Wojciech Lisik, Maciej Kosieradzki, and Ewa Barcz. Obesity and Pelvic Floor Disorders: A Review of the Literature. *Medical Science Monitor*, 22:1880–1886, 6 2016. ISSN 1643-3750. doi: 10.12659/MSM.896331.
- [40] Gin-Den Chen. Pelvic Floor Dysfunction in Aging Women. *Taiwanese Journal of Obstetrics and Gynecology*, 46:374–378, 12 2007. ISSN 10284559. doi: 10.1016/S1028-4559(08)60006-6.
- [41] Hans P. Dietz, Clara Shek, Jeremiah De Leon, and Anneke B. Steensma. Ballooning of the levator hiatus. *Ultrasound in Obstetrics & Gynecology*, 31:676–680, 6 2008. ISSN 09607692. doi: 10.1002/uog.5355.
- [42] Christina Lewicky-Gaupp, Quinn Hamilton, James Ashton-Miller, Markus Huebner, John O.L. DeLancey, and Dee E. Fenner. Anal sphincter structure and function relationships in aging and fecal incontinence. *American Journal of Obstetrics & Gynecology*, 200:559.e1–559.e5, 5 2009. ISSN 00029378. doi: 10.1016/j.ajog.2008.11.009.
- [43] Katarzyna Borycka-Kiciak, Marcel Młyńczak, Adam Kiciak, Piotr Pietrzak, and Adam Dziki. Non-invasive obstetric anal sphincter injury diagnostics using impedance spectroscopy. *Scientific Reports*, 9:7097, 12 2019. ISSN 2045-2322. doi: 10.1038/s41598-019-43637-1.
- [44] Johan Pollack, Johan Nordenstam, Sophia Brismar, Annika Lopez, Daniel Altman, and Jan Zetterstrom. Anal incontinence after vaginal delivery: A five-year prospective cohort study. *Obstetrics & Gynecology*, 104, 2004. ISSN 00297844. doi: 10.1097/01.AOG.0000147597.45349.e8.
- [45] D. N. Samarasekera, M. T. Bekhit, Y. Wright, R. H. Lowndes, K. P. Stanley, J. P. Preston, P. Preston, and C. T.M. Speakman. Long-term anal continence and quality of life following postpartum anal sphincter injury. *Colorectal Disease*, 10, 2008. ISSN 14628910. doi: 10.1111/j.1463-1318.2007.01445.x.
- [46] José Marcio N. Jorge and Leonardo A. Bustamante-Lopez. Pelvic floor anatomy. *Annals of Laparoscopic and Endoscopic Surgery*, 7(0), 2022. ISSN 2518-6973. URL <https://ales.amegroups.com/article/view/8337>.

- [47] Maria Nyhus, Sissel H. Oversand, Oyvind Salvesen, Kjell A. Salvesen, Seema Mathew, and Ingrid Volløyhaug. Ultrasound assessment of pelvic floor muscle contraction: reliability and development of an ultrasound-based contraction scale. *Ultrasound in Obstetrics & Gynecology*, 55, 2020. ISSN 14690705. doi: 10.1002/uog.20382.
- [48] Hans P. Dietz. Ultrasound imaging of the pelvic floor. Part I: two-dimensional aspects. *Ultrasound in Obstetrics & Gynecology*, 23:80–92, 1 2004. ISSN 09607692. doi: 10.1002/uog.939.
- [49] Paul Suetens. *Fundamentals of Medical Imaging*. Cambridge University Press, 2017. ISBN 9781316671849. doi: 10.1017/9781316671849.
- [50] Ricardo G. Dantas, Eduardo T. Costa, and Sidney Leeman. Ultrasound speckle and equivalent scatterers. *Ultrasonics*, 43, 2005. ISSN 0041624X. doi: 10.1016/j.ultras.2004.11.003.
- [51] Thomas L. Szabo. Chapter 1 - Introduction. In Thomas L. Szabo, editor, *Diagnostic Ultrasound Imaging: Inside Out (Second Edition)*, pages 1–37. Academic Press, Boston, second edition edition, 2014. ISBN 978-0-12-396487-8. doi: <https://doi.org/10.1016/B978-0-12-396487-8.00001-X>. URL <https://www.sciencedirect.com/science/article/pii/B978012396487800001X>.
- [52] Hans P. Dietz. Ultrasound imaging of the pelvic floor. Part II: three-dimensional or volume imaging. *Ultrasound in Obstetrics & Gynecology*, 23:615–625, 6 2004. ISSN 09607692. doi: 10.1002/uog.1072.
- [53] Frank G. Sommer, Roy A. Filly, and Micheal J. Minton. Acoustic shadowing due to refractive and reflective effects. *American Journal of Roentgenology*, 132:973–979, 6 1979. ISSN 0361-803X. doi: 10.2214/ajr.132.6.973.
- [54] Andrzej P. Wieczorek, Magdalena Maria Woźniak, Jacek Piłat, and Giulio A. Santoro. Principles and Technical Aspects of Integrated Pelvic Floor Ultrasound. In Giulio A. Santoro, Andrzej P. Wieczorek, and Abdul H. Sultan, editors, *Pelvic Floor Disorders: A Multidisciplinary Textbook*, pages 73–87. Springer International Publishing, Cham, 2021. ISBN 978-3-030-40862-6. doi: 10.1007/978-3-030-40862-6_5. URL https://doi.org/10.1007/978-3-030-40862-6_5.
- [55] Hans P. Dietz. Transperineal Ultrasonography: Methodology and Normal Pelvic Floor Anatomy. In Giulio A. Santoro, Andrzej P. Wieczorek, and Abdul H. Sultan, editors, *Pelvic Floor Disorders*. Springer International Publishing, 2021. ISBN 978-3-030-40861-9. doi: 10.1007/978-3-030-40862-6.
- [56] Hans P. Dietz. Exoanal imaging of the anal sphincters. *Journal of Ultrasound in Medicine*, 37:263–280, 2018. doi: <https://doi.org/10.1002/jum.14246>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/jum.14246>.
- [57] H P Dietz, C Shek, and B Clarke. Biometry of the pubovisceral muscle and levator hiatus by three-dimensional pelvic floor ultrasound. *Ultrasound in Obstetrics & Gynecology*, 25: 580–585, 2005. doi: <https://doi.org/10.1002/uog.1899>. URL <https://obgyn.onlinelibrary.wiley.com/doi/abs/10.1002/uog.1899>.

