Preserving Known Anatomical Topology in Medical Image Segmentation Using Deep Learning

Madeleine K Wyburd

Hertford College
University of Oxford

Submitted in partial fulfilment of the degree of Doctor of Philosophy

Michaelmas 2022

Abstract

Since the rise of deep learning, new medical image segmentation methods have rapidly been proposed with promising results, with each one reporting marginal improvements on the previous state-of-the-art (SOTA) method. However, on visual inspection, errors are often revealed, such as topological mistakes (e.g. holes or folds), that are not detected using traditional evaluation metrics, such as Dice. Moreover, correct topology is often essential in ensuring segmentations are anatomically and pathologically plausible and, ultimately, suitable for downstream image processing tasks. Therefore, there is a need for methods to focus on ensuring that the predicted segmentations are topologically correct, rather than just optimising the pixel-wise accuracy. In this thesis, I propose a method that utilises prior knowledge of anatomy to segment structures, whilst preserving the known topology.

The presented model, Topology Encouraging Deformation Segmentation Network (TEDS-Net), performs segmentation by deforming a prior shape with the same topological features as the anatomy of interest using learnt topology-preserving deformation fields. However, here I show that such fields only guarantee topology preservation in the continuous domain and that their properties begin to break down when applied in discrete spaces. To overcome this effect, I introduced additional modifications in TEDS-Net to more strictly enforce topology preservation, a step often overlooked in previous work.

Across this thesis, TEDS-Net is applied to a range of natural and medical image segmentation tasks. I show how it can be used for multiple topology types, multiple structures and in both two- and three-dimensions. Further, I show how TEDS-Net can be used to segment whole volumes using minimally annotated training data. Across these experiments, TEDS-Net outperforms all SOTA baselines on topology, whilst maintaining competitive pixel-wise accuracy.

Finally, TEDS-Net is integrated into a whole medical imaging pipeline, to illustrate the importance of topologically correct segmentations for downstream tasks. TEDS-Net is used to segment the developing cortical plate from in-utero fetal brain ultrasound scans in 3D, to enable the characterisation of its complex growth and development during gestation. To the best of my knowledge, this task has only been previously attempted from magnetic resonance imaging (MRI), despite ultrasound being the modality of choice in prenatal care. This is likely due to large acoustic shadows obstructing key brain regions in ultrasound. Due to TEDS-Net anatomical constraints, it can anatomically guide the cortical plate segmentation in regions of shadows, producing a complete segmentation that enables accurate downstream analysis.

Preserving Known Anatomical Topology in Medical Image Segmentation Using Deep Learning



Madeleine K Wyburd

Hertford College

University of Oxford

Submitted in partial fulfilment of the degree of

Doctor of Philosophy

Michaelmas 2022

This thesis was written in the Computer Science Department, Hertford Library, the Radcliffe Camera, Vere Harmsworth Library, Oxford Botanical Gardens, Herne Hill, Broughton Astley, Brighton, Bristol, After Glastonbury, Devizes, Ely, The Thameslink: London $\rightarrow$ Brighton (and back), The Oxford Tube, Society Cafe, Missing Bean, those lovely Great Western Trains to Paddington, on my bike, in my head and everywhere else.

Acknowledgements

Although I am obliged to say that this thesis was written by me (which it was, and I have tendonitis to prove it) I have been helped by so many people along the way, whom I would like to thank.

Firstly, to my two brilliant supervisors, Ana and Mark. When I came to Oxford, I didn't know the first thing about fetuses, deep learning or topology, and you both invested so much time and energy into teaching me, which I am so grateful for. Perhaps more though, you have both been so supportive of me personally, I have genuinely enjoyed my whole DPhil, which is a credit to you both. Ana, you have been a fantastic and inspiring mentor throughout, I am so excited to continue our work together. Mark, thank you for teaching me everything I know about the curious world of deformations and statistics. Your ability to teach the most complex topics in such a simple way will always amaze me. Most importantly, thank you for always kindly pointing out my outrageous spelling mistakes[1].

Throughout my DPhil, I have also been supported by so many fabulous colleagues, who have become great pals (I hope you feel the same). To my Nambureteam, thank you for sharing ideas, solving my problems and putting up with my incessant topology jokes. Special mentions to Nicola, for stepping in as my temporary supervisor, despite being in the final 6 months of your own PhD; Felipe, for always fixing my PC before I have even had time to break it; Hugo for always being one step ahead of arXiv (and having the best grin) and of course Linde, although our project didn't make it into my thesis, you certainly had a huge impact on my DPhil. Our grumbling stomachs have sparked a wonderful friendship, thank you for keeping me sane. The MC K Cajones, you made my brief time at the IBME so enjoyable and you've forever changed how I look at laybys. Huge mentions to Eoin, Andy, Stef, Olly and Nav.

Whilst doing my DPhil I have surrounded myself with some fabulous people. Thank you OxWoCS, ONBI and Oxpals for all the laughs and support. A large and honourable mention to Rachel, who has made my write-up much less painful.

I want to thank all my friends outside of Oxford, who have patiently put up with me ignoring messages and becoming a hermit over the last four years[2]. A special

[1]Some of my favourites: disturbed → distributed, hallow → hollow, mythological → morphological

[2]and Jamie, who was less patient but always supportive.

shout out to Lara, Georgie and Becky, my greatest of pals. To Attie and Lucy, who have both saved and endangered my life countless times - without you I may or may not be alive, so there is a chance you deserve a spot here. My grandmothers, Anne and Sue, two of my biggest cheerers, hearing about both of your Oxford antics always brings me joy. To all my sisters, Ellie, Sophie, Phoebe and Mimi, and our current iteration of partners (December 2022), I hope you all stick around. Wine-fuelled Sundays at Stradella have definitely contributed in one way or another to the making of this. Bernie, you are the greatest of men and have made everything better.

To George, I'm so grateful that it was you who found me in that bush on my first night in Oxford, who knew that's how the best of friendships would start? Perhaps I could have done this without you, but it certainly would have been less fun. Thank you for my endless teas, oven dinners, and for always being there, I promise to do the same for you.

Dr Luke, your infectious enthusiasm to learn is what led me to do this DPhil in the first place, thank you for being here every step of the way and for everything else. Without you, this thesis would certainly contain many more typos and perhaps more tear-stained pages.

And finally to my fantastic parents, Kate and Francis. You both have always been my biggest cheerleaders and my constant emotional support. You are the ones who have really made this possible, this work is dedicated to you.

Abstract

Since the rise of deep learning, new medical image segmentation methods have rapidly been proposed with promising results, with each one reporting marginal improvements on the previous state-of-the-art (SOTA) method. However, on visual inspection, errors are often revealed, such as topological mistakes (e.g. holes or folds), that are not detected using traditional evaluation metrics, such as Dice. Moreover, correct topology is often essential in ensuring segmentations are anatomically and pathologically plausible and, ultimately, suitable for downstream image processing tasks. Therefore, there is a need for methods to focus on ensuring that the predicted segmentations are topologically correct, rather than just optimising the pixel-wise accuracy. In this thesis, I propose a method that utilises prior knowledge of anatomy to segment structures, whilst preserving the known topology.

The presented model, Topology Encouraging Deformation Segmentation Network (TEDS-Net), performs segmentation by deforming a prior shape with the same topological features as the anatomy of interest using learnt topology-preserving deformation fields. However, here I show that such fields only guarantee topology preservation in the continuous domain and that their properties begin to break down when applied in discrete spaces. To overcome this effect, I introduced additional modifications in TEDS-Net to more strictly enforce topology preservation, a step often overlooked in previous work.

Across this thesis, TEDS-Net is applied to a range of natural and medical image segmentation tasks. I show how it can be used for multiple topology types, multiple structures and in both two- and three-dimensions. Further, I show how TEDS-Net can be used to segment whole volumes using minimally annotated training data. Across these experiments, TEDS-Net outperforms all SOTA baselines on topology, whilst maintaining competitive pixel-wise accuracy.