Chapter 11. Video links

- [58] Laura Cattani, Dominique Van Schoubroeck, Susanne Housmans, Geertje Callewaert, Erika Werbrouck, Jan Y. Verbakel, and Jan Deprest. Exo-anal imaging of the anal sphincter: a comparison between introital and transperineal image acquisition. *International urogynecology journal*, 31:1107–1113, 6 2020. ISSN 0937-3462. doi: 10.1007/s00192-019-04122-5. URL <https://doi.org/10.1007/s00192-019-04122-5>.
- [59] Greggory R. DeVore and Bardo Polanko. Tomographic Ultrasound Imaging of the Fetal Heart. *Journal of Ultrasound in Medicine*, 24:1685–1696, 12 2005. ISSN 02784297. doi: 10.7863/jum.2005.24.12.1685.
- [60] Hans P. Dietz. Quantification of major morphological abnormalities of the levator ani. *Ultrasound in Obstetrics & Gynecology*, 29, 2007. ISSN 09607692. doi: 10.1002/uog.3951.
- [61] Hans P. Dietz, Maria J. Bernardo, Adrienne Kirby, and Ka Lai Shek. Minimal criteria for the diagnosis of avulsion of the puborectalis muscle by tomographic ultrasound. *International Urogynecology Journal*, 22:699–704, 6 2011. ISSN 0937-3462. doi: 10.1007/s00192-010-1329-4.
- [62] Rodrigo A. Guzmán Rojas, Ka Lai Shek, Susanne M. Langer, and Hans P. Dietz. The prevalence of anal sphincter injury in primiparous women. *Ultrasound in Obstetrics & Gynecology*, 3 2013. ISSN 09607692. doi: 10.1002/uog.12481.
- [63] Rodrigo A. Guzmán Rojas, Kjell Salvesen, and Ingrid Volløyhaug. Anal sphincter defects and fecal incontinence 15–24 years after first delivery: a cross-sectional study. *Ultrasound in Obstetrics & Gynecology*, 51, 2018. ISSN 14690705. doi: 10.1002/uog.18827.
- [64] Hans P. Dietz and Judy M. Simpson. Levator trauma is associated with pelvic organ prolapse. *BJOG: An International Journal of Obstetrics & Gynaecology*, 115:979–984, 2008.
- [65] Sourabh Kulhare, Xinliang Zheng, Curosh Mehanian, Cynthia Gregory, Meihua Zhu, Kenton Gregory, Hua Xie, James McAndrew Jones, and Benjamin Wilson. Ultrasound-based detection of lung abnormalities using single shot detection convolutional neural networks. In Danail Stoyanov, Zeike Taylor, Stephen Aylward, João Manuel R.S. Tavares, Yiming Xiao, Amber Simpson, Anne Martel, Lena Maier-Hein, Shuo Li, Hassan Rivaz, Ingerid Reinertsen, Matthieu Chabanas, and Keyvan Farahani, editors, *Simulation, Image Processing, and Ultrasound Systems for Assisted Diagnosis and Navigation*, pages 65–73, Cham, 2018. Springer International Publishing. ISBN 978-3-030-01045-4.
- [66] Yi-Cheng Zhu, Peng-Fei Jin, Jie Bao, Quan Jiang, and Ximing Wang. Thyroid ultrasound image classification using a convolutional neural network. *Annals of translational medicine*, 9:1526, 10 2021. ISSN 2305-5839. doi: 10.21037/atm-21-4328.
- [67] Hao Chen, Lingyun Wu, Qi Dou, Jing Qin, Shengli Li, Jie-Zhi Cheng, Dong Ni, and Pheng-Ann Heng. Ultrasound Standard Plane Detection Using a Composite Neural Network Framework. *IEEE Transactions on Cybernetics*, 47:1576–1586, 6 2017. ISSN 2168-2267. doi: 10.1109/TCYB.2017.2685080.
- [68] Shengfeng Liu, Yi Wang, Xin Yang, Baiying Lei, Li Liu, Shawn Xiang Li, Dong Ni, and Tianfu Wang. Deep Learning in Medical Ultrasound Analysis: A Review. *Engineering*, 5: 261–275, 4 2019. ISSN 20958099. doi: 10.1016/j.eng.2018.11.020.

- [69] Kimberly A. Kenne, Linder Wendt, and J. Brooks Jackson. Prevalence of pelvic floor disorders in adult women being seen in a primary care setting and associated risk factors. *Scientific Reports*, 12:9878, 12 2022. ISSN 2045-2322. doi: 10.1038/s41598-022-13501-w.
- [70] Ryan Therrien and Scott Doyle. Role of training data variability on classifier performance and generalizability. In John E. Tomaszewski and Metin N. Gurcan, editors, *Medical Imaging 2018: Digital Pathology*, volume 10581, page 1058109. International Society for Optics and Photonics, SPIE, 2018. doi: 10.1117/12.2293919. URL <https://doi.org/10.1117/12.2293919>.
- [71] Trishan Panch, Heather Mattie, and Rifat Atun. Artificial intelligence and algorithmic bias: Implications for health systems. *Journal of Global Health*, 9, 2019. ISSN 20472986. doi: 10.7189/jogh.09.020318.
- [72] Natalia Norori, Qiyang Hu, Florence Marcelle Aellen, Francesca Dalia Faraci, and Athina Tzovara. Addressing bias in big data and AI for health care: A call for open science. *Patterns*, 2, 2021. ISSN 26663899. doi: 10.1016/j.patter.2021.100347.
- [73] Laleh Seyyed-Kalantari, Haoran Zhang, Matthew B.A. McDermott, Irene Y. Chen, and Marzyeh Ghassemi. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. *Nature Medicine*, 27, 2021. ISSN 1546170X. doi: 10.1038/s41591-021-01595-0.
- [74] Hannah E. Knight, Sarah R. Deeny, Kathryn Dreyer, Jorgen Engmann, Maxine Mackintosh, Sobia Raza, Mai Stafford, Rachel Tesfaye, and Adam Steventon. Challenging racism in the use of health data. *The Lancet Digital Health*, 3, 2021. ISSN 25897500. doi: 10.1016/S2589-7500(21)00019-4.
- [75] Avinash V. Varadarajan, Pinal Bavishi, Paisan Ruamviboonsuk, Peranut Chotcomwongse, Subhashini Venugopalan, Arunachalam Narayanaswamy, Jorge Cuadros, Kuniyoshi Kanai, George Bresnick, Mongkol Tadarati, Sukhum Silpa-archa, Jirawut Limwattanayyingyong, Variya Nganthavee, Joseph R. Ledsam, Pearse A. Keane, Greg S. Corrado, Lily Peng, and Dale R. Webster. Predicting optical coherence tomography-derived diabetic macular edema grades from fundus photographs using deep learning. *Nature Communications*, 11, 2020. ISSN 20411723. doi: 10.1038/s41467-019-13922-8.
- [76] Margarita Sordo and Qing Zeng. On sample size and classification accuracy: A performance comparison. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 3745 LNBI, 2005. doi: 10.1007/11573067_20.
- [77] Alhanoof Althnian, Duaa AlSaeed, Heyam Al-Baity, Amani Samha, Alanoud Bin Dris, Najla Alzakari, Afnan Abou Elwafa, and Heba Kurdi. Impact of Dataset Size on Classification Performance: An Empirical Evaluation in the Medical Domain. *Applied Sciences*, 11:796, 1 2021. ISSN 2076-3417. doi: 10.3390/app11020796.
- [78] Ester Bonmati, Yipeng Hu, Nikhil Sindhwan, Hans P. Dietz, Jan D'hooge, Dean Barratt, Jan Deprest, and Tom Vercauteren. Automatic segmentation method of pelvic floor levator hiatus in ultrasound using a self-normalising neural network. *Journal of Medical Imaging*, 5, 5 2017. doi: 10.1117/1.JMI.5.2.021206.
- [79] Yujie Yang, Pinli Yang, and Bo Zhang. Automatic segmentation in fetal ultrasound images based on improved U-net. *Journal of Physics: Conference Series*, 1693:012183, 12 2020. ISSN 1742-6588. doi: 10.1088/1742-6596/1693/1/012183.

Chapter 11. Video links

- [80] Shakiba Moradi, Mostafa Ghelich Oghli, Azin Alizadehasl, Isaac Shiri, Niki Oveis, Mehrdad Oveis, Majid Maleki, and Jan D'hooge. MFP-Unet: A novel deep learning based approach for left ventricle segmentation in echocardiography. *Physica Medica*, 67: 58–69, 11 2019. ISSN 11201797. doi: 10.1016/j.ejmp.2019.10.001.
- [81] Bjoern H. Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, Levente Lanczi, Elizabeth Gerstner, Marc André Weber, Tal Arbel, Brian B. Avants, Nicholas Ayache, Patricia Buendia, D. Louis Collins, Nicolas Cordier, Jason J. Corso, Antonio Criminisi, Tilak Das, Hervé Delingette, Çağatay Demiralp, Christopher R. Durst, Michel Dojat, Senan Doyle, Joana Festa, Florence Forbes, Ezequiel Geremia, Ben Glocker, Polina Golland, Xiaotao Guo, Andac Hamamci, Khan M. Iftekharuddin, Raj Jena, Nigel M. John, Ender Konukoglu, Danial Lashkari, José António Mariz, Raphael Meier, Sérgio Pereira, Doina Precup, Stephen J. Price, Tammy Riklin Raviv, Syed M.S. Reza, Michael Ryan, Duygu Sarikaya, Lawrence Schwartz, Hoo Chang Shin, Jamie Shotton, Carlos A. Silva, Nuno Sousa, Nagesh K. Subbanna, Gabor Szekely, Thomas J. Taylor, Owen M. Thomas, Nicholas J. Tustison, Gozde Unal, Flor Vasseur, Max Wintermark, Dong Hye Ye, Liang Zhao, Binsheng Zhao, Darko Zikic, Marcel Prastawa, Mauricio Reyes, and Koen Van Leemput. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Transactions on Medical Imaging*, 34, 2015. ISSN 1558254X. doi: 10.1109/TMI.2014.2377694.
- [82] Rikiya Yamashita, Mizuho Nishio, Richard Kinh Gian Do, and Kaori Togashi. Convolutional neural networks: an overview and application in radiology. *Insights into Imaging*, 9:611–629, 8 2018. ISSN 18694101. doi: 10.1007/s13244-018-0639-9.
- [83] Honglak Lee, Roger Grosse, Rajesh Ranganath, and Andrew Y. Ng. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In *Proceedings of the 26th International Conference On Machine Learning, ICML 2009*, 2009. doi: 10.1145/1553374.1553453.
- [84] Rasoul Sali, Sodiq Adewole, Lubaina Ehsan, Lee A. Denson, Paul Kelly, Beatrice C. Amadi, Lori Holtz, Syed Asad Ali, Sean R. Moore, Sana Syed, and Donald E. Brown. Hierarchical Deep Convolutional Neural Networks for Multi-category Diagnosis of Gastrointestinal Disorders on Histopathological Images. In *2020 IEEE International Conference on Healthcare Informatics, ICHI 2020*, 2020. doi: 10.1109/ICHI48887.2020.9374332.
- [85] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521:436–444, 5 2015. ISSN 0028-0836. doi: 10.1038/nature14539.
- [86] Tien-Bach-Thanh Do, Dang-Linh Trinh, Minh-Trieu Tran, Guee-Sang Lee, Soo-Hyung Kim, and Hyung-Jeong Yang. Deep learning based ensemble approach for 3d mri brain tumor segmentation. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 7th International Workshop, BrainLes 2021, Held in Conjunction with MICCAI 2021, Virtual Event, September 27, 2021, Revised Selected Papers, Part II*, pages 210–221, Berlin, Heidelberg, 2021. Springer-Verlag. doi: 10.1007/978-3-031-09002-8_19.
- [87] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. In *Proceedings of the IEEE*, volume 86, pages 2278–2324, 1998. doi: 10.1109/5.726791.

- [88] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60:84–90, 5 2017. ISSN 0001-0782. doi: 10.1145/3065386.
- [89] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9. IEEE, 6 2015. ISBN 978-1-4673-6964-0. doi: 10.1109/CVPR.2015.7298594.
- [90] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778. IEEE, 6 2016. ISBN 978-1-4673-8851-1. doi: 10.1109/CVPR.2016.90.
- [91] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. *CoRR*, abs/1505.04597, 2015. URL <http://arxiv.org/abs/1505.04597>.
- [92] Çiçek Özgün, Abdulkadir Ahmed, Lienkamp Soeren S., Brox Thomas, and Ronneberger Olaf. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. In Sébastien Ourselin, Leo Joskowicz, Mert R. Sabuncu, Gozde Unal, and William Wells, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, pages 424–432, Cham, 2016. Springer International Publishing. ISBN 978-3-319-46723-8.
- [93] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In *2016 Fourth International Conference on 3D Vision (3DV)*, pages 565–571. IEEE, 10 2016. ISBN 978-1-5090-5407-7. doi: 10.1109/3DV.2016.79.
- [94] Wenqi Li, Guotai Wang, Lucas Fidon, Sébastien Ourselin, M. Jorge Cardoso, and Tom Vercauteren. On the Compactness, Efficiency, and Representation of 3D Convolutional Networks: Brain Parcellation as a Pretext Task. *CoRR*, abs/1707.01992, 2017. URL <http://arxiv.org/abs/1707.01992>.
- [95] Xiaomeng Li, Hao Chen, Xiaojuan Qi, Qi Dou, Chi Wing Fu, and Pheng Ann Heng. H-DenseUNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation from CT Volumes. *IEEE Transactions on Medical Imaging*, 37, 2018. ISSN 1558254X. doi: 10.1109/TMI.2018.2845918.
- [96] Fabian Isensee, Philipp Kickingereder, Wolfgang Wick, Martin Bendszus, and Klaus H Maier-Hein. No New-Net. *CoRR*, abs/1809.10483, 2018. URL <http://arxiv.org/abs/1809.10483>.
- [97] Franziska Siafarikas, Jette Staer-Jensen, Ingeborg H. Braekken, Kari Bø, and Marie Ellström Engh. Learning process for performing and analyzing 3D/4D transperineal ultrasound imaging and interobserver reliability study. *Ultrasound in Obstetrics & Gynecology*, 41:312–317, 3 2013. ISSN 09607692. doi: 10.1002/uog.11192.
- [98] Zeelha Abdool, Abdul H Sultan, and Ranee Thakar. Ultrasound imaging of the anal sphincter complex: a review. *The British Journal of Radiology*, 85:865–875, 7 2012. ISSN 0007-1285. doi: 10.1259/bjr/27314678.

Chapter 11. Video links

- [99] F. van den Noort, A. T. M. Grob, C. H. Slump, C. H. van der Vaart, and M. van Stralen. Automatic segmentation of puborectalis muscle on three-dimensional transperineal ultrasound. *Ultrasound in Obstetrics & Gynecology*, 52:97–102, 7 2018. ISSN 0960-7692. doi: 10.1002/uog.18927.
- [100] Nikhil Sindhwanı, Daniel Barbosa, Martino Alessandrini, Brecht Heyde, Hans P. Dietz, Jan D’hooge, and Jan Deprest. Semi-Automatic Outlining of Levator Hiatus. *Ultrasound in Obstetrics & Gynecology*, 48:98–105, 2016. ISSN 0960-7692.
- [101] Zeping Huang, Enze Qu, Yishuang Meng, Man Zhang, Qiuwen Wei, Xianghui Bai, and Xinling Zhang. Deep learning-based pelvic levator hiatus segmentation from ultrasound images. *European Journal of Radiology Open*, 9:100412, 2022. ISSN 23520477. doi: 10.1016/j.ejro.2022.100412.
- [102] Xu Li, Yuan Hong, Dexing Kong, and Xinling Zhang. Automatic segmentation of levator hiatus from ultrasound images using U-net with dense connections. *Physics in Medicine & Biology*, 64:075015, 4 2019. ISSN 1361-6560. doi: 10.1088/1361-6560/ab0ef4.
- [103] Na Wang, Yi Wang, Huifang Wang, Baiying Lei, Tianfu Wang, and Dong Ni. Auto-context fully convolutional network for levator hiatus segmentation in ultrasound images. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 1479–1482. IEEE, 4 2018. ISBN 978-1-5386-3636-7. doi: 10.1109/ISBI.2018.8363852.
- [104] F. van den Noort, C. H. van der Vaart, A. T. M. Grob, K. van de Waarsenburg, C. H. Slump, and M. van Stralen. Deep learning enables automatic quantitative assessment of puborectalis muscle and urogenital hiatus in plane of minimal hiatal dimensions. *Ultrasound in Obstetrics & Gynecology*, 54:270–275, 2019. doi: <https://doi.org/10.1002/uog.20181>. URL <https://obgyn.onlinelibrary.wiley.com/doi/abs/10.1002/uog.20181>.
- [105] F. van den Noort, C. Manzini, C. H. van der Vaart, M. A. J. Limbeek, C. H. Slump, and A. T. M. Grob. Automatic identification and segmentation of the slice of minimal hiatal dimensions in transperineal ultrasound volumes. *Ultrasound in Obstetrics & Gynecology*, 11 2021. ISSN 0960-7692. doi: 10.1002/uog.24810.
- [106] Helena Williams, Laura Cattani, Wenqi Li, Mahdi Tabassian, Tom Vercauteran, Jan Deprest, and Jan 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.
- [107] Helena Williams, Laura Cattani, Mohammad Yaqub, Carole Sudre, Tom Vercauteran, Jan Deprest, and Jan D’hooge. Automatic C-Plane Detection in Pelvic Floor Transperineal Volumetric Ultrasound. In Roxane Cham, J Alison Noble, Jana Hutter, Stephen Aylward, Andrew Melbourne, Turk Esra Abaci, Torrents Barrena Jordina Hu Yipeng, and Licandro, editors, *Medical Ultrasound, and Preterm, Perinatal and Paediatric Image Analysis*, pages 136–145. Springer International Publishing, 2020. ISBN 978-3-030-60334-2.
- [108] Helena Williams, Laura Cattani, Dominique Van Schoubroeck, Mohammad Yaqub, Carole Sudre, Tom Vercauteran, Jan D’hooge, and Jan Deprest. Automatic Extraction of Hiatal Dimensions in 3-D Transperineal Pelvic Ultrasound Recordings. *Ultrasound in Medicine & Biology*, 47:3470–3479, 12 2021. ISSN 03015629. doi: 10.1016/j.ultrasmedbio.2021.08.009.