Finally, TEDS-Net is integrated into a whole medical imaging pipeline, to illustrate the importance of topologically correct segmentations for downstream tasks. TEDS-Net is used to segment the developing cortical plate from in-utero fetal brain ultrasound scans in 3D, to enable the characterisation of its complex growth and development during gestation. To the best of my knowledge, this task has only been previously attempted from magnetic resonance imaging (MRI), despite ultrasound being the modality of choice in prenatal care. This is likely

due to large acoustic shadows obstructing key brain regions in ultrasound. Due to TEDS-Net anatomical constraints, it can anatomically guide the cortical plate segmentation in regions of shadows, producing a complete segmentation that enables accurate downstream analysis.

Contents

List of Figures **xiii**

List of Tables **xvii**

List of Abbreviations **xxi**

Notation **xxii**

1 Introduction **1**
1.1 Motivation 1
1.2 How This Thesis Came To Be 4
1.3 Contributions 5
1.4 Thesis Layout 6
1.5 Originality 8
1.6 List of Publications 10

2 Review of Medical Segmentations and Topology **13**
2.1 Segmentation Methods 14
2.1.1 Semantic Segmentation 14
2.1.2 Traditional Segmentation Methods 14
2.1.3 Deep learning Segmentation 17
2.2 Segmentation Assessment 24
2.2.1 Segmentation Quality 28
2.2.2 Loss Functions 30
2.3 Topology in Medical Segmentation 31
2.3.1 Pixel labelling 31
2.3.2 Segmenting through Deformation 33
2.4 Conclusion 42

3 Review of the Fetal Cortical Plate **43**
3.1 Fetal Brain Development 44
3.1.1 The Cortical Plate 45
3.1.2 Cortical Folding 48

3.1.3 The Sylvian Fissure . . . 52
3.1.4 Abnormal Gyrification . . . 53
3.2 Clinical Assessment . . . 55
3.2.1 Routine Care . . . 55
3.2.2 Ultrasound Imaging . . . 57
3.2.3 Fissure Screening . . . 58
3.3 Assessing the Developing Cortex . . . 62
3.3.1 Segmenting the Cortical Plate . . . 63
3.3.2 Cortical Analysis . . . 69
3.4 Conclusion . . . 75

4 Topology Preserving Segmentation Method 77
4.1 Overview . . . 77
4.2 Introduction . . . 79
4.3 Theory . . . 81
4.3.1 Topology . . . 81
4.3.2 Topology-preserving fields . . . 84
4.4 Method . . . 86
4.4.1 Problem Definition . . . 86
4.4.2 Architectural Design . . . 86
4.4.3 Loss Functions . . . 90
4.4.4 Evaluation Metrics . . . 91
4.5 Experimental Set-up . . . 92
4.5.1 Dataset . . . 93
4.5.2 Implementation details . . . 93
4.5.3 Experiments . . . 94
4.6 Results and Discussion . . . 97
4.6.1 Ablation Analysis . . . 97
4.6.2 Hyper-parameter Exploration . . . 101
4.6.3 Prior exploration . . . 107
4.6.4 Baseline Comparison . . . 108
4.6.5 Myocardium Perimeter . . . 112
4.6.6 Limitations . . . 113
4.7 Conclusion . . . 115

5 Application of TEDS-Net **117**
5.1 Overview . . . 117
5.2 Introduction . . . 118
5.3 Multi-Prior Segmentation . . . 119
5.3.1 Methods . . . 119
5.3.2 Results . . . 121
5.3.3 Discussion . . . 125
5.4 Multi-Structure Segmentation . . . 127
5.4.1 Methods . . . 128
5.4.2 Results . . . 132
5.4.3 Discussion . . . 134
5.5 3D Liver and Few Slice Labelling . . . 135
5.5.1 Methods . . . 136
5.5.2 Results . . . 143
5.5.3 Discussion . . . 147
5.6 Conclusion . . . 148

6 Characterising the Healthy Cortical Plate Development **151**
6.1 Overview . . . 151
6.2 Introduction . . . 152
6.2.1 Contributions . . . 154
6.3 Image Dataset . . . 156
6.3.1 INTERGROWTH-21st . . . 156
6.3.2 Ultrasound Atlas . . . 157
6.4 Segmenting the Cortical Plate . . . 160
6.4.1 Methods . . . 160
6.4.2 Segmentation Results . . . 167
6.5 Morphological Measurements . . . 172
6.5.1 Global Properties . . . 173
6.5.2 Local Properties . . . 178
6.6 Healthy Population Averages . . . 180
6.6.1 Methods . . . 180
6.6.2 Results and Discussion . . . 182
6.7 Limitations . . . 195
6.8 Conclusion . . . 197