- [109] Helena Williams, João Pedrosa, Laura Cattani, Susanne Housmans, Tom Vercauteren, Jan Deprest, and Jan D’hooge. Interactive Segmentation via Deep Learning and B-Spline Explicit Active Surfaces. In Marleen de Bruijne, Philippe C. Cattin, Stéphane Cotin, Nicolas Padoy, Stefanie Speidel, Yefeng Zheng, and Caroline Essert, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, pages 315–325, Cham, 2021. Springer International Publishing. ISBN 978-3-030-87193-2.
- [110] Helena Williams, Laura Cattani, Tom Vercauteren, Jan Deprest, and Jan D’hooge. Automatic Tomographic Ultrasound Imaging Sequence Extraction of the Anal Sphincter. In J. Alison Noble, Stephen Aylward, Alexander Grimwood, Zhe Min, Su-Lin Lee, and Yipeng Hu, editors, *Simplifying Medical Ultrasound*, pages 35–44, Cham, 2021. Springer International Publishing. ISBN 978-3-030-87583-1.
- [111] Neeraj Sharma and Lalit Aggarwal. Automated medical image segmentation techniques. *Journal of medical physics / Association of Medical Physicists of India*, 35:3–14, 5 2010. doi: 10.4103/0971-6203.58777.
- [112] Shenlong Wang, Kaixin Han, and Jiafeng Jin. Review of image low-level feature extraction methods for content-based image retrieval. *Sensor Review*, 39:783–809, 11 2019. ISSN 0260-2288. doi: 10.1108/SR-04-2019-0092.
- [113] Juan Shan, H.D. Cheng, and Yuxuan Wang. A completely automatic segmentation method for breast ultrasound images using region growing. In *Proceedings of the 11th Joint Conference on Information Sciences (JCIS)*. Atlantis Press, 2008. ISBN 9789078677185. doi: 10.2991/jcis.2008.57.
- [114] S. Poonguzhal and G. Ravindran. A complete automatic region growing method for segmentation of masses on ultrasound images. In *2006 International Conference on Biomedical and Pharmaceutical Engineering*, pages 88–92, 2006.
- [115] Lingling Fang, Lirong Zhang, Yibo Yao, and Le Chen. Ultrasound image segmentation using an active contour model and learning-structured inference. *Multimedia Tools and Applications*, 81:13389–13407, 4 2022. ISSN 1380-7501. doi: 10.1007/s11042-021-11088-4.
- [116] Chenyang Xu and J.L. Prince. Snakes, shapes, and gradient vector flow. *IEEE Transactions on Image Processing*, 7:359–369, 3 1998. ISSN 10577149. doi: 10.1109/83.661186.
- [117] Roushanak Rahmat and David Harris-Birtill. Comparison of level set models in image segmentation. *IET Image Processing*, 12:2212–2221, 12 2018. ISSN 1751-9667. doi: 10.1049/iet-ipr.2018.5796.
- [118] Daniel Barbosa, Thomas Dietenbeck, Joël Schaerer, Jan D’hooge, Denis Friboulet, and Olivier Bernard. B-Spline Explicit Active Surfaces: An Efficient Framework for Real-Time 3-D Region-Based Segmentation. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, 21:241–251, 5 2012. doi: 10.1109/TIP.2011.2161484.
- [119] Chuanjiang He, Yan Wang, and Qiang Chen. Active contours driven by weighted region-scalable fitting energy based on local entropy. *Signal Processing*, 92:587–600, 2 2012. ISSN 01651684. doi: 10.1016/j.sigpro.2011.09.004.
- [120] Shawn Lankton and Allen Tannenbaum. Localizing Region-Based Active Contours. *IEEE Transactions on Image Processing*, 17, 5 2008. doi: 10.1109/TIP.2008.2004611.

Chapter 11. Video links

- [121] Chunming Li, Chiu-Yen Kao, J.C. Gore, and Zhaohua Ding. Minimization of Region-Scalable Fitting Energy for Image Segmentation. *IEEE Transactions on Image Processing*, 17:1940–1949, 10 2008. ISSN 1057-7149. doi: 10.1109/TIP.2008.2002304.
- [122] Li Wang, Li Chunming, Sun Quansen, Xia Deshen, and Chiu-Yen Kao. Brain MR Image Segmentation Using Local and Global Intensity Fitting Active Contours/Surfaces. In Dimitris Metaxas, Leon Axel, Gabor Fichtinger, and Gábor Székely, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2008*, pages 384–392, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg. ISBN 978-3-540-85988-8.
- [123] Jing-Jing Zong, Tian-Shuang Qiu, Wei-Dong Li, and Dong-Mei Guo. Automatic ultrasound image segmentation based on local entropy and active contour model. *Computers & Mathematics with Applications*, 78:929–943, 8 2019. ISSN 08981221. doi: 10.1016/j.camwa.2019.03.022.
- [124] Qi Duan, Elsa D. Angelini, and Andrew F. Laine. Real-time segmentation by Active Geometric Functions. *Computer Methods and Programs in Biomedicine*, 98:223–230, 6 2010. ISSN 01692607. doi: 10.1016/j.cmpb.2009.09.001.
- [125] Daniel Barbosa, Thomas Dietenbeck, Brecht Heyde, Helene Houle, Denis Frioulet, Jan D’hooge, and Olivier Bernard. Fast and fully automatic 3D echocardiographic segmentation using B-spline explicit active surfaces. In *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 1088–1091. IEEE, 5 2012. ISBN 978-1-4577-1858-8. doi: 10.1109/ISBI.2012.6235748.
- [126] Helena R. Torres, Sandro Queiros, Pedro Morais, Bruno Oliveira, Joao Gomes-Fonseca, Paulo Mota, Estevao Lima, Jan D’Hooge, Jaime C. Fonseca, and Joao L. Vilaca. Kidney Segmentation in 3-D Ultrasound Images Using a Fast Phase-Based Approach. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 68:1521–1531, 5 2021. ISSN 0885-3010. doi: 10.1109/TUFFC.2020.3039334.
- [127] Joao Pedrosa, Sandro Queirós, Olivier Bernard, Jan Engvall, Thor Edvardsen, Eike Nagel, and Jan Dhooge. Fast and Fully Automatic Left Ventricular Segmentation and Tracking in Echocardiography Using Shape-Based B-Spline Explicit Active Surfaces. *IEEE Transactions on Medical Imaging*, PP:1, 5 2017. doi: 10.1109/TMI.2017.2734959.
- [128] Yuan Wang, Wei Zhuo, Yucong Li, Zhi Wang, Qi Ju, and Wenwu Zhu. Fully Self-Supervised Learning for Semantic Segmentation. *CoRR*, 2022. doi: 10.48550/ARXIV.2202.11981. URL <https://arxiv.org/abs/2202.11981>.
- [129] Mina Amiri, Rupert Brooks, Bahareh Behboodi, and Hassan Rivaz. Two-stage ultrasound image segmentation using U-Net and test time augmentation. *International Journal of Computer Assisted Radiology and Surgery*, 15:981–988, 6 2020. ISSN 1861-6410. doi: 10.1007/s11548-020-02158-3.
- [130] Michal Byra, Piotr Jarosik, Aleksandra Szubert, Michael Galperin, Haydee Ojeda-Fournier, Linda Olson, Mary O’Boyle, Christopher Comstock, and Michael Andre. Breast mass segmentation in ultrasound with selective kernel U-Net convolutional neural network. *Biomedical Signal Processing and Control*, 61:102027, 8 2020. ISSN 17468094. doi: 10.1016/j.bspc.2020.102027.
- [131] Ji Yang, Mehdi Faraji, and Anup Basu. Robust segmentation of arterial walls in intravascular ultrasound images using Dual Path U-Net. *Ultrasonics*, 96:24–33, 7 2019. ISSN 0041624X. doi: 10.1016/j.ultras.2019.03.014.

- [132] Deepthy Mary Alex and D. Abraham Chandy. Investigations on Performances of Pre-trained U-Net Models for 2D Ultrasound Kidney Image Segmentation. In Mahdi H. Miraz, Peter S. Excell, Andrew Ware, Safeeullah Soomro, and Maaruf Ali, editors, *Emerging Technologies in Computing*, pages 185–195, Cham, 2020. Springer International Publishing. ISBN 978-3-030-60036-5.
- [133] Jian Zhu, Will Styler, and Ian C. Calloway. A CNN-based tool for automatic tongue contour tracking in ultrasound images. *CoRR*, abs/1907.10210, 2019.
- [134] Yi Wang, Na Wang, Min Xu, Junxiong Yu, Chenchen Qin, Xiao Luo, Xin Yang, Tianfu Wang, Anhua Li, and Dong Ni. Deeply-Supervised Networks With Threshold Loss for Cancer Detection in Automated Breast Ultrasound. *IEEE Transactions on Medical Imaging*, 39:866–876, 4 2020. ISSN 0278-0062. doi: 10.1109/TMI.2019.2936500.
- [135] Hongxu Yang, Caifeng Shan, Tao Tan, Alexander F. Kolen, and Peter H. N. de With. Transferring from ex-vivo to in-vivo: Instrument Localization in 3D Cardiac Ultrasound Using Pyramid-UNet with Hybrid Loss. In Dinggang Shen, Tianming Liu, Terry M. Peters, Lawrence H. Staib, Caroline Essert, Sean Zhou, Pew-Thian Yap, and Ali Khan, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, pages 263–271, Cham, 2019. Springer International Publishing. ISBN 978-3-030-32254-0.
- [136] Zachary Szentimrey, Sandrine de Ribaupierre, Aaron Fenster, and Eranga Ukwatta. Automatic deep learning-based segmentation of neonatal cerebral ventricles from 3D ultrasound images. In Barjor S. Gimi and Andrzej Krol, editors, *Medical Imaging 2021: Biomedical Applications in Molecular, Structural, and Functional Imaging*, page 24. SPIE, 2 2021. ISBN 9781510640290. doi: 10.1117/12.2581749.
- [137] Hongxu Yang, Caifeng Shan, Alexander F. Kolen, and Peter H. N. de With. Automated Catheter Localization in Volumetric Ultrasound Using 3D Patch-Wise U-Net with Focal Loss. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 1346–1350. IEEE, 9 2019. ISBN 978-1-5386-6249-6. doi: 10.1109/ICIP.2019.8803045.
- [138] Praful Mathur, Karan Kakwani, Diplay, Srinivas Kudavelly, and Ramaraju GA. Deep Learning based Quantification of Ovary and Follicles using 3D Transvaginal Ultrasound in Assisted Reproduction. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 2109–2112. IEEE, 7 2020. ISBN 978-1-7281-1990-8. doi: 10.1109/EMBC44109.2020.9176703.
- [139] Xiaofeng Yang, Yang Lei, Sibo Tian, Tonghe Wang, Bo Wang, Ashesh B. Jani, Pretesh Patel, Hui Mao, Walter J. Curran, Tian Liu, and Xiuxiu He. Ultrasound prostate segmentation based on 3D V-Net with deep supervision. In Nicole V. Ruiter and Brett C. Byram, editors, *Medical Imaging 2019: Ultrasonic Imaging and Tomography*, page 30. SPIE, 3 2019. ISBN 9781510625570. doi: 10.1117/12.2512558.
- [140] Yang Lei, Sibo Tian, Xiuxiu He, Tonghe Wang, Bo Wang, Pretesh Patel, Ashesh B. Jani, Hui Mao, Walter J. Curran, Tian Liu, and Xiaofeng Yang. Ultrasound prostate segmentation based on multidirectional deeply supervised V-Net. *Medical Physics*, 46: 3194–3206, 7 2019. ISSN 0094-2405. doi: 10.1002/mp.13577.
- [141] Yan Zeng, Po-Hsiang Tsui, Weiwei Wu, Zhuhuang Zhou, and Shuicai Wu. Fetal Ultrasound Image Segmentation for Automatic Head Circumference Biometry Using Deeply Supervised Attention-Gated V-Net. *Journal of Digital Imaging*, 34:134–148, 2 2021. ISSN 0897-1889. doi: 10.1007/s10278-020-00410-5.

Chapter 11. Video links

- [142] Fuat Türk, Murat Lüy, and Necattin Barışçı. Kidney and Renal Tumor Segmentation Using a Hybrid V-Net-Based Model. *Mathematics*, 8:1772, 10 2020. ISSN 2227-7390. doi: 10.3390/math8101772.
- [143] Lingling Fang, Xiaohang Pan, Yibo Yao, Lirong Zhang, and Dongmei Guo. A hybrid active contour model for ultrasound image segmentation. *Soft Computing*, 24:18611–18625, 12 2020. ISSN 1432-7643. doi: 10.1007/s00500-020-05097-y.
- [144] Revathy Sivanandan and J Jayakumari. An Improved Ultrasound Tumor Segmentation Using CNN Activation Map Clustering and Active Contours. In *2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA)*, pages 263–268. IEEE, 10 2020. ISBN 978-1-7281-6324-6. doi: 10.1109/ICCCA49541.2020.9250909.
- [145] Yibiao Rong, Dehui Xiang, Weifang Zhu, Fei Shi, Enting Gao, Zhun Fan, and Xinjian Chen. Deriving external forces via convolutional neural networks for biomedical image segmentation. *Biomedical Optics Express*, 10:3800, 8 2019. ISSN 2156-7085. doi: 10.1364/BOE.10.003800.
- [146] Agus Pratondo, Chee-Kong Chui, and Sim-Heng Ong. Integrating machine learning with region-based active contour models in medical image segmentation. *Journal of Visual Communication and Image Representation*, 43:1–9, 2 2017. ISSN 10473203. doi: 10.1016/j.jvcir.2016.11.019.
- [147] Mo Zhang, Bin Dong, and Quanzheng Li. Deep Active Contour Network for Medical Image Segmentation. In Anne L. Martel, Purang Abolmaesumi, Danail Stoyanov, Diana Mateus, Maria A. Zuluaga, S. Kevin Zhou, Daniel Racoceanu, and Leo Joskowicz, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, pages 321–331, Cham, 2020. Springer International Publishing. ISBN 978-3-030-59719-1.
- [148] Yunyun Yang, Chong Feng, and Ruofan Wang. Automatic segmentation model combining U-Net and level set method for medical images. *Expert Systems with Applications*, 153: 113419, 9 2020. ISSN 09574174. doi: 10.1016/j.eswa.2020.113419.
- [149] Tomas Sakinis, Fausto Milletari, Holger Roth, Panagiotis Korfiatis, Petro M Kostandy, Kenneth Philbrick, Zeynettin Akkus, Ziyue Xu, Daguang Xu, and Bradley J Erickson. Interactive segmentation of medical images through fully convolutional neural networks. *CoRR*, abs/1903.08205, 2019. URL <http://arxiv.org/abs/1903.08205>.
- [150] Roberto Ardon, Remi Cuingnet, Ketan Bacchuwar, and Vincent Auvray. Fast kidney detection and segmentation with learned kernel convolution and model deformation in 3D ultrasound images. In *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, pages 268–271. IEEE, 4 2015. ISBN 978-1-4799-2374-8. doi: 10.1109/ISBI.2015.7163865.
- [151] Daniel Barbosa, Brecht Heyde, Maja Cikes, Thomas Dietenbeck, Piet Claus, Denis Friboulet, Olivier Bernard, and Jan D’hooge. Real-Time 3D Interactive Segmentation of Echocardiographic Data Through User-based Deformation of B-Spline Explicit Active Surfaces. *Computerized Medical Imaging and Graphics*, 38:57–67, 5 2014. doi: 10.1016/j.compmedimag.2013.10.002.
- [152] Martin Rajchl, Matthew C H Lee, Ozan Oktay, Konstantinos Kamnitsas, Jonathan Passerat-Palmbach, Wenjia Bai, Bernhard Kainz, and Daniel Rueckert. DeepCut: Object Segmentation from Bounding Box Annotations using Convolutional Neural Networks. *CoRR*, abs/1605.07866, 2016. URL <http://arxiv.org/abs/1605.07866>.