7 Conclusion **199**
7.1 Summary . . . 199
7.2 Contributions . . . 200
7.3 Limitations and Future Work . . . 203

Appendices

A Baseline Training **211**
A.1 VoxelMorph 211
A.2 TopoLoss 212

B Statistical Tests **215**

C Myocaridum Segmentation Examples **219**

D Cortical Plate Segmentations **223**
D.1 Initial Segmentation Methods 223
D.2 CP Growth on new data 224
D.3 Domain Adaption 224

References **227**

List of Figures

1.1 The role of segmentations in medical imaging pipelines. 2

2.1 An schematic of supervised deep learning segmentation networks. . 17
2.2 The building blocks of convolutions neural networks (CNN). 18
2.3 An overview of the popular U-Net architecture 21
2.4 An overview of transformer networks 23
2.5 An overview of topology and Betti numbers for 2D and 3D shapes. . 27
2.6 A set of example segmentations 29
2.7 An overview of spatial transformer networks 33
2.8 Displacement and Deformations fields in 1D 35
2.9 Potential topology changes through resampling 39

3.1 The appearance of the fetal brain across gestation 44
3.2 The organisation of the migrating neurones 45
3.3 Fetal brain terminology 46
3.4 The distinct brain layers across three modalities 47
3.5 The emergence of primary and secondary sulci, gyri and fissures . . 49
3.6 Routine fetal brain ultrasound assessment 56
3.7 Using ultrasound to image the fetal brain 58
3.8 Assessment of the Sylvian fissure across gestation 59
3.9 The cortical plate volume and boundaries in MRI and US 69
3.10 Complex cortical growth across the second trimester 71

4.1 Topological errors found when segmenting with CNNs. 79
4.2 Schematic of different types of connectivity 81
4.3 An overview of measuring topology from 2D digital images. 83
4.4 Limitations of performing Scaling and Squaring in the discrete domain. 84
4.5 Schematic of TEDS-Net architecture 87
4.6 Visualisation of the scaling and squaring approach. 89
4.7 Myocaridum segmentation metrics 91
4.8 Prior shapes used in myocardium segmentation 95
4.9 Qualitative examples of the ablation study. 99

4.10 The effect that σ within the Gaussian smoothing kernel played on segmentation performance . . . 102
4.11 Investigation the weighting of the field regularisation loss term. . . . 103
4.12 Segmentation performance and percentage of folding pixels achieved under different interpolation methods. . . . 105
4.13 The effect that the number of integration layers had on segmentation performance and the topology-preserving nature of the generated fields 106
4.14 The effect that the radius of the prior (r) had on segmentation performance and the nature of the generated fields. . . . 107
4.15 Qualitative performance of myocardium segmentation for TEDS-Net and across the baseline comparison methods for four different subjects 109
4.16 The relative error of automated myocardium perimeter measurements across the five methods. . . . 112
4.17 Qualitative examples to show TEDS-Net over-segmenting in regions of thin boundaries. . . . 114
4.18 The effect of smoothing on perimeter and volume measures. . . . 115

5.1 MNIST digits and their assigned prior shapes . . . 120
5.2 MNIST segmentation performance across the digits when using the two sets of priors. . . . 122
5.3 Qualitative MNIST segmentation performance . . . 123
5.4 A statistical comparison of segmentation performance between digits segmented with the same prior. . . . 124
5.5 Examples of where TEDS-Net has not preserved the topology of the prior . . . 124
5.6 Results obtained by replacing the prior shape at inference time. . . 126
5.7 Topology of cardiac imaging in 2D and 3D. . . . 128
5.8 Priors used in multi-structure cardiac segmentation . . . 129
5.9 Multi-structure segmentation when using different priors. . . . 131
5.10 Qualitative examples of cardiac multi-structure segmentation. . . . 133
5.11 An overview of the three TEDS-Net training strategies for minimal liver labels. . . . 138
5.12 Comparing the liver segmentation performance when training TEDS-Net using the different labelling scenarios. . . . 142
5.13 Qualitative comparison of the Liver segmentation using the two fully supervised approaches and TEDS-Net methods that use a single labelled slice. . . . 146

6.1 An overview of the CP analysis pipeline . . . 155

6.2 Qualitative overview of scans included and those excluded based on quality. . . . 159
6.3 An overview of the generation of the INTERGROWTH-21st weekly atlas. . . . 159
6.4 An overview of the pipeline used to generate the ground-truth cortical plate labels. . . . 161
6.5 Cortical plate labelling protocol for each of the weekly ultrasound atlases. . . . 162
6.6 The manually annotated cortical plate labels ($\mathbf{S}_R$) are shown at each gestational week in the transventricular plane. . . . 162
6.7 INTERGROWTH-21st training data split across gestation. . . . 164
6.8 The set of priors investigated for TEDS-Net cortical plate segmentation. 166
6.9 TEDS-Net segmentation across the different priors. . . . 168
6.10 Example segmentations across the different networks. . . . 170
6.11 Qualitative examples of surface meshes generated from predicted CP volumes. . . . 173
6.12 An overview of the Sylvian Fissure automated and manual labels . . 175
6.13 Automated pipeline for measuring SF depth. . . . 175
6.14 Comparing the SF depths measured for manually obtained SF labels and from our automated method. . . . 177
6.15 Pipeline to measure the depth at each location on the predicted cortical plate. . . . 178
6.16 An overview of the thickness measurement pipeline 179
6.17 Measures of cortical volume and surface area of the cortical plate across gestation. . . . 182
6.18 The Sylvian Fissure depth measured using our automated method. . 184
6.19 The average depth of the inner cortical plate boundary projected onto the atlas CP template for each gestational week. . . . 187
6.20 The average depth of the inner cortical plate boundary projected onto the TEDS-Net prior. . . . 188
6.21 The average depth of the outer cortical plate boundary projected onto the atlas CP template for each gestational week. . . . 189
6.22 The average depth of the outer cortical plate boundary projected onto the TEDS-Net prior. . . . 189
6.23 An overview of the central sulcus at 26 GW 190
6.24 The average thickness values at each location across the CP at each gestational week projected onto the prior shape surface. . . . 191
6.25 The thickness values averaged in the atlas space and shown in the transventricular axial plane. . . . 192

6.26 The thickness values averaged in the atlas space and shown in the coronal plane. 193
6.27 Visualisation of the average Jacobian determinants on the prior template. 194
6.28 Visualisation of the average Jacobian determinants on the prior template from 2D slices. 195

A.1 Parameter tuning VoxelMorph 212
A.2 Parameter tuning TopoLoss 213

B.1 Statistically significant results for ACDC segmentation 216
B.2 Statistically significant results for liver segmentation 217
B.3 Statistically significant results for CP segmentation 218

C.1 The five best performing TEDS-Net segmentation in terms of Dice compared to the other baselines. 220
C.2 The five mid-preforming TEDS-Net segmentation in terms of Dice compared, to the other baselines. 220
C.3 The five worse performing TEDS-Net segmentations in terms of Dice, compared to the other baselines 221
C.4 The average Dice scores across the networks for the 3 groups displayed in Fig. C.1, Fig. C.2 and Fig. C.3. 221

D.1 The initial architecture used for cortical plate segmentation. 223
D.2 Volume and surface area for a larger validation dataset 224

List of Tables

2.1 Evaluating the segmentations across a range of metrics 29

3.1 An overview of the studies that describe the folding timetabling . . 50
3.2 Methods for cortical plate segmentation. 64
3.3 An overview of the studies that have characterised the fetal brain in 3D using different morphological properties. 74

4.1 Ablation studies of TEDS-Net 98
4.2 Quantitative performance of myocardium segmentation 110
4.3 Network training time and number of parameters 111