- [153] Guotai Wang, Michael Aertsen, Jan Deprest, Sébastien Ourselin, Tom Vercauteren, and Shaoting Zhang. Uncertainty-Guided Efficient Interactive Refinement of Fetal Brain Segmentation from Stacks of MRI Slices. *CoRR*, pages 279–288, 5 2020. doi: 10.1100/978-3-030-59719-1_28.
- [154] Kevis-Kotiksi Maninis, Sergi Caelles, Jordi Pont-Tuset, and Luc Van Gool. Deep Extreme Cut: From Extreme Points to Object Segmentation. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 616–625. IEEE, 6 2018. ISBN 978-1-5386-6420-9. doi: 10.1109/CVPR.2018.00071.
- [155] Won-Dong Jang and Chang-Su Kim. Interactive Image Segmentation via Backpropagating Refinement Scheme. In *Proceedings of The IEEE Conference on Computer Vision and Pattern Recognition*, 2019.
- [156] Xuan Chen, Boyang Zhou, Linfei Xiong, Chongke Zhao, Lifan Wang, Yongwei Zhang, and Huixiong Xu. Balancing regional and global information: An interactive segmentation framework for ultrasound breast lesion. *Biomedical Signal Processing and Control*, 77: 103723, 8 2022. ISSN 17468094. doi: 10.1016/j.bspc.2022.103723.
- [157] Guotai Wang, Maria A. Zuluaga, Wenqi Li, Rosalind Pratt, Premal A. Patel, Michael Aertsen, Tom Doel, Anna L. David, Jan Deprest, Sébastien Ourselin, and Tom Vercauteren. DeepIGeoS: A Deep Interactive Geodesic Framework for Medical Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41:1559–1572, 7 2019. ISSN 0162-8828. doi: 10.1109/TPAMI.2018.2840695.
- [158] Guotai Wang, Wenqi Li, Maria A Zuluaga, Rosalind Pratt, Premal A Patel, Michael Aertsen, Tom Doel, Anna L David, Jan Deprest, Sébastien Ourselin, and Tom Vercauteren. Interactive Medical Image Segmentation using Deep Learning with Image-specific Fine-tuning. *CoRR*, abs/1710.04043, 2017. URL <http://arxiv.org/abs/1710.04043>.
- [159] Xiangde Luo, Guotai Wang, Tao Song, Jingyang Zhang, Michael Aertsen, Jan Deprest, Sébastien Ourselin, Tom Vercauteren, and Shaoting Zhang. MIDeepSeg: Minimally interactive segmentation of unseen objects from medical images using deep learning. *Medical Image Analysis*, 72:102102, 8 2021. ISSN 13618415. doi: 10.1016/j.media.2021.102102.
- [160] Muhammad Asad, Lucas Fidon, and Tom Vercauteren. ECONet: Efficient Convolutional Online Likelihood Network for Scribble-based Interactive Segmentation. *CoRR*, 2022. doi: 10.48550/ARXIV.2201.04584. URL <https://arxiv.org/abs/2201.04584>.
- [161] Eric N. Mortensen and William A. Barrett. Interactive Segmentation with Intelligent Scissors. *Graphical Models and Image Processing*, 60:349–384, 9 1998. ISSN 10773169. doi: 10.1006/gmip.1998.0480.
- [162] Kiryl Chykeyuk, Mohammad Yaqub, and J. Alison Noble. Class-Specific Regression Random Forest for Accurate Extraction of Standard Planes from 3D Echocardiography. In Bjoern Menze, Georg Langs, Albert Montillo, Michael Kelm, Henning Müller, and Zhuowen Tu, editors, *Medical Computer Vision. Large Data in Medical Imaging*, pages 53–62, Cham, 2014. Springer International Publishing. ISBN 978-3-319-05530-5.
- [163] Leo Breiman. Random Forests. *Machine Learning*, 45:5–32, 2001. ISSN 08856125. doi: 10.1023/A:1010933404324.

Chapter 11. Video links

- [164] Mohammad Yaqub, Anil Kopuri, Sylvia Rueda, Peter Sullivan, Kenneth McCormick, and Julia Noble. A Constrained Regression Forests Solution to 3D Fetal Ultrasound Plane Localization for Longitudinal Analysis of Brain Growth and Maturation. In *Mach. Learn. Med. Imaging*, volume LNCS 8679, pages 109–116, 5 2014. doi: 10.1007/978-3-319-10581-9_14.
- [165] Yuanwei Li, Bishesh Khanal, Benjamin Hou, Amir Alansary, Juan J Cerrolaza, Matthew Sinclair, Jacqueline Matthew, Chandni Gupta, Caroline Knight, Bernhard Kainz, and et al. Standard Plane Detection in 3D Fetal Ultrasound Using an Iterative Transformation Network. *Lecture Notes in Computer Science*, pages 392–400, 2018. ISSN 1611-3349. doi: 10.1007/978-3-030-00928-1_45. URL http://dx.doi.org/10.1007/978-3-030-00928-1_45.
- [166] Peifei Zhu and Zisheng Li. Guideline-Based Machine Learning for Standard Plane Extraction in 3D Cardiac Ultrasound. In Henning Müller, B. Michael Kelm, Tal Arbel, Weidong Cai, M. Jorge Cardoso, Georg Langs, Bjoern Menze, Dimitris Metaxas, Albert Montillo, William M. Wells III, Shaoting Zhang, Albert C.S. Chung, Mark Jenkinson, and Annemie Ribbens, editors, *Medical Computer Vision and Bayesian and Graphical Models for Biomedical Imaging*, pages 137–147, Cham, 2017. Springer International Publishing. ISBN 978-3-319-61188-4.
- [167] Eli Gibson, Wenqi Li, Carole H Sudre, Lucas Fidon, Dzhoshkun I Shakir, Guotai Wang, Zach Eaton-Rosen, Robert Gray, Tom Doel, Yipeng Hu, Tom Whyntie, Parashkev Nachev, Dean C Barratt, Sébastien Ourselin, M. Jorge Cardoso, and Tom Vercauteren. NiftyNet: a deep-learning platform for medical imaging. *CoRR*, abs/1709.03485, 2017. URL <http://arxiv.org/abs/1709.03485>.
- [168] Konstantinos Kamnitsas, Christian Ledig, Virginia F.J. Newcombe, Joanna P. Simpson, Andrew D. Kane, David K. Menon, Daniel Rueckert, and Ben Glocker. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Medical Image Analysis*, 36:61–78, 2 2017. ISSN 13618415. doi: 10.1016/j.media.2016.10.004.
- [169] Lorenz Berger, Hyde Eoin, M. Jorge Cardoso, and Sébastien Ourselin. An Adaptive Sampling Scheme to Efficiently Train Fully Convolutional Networks for Semantic Segmentation. In Mark Nixon, Sasan Mahmoodi, and Reyer Zwiggelaar, editors, *Medical Image Understanding and Analysis*, pages 277–286, Cham, 2018. Springer International Publishing. ISBN 978-3-319-95921-4.
- [170] Carole H. Sudre, Wenqi Li, Tom Vercauteren, Sébastien Ourselin, and M. Jorge Cardoso. Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 240–248. Springer International Publishing, 2017. doi: 10.1007/978-3-319-67558-9_28. URL https://doi.org/10.1007%2F978-3-319-67558-9_28.
- [171] D.P. Huttenlocher, G.A. Klanderman, and W.J. Rucklidge. Comparing images using the Hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15 (9):850–863, 1993. doi: 10.1109/34.232073.
- [172] Yunku Kang, Seung-Hyun Yoon, Min-Ho Kyung, and Myung-Soo Kim. Fast and robust computation of the Hausdorff distance between triangle mesh and quad mesh for near-zero cases. *Computers & Graphics*, 81:61–72, 6 2019. ISSN 00978493. doi: 10.1016/j.cag.2019.03.014.