5.1 Performance for multiple structure cardiac segmentation. 133
5.2 A comparison of liver segmentation performance. 144

6.1 Cortical Plate Segmentation Performance 171
6.2 Hemisphere Asymmetry 185

List of Abbreviations

1D, 2D, 3D	One-, two- or three-dimensions.
ACDC	Automated Cardiac Diagnosis Challenge
ANCOVA	Analysis of Covariance
ANOVA	Analysis of Variance
ASSD	Average Surface Symmetrical Distance
BCE	Binary Cross Entropy
CAD	Coronary Artery Disease
CNN	Convolutional Neural Network
CNS	Central Nervous System
CP	Cortical Plate
CT	Computer Tomography
FDA	Food and Drug Administration
FGLS	Fetal Growth Longitudinal Study
fMRI	Functional Magnetic Resonance Imaging
FSL	FMRIB Software Library
GW	Gestational Week
HD	Hausdorff Distance
HPLS	Hypoplastic Left Heart Syndrome
HU	Hounsfield Unit
INTERGROWTH	International Fetal and Newborn Growth Consortium
ISUOG	The International Society of Ultrasound in Obstetrics and Gynecology
MICCAI	Medical Image Computing and Computer Assisted Interventions
MNIST	Modified National Institute of Standards and Technology database
MRI	Magnetic Resonance Imaging
NHS	National Health Service
ODE	Ordinary Differential Equation
ReLU	Rectified Linear Unit
SOTA	State of the art
SF	Sylvian fissure
STN	Spatial Transformer Network
TEDS-Net	Topology Encouraging Deformation Segmentation Network
US	Ultrasound

TEDS-Net Notation

$\mathbf{X}$	Network input
$\mathbf{Y}$	Ground truth label
$\hat{\mathbf{Y}}$	Predicted segmentation
$\mathbf{P}$	Prior shape
$\mathbf{x}$	Positions at which the deformation was described
$\mathbf{u}$	Displacement matrix
$\mathbf{\Phi}$	Deformation field
$\mathbf{\Phi}_{\text{bulk}}$	Bulk deformation field
$\mathbf{\Phi}_{\text{ft}}$	Fine-tuning deformation field
g_θ	Encoder-decoder network, described by parameters θ
l	Number of layers within the network
f	Number of feature maps in the first layer
l_{bulk}	The decoder layer were the bulk field is generated from
l_{ft}	The decoder layer were the fine-tuning field is generated from
h	Number of composition layers
σ	The sigma used in the Gaussian kernels for smoothing
T	The threshold applied to the predicted segmentation
α, β	The weighting between field regularisation and Dice loss

1 Introduction

1.1 Motivation

Over the last decade, medical imaging has become more commonplace in routine care, with the number of computer tomography (CT) and magnetic resonance imaging (MRI) scans increasing on average by 7% each year [1]. In 2019, 45.4 million imaging tests were performed by the National Health Service (NHS) in England alone [2]. In medical imaging pipelines, an image is acquired and then analysed by a radiographer or specialist to obtain an outcome, i.e. a diagnosis. One fundamental step in many of these analysis pipelines is structural segmentation, which is used frequently in diagnostics, treatment planning and monitoring disease progression [3–5].

Segmentation is the process of reducing an image into meaningful categories by grouping pixels that belong to the same structure together. Once the label of a structure of interest is obtained, clinical measurements, such as volume or perimeter, can be automatically performed [6, 7]. For example, in radiotherapy treatment planning, the target organ is usually segmented to facilitate measures of volume or distances to other key organs, in order to calculate dose [8, 9]. Additionally, the muscle surrounding the heart's ventricles, the *myocardium*, is often delineated to measure its thickness or perimeter, which can be used to detect diseases such as coronary artery disease (CAD) [6, 10, 11]. An example pipeline of this scenario is