- [173] Dong-Gyu Sim, Oh-Kyu Kwon, and Rae-Hong Park. Object matching algorithms using robust Hausdorff distance measures. *IEEE Transactions on Image Processing*, 8:425–429, 3 1999. ISSN 10577149. doi: 10.1109/83.748897.
- [174] Zach Eaton-Rosen, Felix J. S. Bragman, Sébastien Ourselin, and M. Jorge Cardoso. Improving Data Augmentation for Medical Image Segmentation. In *1st Conference on Medical Imaging with Deep Learning (MIDL)*, 2018.
- [175] Hongyi Zhang, Moustapha Cissé, Yann N Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. *CoRR*, abs/1710.09412, 2017. URL <http://arxiv.org/abs/1710.09412>.
- [176] Karin Lammers, Jurgen Jacobus Fütterer, Joanna Inthout, Mathias Prokop, Mark Ewout Vierhout, and Kirsten Birgit Kluivers. Correlating signs and symptoms with pubo-visceral muscle avulsions on magnetic resonance imaging. *American Journal of Obstetrics & Gynecology*, 208:148.e1 – 148.e7, 2013. ISSN 0002-9378. doi: <https://doi.org/10.1016/j.ajog.2012.12.006>. URL <http://www.sciencedirect.com/science/article/pii/S0002937812022016>.
- [177] AIUM/IUGA. 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*, 38:851–864, 4 2019. ISSN 0278-4297. doi: 10.1002/jum.14953.
- [178] Xiaoguang Lu, Bogdan Georgescu, Yefeng Zheng, Joanne Otsuki, and Dorin Comaniciu. AutoMPR: Automatic detection of standard planes in 3D echocardiography. In *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Proceedings, ISBI*, pages 1279 – 1282, 5 2008. doi: 10.1109/ISBI.2008.4541237.
- [179] Y Xuan, T Friedman, and H P Dietz. Does levator ani hiatal area configuration affect pelvic organ prolapse? *Ultrasound in Obstetrics & Gynecology*, 54:124—127, 7 2019. ISSN 0960-7692. doi: 10.1002/uog.20210. URL <https://doi.org/10.1002/uog.20210>.
- [180] Susan V Brooks, Eileen Zerba, and John A Faulkner. Injury to muscle fibres after single stretches of passive and maximally stimulated muscles in mice. *The Journal of physiology*, 488:459–469, 1995.
- [181] Kuo-Cheng Lien, Brian Mooney, John O L DeLancey, and James A Ashton-Miller. Levator ani muscle stretch induced by simulated vaginal birth. *Obstetrics & Gynecology*, 103:31, 2004.
- [182] Ka Lai Shek and Hans P. Dietz. Intrapartum risk factors for levator trauma. *BJOG: An International Journal of Obstetrics & Gynaecology*, 117:1485–1492, 2010.
- [183] Kuntal K. Pal and K. S. Sudeep. Preprocessing for image classification by convolutional neural networks. In *2016 IEEE International Conference on Recent Trends in Electronics, Information Communication Technology (RTEICT)*, pages 1778–1781, 2016. doi: 10.1109/RTEICT.2016.7808140.
- [184] Andriy Fedorov, Reinhard Beichel, Jayashree Kalpathy-Cramer, Julien Finet, Jean-Christophe Fillion-Robin, Sonia Pujol, Christian Bauer, Dominique Jennings, Fiona Fennessy, Milan Sonka, John Buatti, Stephen Aylward, James V Miller, Steve Pieper, and Ron Kikinis. 3D Slicer as an image computing platform for the Quantitative Imaging

Chapter 11. Video links

- Network. *Magnetic Resonance Imaging*, 30:1323–1341, 2012. ISSN 0730-725X. doi: <https://doi.org/10.1016/j.mri.2012.05.001>. URL <https://www.sciencedirect.com/science/article/pii/S0730725X12001816>.
- [185] Saeid Asgari Taghanaki, Kumar Abhishek, Joseph Paul Cohen, Julien Cohen-Adad, and Ghassan Hamarneh. Deep Semantic Segmentation of Natural and Medical Images: A Review. *CoRR*, abs/1910.07655, 2019. URL <http://arxiv.org/abs/1910.07655>.
- [186] Dzung L. Pham, Chenyang Xu, and Jerry L. Prince. Current Methods in Medical Image Segmentation. *Annual Review of Biomedical Engineering*, 2:315–337, 2000. doi: 10.1146/annurev.bioeng.2.1.315. URL <https://doi.org/10.1146/annurev.bioeng.2.1.315>. PMID: 11701515.
- [187] Stan Benjamins, Pranavsingh Dhunnoo, and Bertalan Mesko. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. *npj Digital Medicine*, 3, 5 2020. doi: 10.1038/s41746-020-00324-0.
- [188] Hans P. Dietz, Fernandi Moegni, and Ka Lai Shek. Diagnosis of levator avulsion injury: a comparison of three methods. *Ultrasound in Obstetrics & Gynecology*, 40:693–698, 2012. doi: <https://doi.org/10.1002/uog.11190>. URL <https://obgyn.onlinelibrary.wiley.com/doi/abs/10.1002/uog.11190>.
- [189] Subbiahpillai N. Kumar, Alfred L. Fred, and Paul S. Varghese. An Overview of Segmentation Algorithms for the Analysis of Anomalies on Medical Images. *Journal of Intelligent Systems*, 29:612–625, 2020. doi: doi:10.1515/jisys-2017-0629. URL <https://doi.org/10.1515/jisys-2017-0629>.
- [190] Hiba Ramadan, Chaymae Lachqar, and H Tairi. A survey of recent interactive image segmentation methods. *Computational Visual Media*, 5 2020. doi: 10.1007/s41095-020-0177-5.
- [191] Di Lin, Jifeng Dai, Jiaya Jia, Kaiming He, and Jian Sun. ScribbleSup: Scribble-Supervised Convolutional Networks for Semantic Segmentation. *CoRR*, abs/1604.05144, 2016. URL <http://arxiv.org/abs/1604.05144>.
- [192] Olivier Bernard, Denis Friboulet, Philippe Thevenaz, and Michael Unser. Variational B-Spline Level-Set: A Linear Filtering Approach for Fast Deformable Model Evolution. *IEEE Transactions on Image Processing*, 18:1179–1191, 2009. doi: 10.1109/TIP.2009.2017343.
- [193] Michael Unser. Splines: A Perfect Fit for Signal and Image Processing. *IEEE Signal Processing Magazine*, pages 22 – 38, 11 1999.
- [194] Anthony Yezzi, Andy Tsai, and Alan Willsky. A Fully Global Approach to Image Segmentation via Coupled Curve Evolution Equations. *Journal of Visual Communication and Image Representation*, 13:195–216, 2002.
- [195] Sandra G. Hart. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50:904–908, 2006. doi: 10.1177/154193120605000909. URL <https://doi.org/10.1177/154193120605000909>.
- [196] Kai Virtanen, Heikki Mansikka, Helmiina Kontio, and Don Harris. Weight watchers: NASA-TLX weights revisited. *Theoretical Issues in Ergonomics Science*, pages 1–24, 11 2021. ISSN 1463-922X. doi: 10.1080/1463922X.2021.2000667.

- [197] Leo Grady. Minimal Surfaces Extend Shortest Path Segmentation Methods to 3D. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32:321–334, 2 2010. ISSN 0162-8828. doi: 10.1109/TPAMI.2008.289.
- [198] Andrew Pressley. Gauss’ Theorema Egregium. In *Elementary Differential Geometry*, pages 247–268. Springer London, London, 2010. ISBN 978-1-84882-891-9. doi: 10.1007/978-1-84882-891-9_10. URL https://doi.org/10.1007/978-1-84882-891-9_10.
- [199] Isabel Benavente-Fernandez, Manuel Lubián-Gutierrez, Gema Jimenez-Gomez, Alfonso M. Lechuga-Sancho, Simon P. Lubián-López, and Neonatal Neurology Foundation (Fundación Nene). Ultrasound lineal measurements predict ventricular volume in posthaemorrhagic ventricular dilatation in preterm infants. *Acta Paediatrica*, 106(2): 211–217, 2017. doi: <https://doi.org/10.1111/apa.13645>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/apa.13645>.
- [200] Lionel Gontard, Joaquín Pizarro, Borja Sanz-Peña, S. López, and Isabel Benavente-Fernández. Automatic segmentation of ventricular volume by 3D ultrasonography in post haemorrhagic ventricular dilatation among preterm infants. *Scientific Reports*, 11, 01 2021. doi: 10.1038/s41598-020-80783-3.
- [201] MONAI Consortium. MONAI: Medical Open Network for AI, 3 2020. URL <https://github.com/Project-MONAI/MONAI>.
- [202] Pedro Furtado. Loss, post-processing and standard architecture improvements of liver deep learning segmentation from Computed Tomography and magnetic resonance. *Informatics in Medicine Unlocked*, 24:100585, 2021. ISSN 23529148. doi: 10.1016/j.imu.2021.100585.
- [203] Eva M. Franco, Cristina Ros, Giulio A. Santoro, Jordi C. Garriga, Lluís A. Tardieu, Daniel Cuadras, and Montserrat Espuña. Transperineal anal sphincter complex evaluation after obstetric anal sphincter injuries: With or without tomographic ultrasound imaging technique? *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 257:70–75, 2021. ISSN 0301-2115. doi: <https://doi.org/10.1016/j.ejogrb.2020.12.009>. URL <https://www.sciencedirect.com/science/article/pii/S0301211520307934>.
- [204] Andrea Stuart, Claes Ignell, and Ann-Kristin Örnö. Comparison of transperineal and endoanal ultrasound in detecting residual obstetric anal sphincter injury. *Acta Obstetricia et Gynecologica Scandinavica*, 98:1624–1631, 2019. doi: <https://doi.org/10.1111/aogs.13701>. URL <https://obgyn.onlinelibrary.wiley.com/doi/abs/10.1111/aogs.13701>.
- [205] Ka Shek, Vincent Zazzera, Ixora Kamisan Atan, Rodrigo Guzmán Rojas, Susanne Langer, and Hans Dietz. The evolution of transperineal ultrasound findings of the external anal sphincter during the first years after childbirth. *International Urogynecology Journal*, 27, 5 2016. doi: 10.1007/s00192-016-3055-z.
- [206] Vivian Bass, Julieta Mateos, Ivan M. Rosado-Mendez, and Jorge Márquez. Ultrasound image segmentation methods: A review. In *AIP Conference Proceedings 2348*, 050018, page 050018, 2021. doi: 10.1063/5.0051110.
- [207] Jordi Cassadó, Marta Simó, Nuria Rodríguez, Oriol Porta, Eva Huguet, Irene Mora, Marta Girvent, Rebeca Fernández, and Ignasi Gich. Prevalence of levator ani avulsion in a multicenter study (PAMELA study). *Archives of Gynecology and Obstetrics*, 302: 273–280, 7 2020. ISSN 0932-0067. doi: 10.1007/s00404-020-05585-4.

Chapter 11. Video links

- [208] Jasper Tromp, David Bauer, Brian L. Claggett, Matthew J. Frost, Matthias B. Iversen, Narayana Prasad, Mark Petrie, Martin G. Larson, Justin A. Ezekowitz, and Scott D. Solomon. A prospective validation of a deep learning-based automated workflow for the interpretation of the echocardiogram. *European Heart Journal - Cardiovascular Imaging*, 23, 2 2022. ISSN 2047-2404. doi: 10.1093/ehjci/jeab289.002.
- [209] Abdul H. Sultan. Editorial: Obstetrical Perineal Injury and Anal Incontinence. *AVMA Medical & Legal Journal*, 5:193–196, 11 1999. ISSN 0959-9304. doi: 10.1177/135626229900500601.
- [210] Lingyun Wu, Jie Zhi Cheng, Shengli Li, Baiying Lei, Tianfu Wang, and Dong Ni. FUIQA: Fetal ultrasound image quality assessment with deep convolutional networks. *IEEE Transactions on Cybernetics*, 47, 2017. ISSN 21682267. doi: 10.1109/TCYB.2017.2671898.
- [211] David Pasdeloup, Sindre Hellum Olaisen, Andreas Østvik, Sigbjørn Sæbø, Håkon Neergaard Pettersen, Espen Holte, Bjørnar Grenne, Stian Bergseng Stølen, Erik Smistad, Svein Arne Aase, Håvard Dalen, and Lasse Lovstakken. Real-Time Echocardiography Guidance for Optimized Apical Standard Views. *TechRxiv*, 6 2022. doi: 10.36227/techrxiv.20180012.v1. URL https://www.techrxiv.org/articles/preprint/Real-Time_Echocardiography_Guidance_for_Optimized_Apical_Standard_VIEWS/20180012.
- [212] Baljash S. Cheema, James Walter, Akhil Narang, and James D. Thomas. Artificial Intelligence–Enabled POCUS in the COVID-19 ICU: A New Spin on Cardiac Ultrasound. *JACC: Case Reports*, 3, 2021. ISSN 26660849. doi: 10.1016/j.jaccas.2020.12.013.
- [213] Karim Lekadir, Gianluca Quaglio, Anna Tselioudis Garmendia, and Catherine Gallin. Artificial intelligence in healthcare; Applications, risks, and ethical and societal impacts. Technical Report PE 729.512, ERPS: European Parliamentary Research Service, Brussels, European Union, 6 2022.
- [214] Hao Guan and Mingxia Liu. Domain Adaptation for Medical Image Analysis: A Survey. *IEEE Transactions on Biomedical Engineering*, 69, 2022. ISSN 15582531. doi: 10.1109/TBME.2021.3117407.
- [215] Wenjun Yan, Yuanyuan Wang, Shengjia Gu, Lu Huang, Fuhua Yan, Liming Xia, and Qian Tao. The Domain Shift Problem of Medical Image Segmentation and Vendor-Adaptation by Unet-GAN. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 11765 LNCS, 2019. doi: 10.1007/978-3-030-32245-8_69.
- [216] CORE-MD. Coordinating Research and Evidence for Medical Devices. About the initiative., 2022. URL <https://www.core-md.eu/about-the-initiative>.
- [217] Kincso Izsak and Apolline Terrier. Advanced Technologies for Industry - Product Watch. Artificial Intelligence-based software as a medical device. Technical report, European Commission, Brussels, European Union, 7 2020.
- [218] COCIR. Artificial Intelligence in EU Medical Device Legislation. Technical report, COCIR, the European Coordination Committee of the Radiological, Electromedical and Healthcare IT Industry, Brussels, Belgium, May 2021.

- [219] Nicola Rieke, Jonny Hancock, Wenqi Li, Fausto Milletari, Holger R. Roth, Shadi Albarqouni, Spyridon Bakas, Mathieu N. Galtier, Bennett A. Landman, Klaus Maier-Hein, Sébastien Ourselin, Micah Sheller, Ronald M. Summers, Andrew Trask, Daguang Xu, Maximilian Baust, and M. Jorge Cardoso. The future of digital health with federated learning. *npj Digital Medicine*, 3, 2020. ISSN 23986352. doi: 10.1038/s41746-020-00323-1.
- [220] Tian Li, Anit Kumar Sahu, Ameet Talwalkar, and Virginia Smith. Federated Learning: Challenges, Methods, and Future Directions. *IEEE Signal Processing Magazine*, 37, 2020. ISSN 15580792. doi: 10.1109/MSP.2020.2975749.
- [221] MDCG. MDCG 2019-11 Guidance on Qualification and Classification of Software in Regulation (EU) 2017/745 – MDR and Regulation (EU) 2017/746 – IVDR . Technical report, Medical Device Coordination Group, Brussels, Belgium, 10 2019.
- [222] Wimmy Choi, Marlies van Eck, Cecile van der Heijden, Theo Hooghiemstra, and Erik Vollebregt. Legal analysis- European legislative proposal draft AI act and MDR/IVDR. Technical report, Hooghiemstra & Partner and AXON science based lawyers, 1 2022.