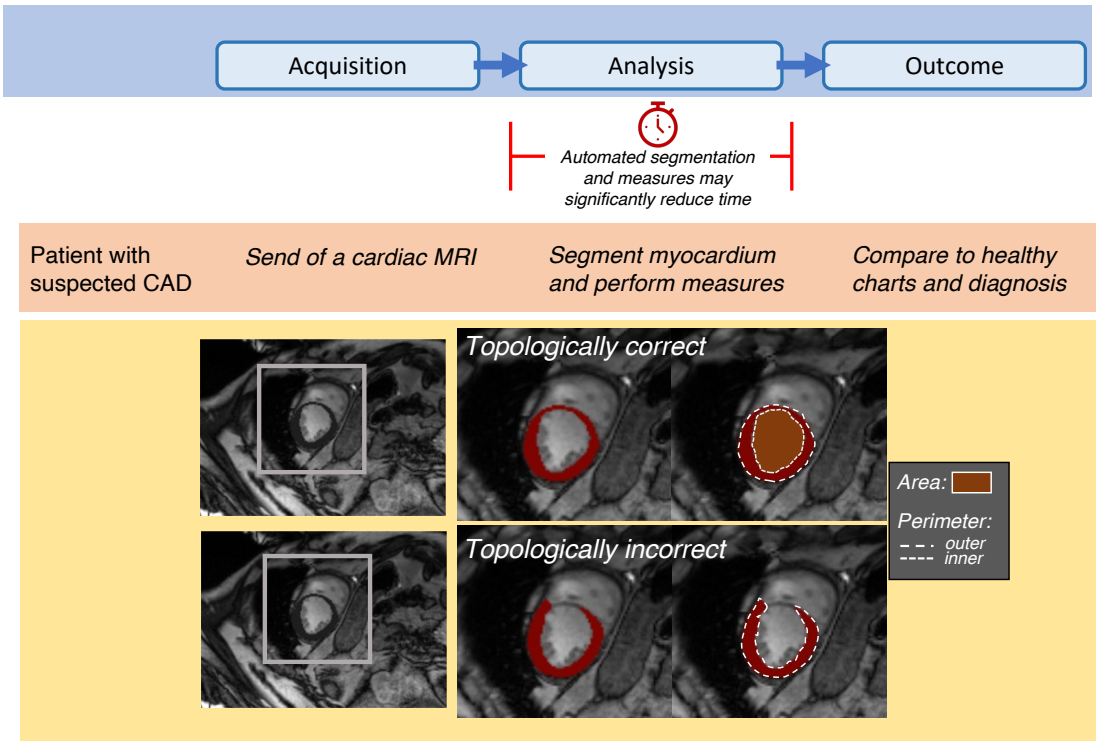


Figure 1.1: The role of segmentations in medical imaging pipelines. In blue shows a *simplified* overview of the standard medical imaging pipeline, with a hypothetical example shown in orange, for the detection of coronary artery disease (CAD) [6]. The bottleneck in such pipelines is the slow process of performing manual segmentations or measurements. Automated deep learning segmentation offers an alternative, however, it requires highly accurate results. In yellow shows, the myocardium is segmented using two automated methods. The top example is topologically correct, allowing the inner and outer boundary to be extracted separately, which may be required for some measurements, such as the inner or outer perimeter or left ventricle area (shown in brown). Whereas, the bottom example is topologically incorrect, making automated measures of boundaries particularly difficult.

shown in Figure 1.1. Therefore, it is crucial that segmentations are highly accurate to facilitate measures further down the imaging pipeline.

In current clinical pipelines, structural segmentation is often performed manually, despite it being time consuming and prone to intra- and interobserver variations [8]. However, due to workforce shortages, there is simply not enough time to perform the increasing load of analysis [1, 12], with current projections estimating a 39% workforce shortfall in consultant radiologist by 2026 [12]. Therefore, as the number of imaging tasks is increasing, combined with the workforce shortfall, there is a need to deliver automated analysis methods to try and relieve some

of this burden from radiologists.

The need for automated medical segmentations has given rise to a large number of machine-learning algorithms. Initial segmentation algorithms applied to medical imaging were often inherited from the computer vision community, such as AlexNet [13] and VGG-Net [14], which have been shown to achieve promising performance in medical imaging challenges [15, 16]. However, medical images have unique issues, such as large shadows or artefacts across the volumes, which, when paired with the traditional optimisation process (discussed further in Chapter 2), can give rise to *topological* errors within the segmentations. The topology of a structure refers to its overall shape, and can be used to assess whether a segmented structure is anatomically and pathologically plausible, and therefore suitable for the subsequent analysis steps. For instance, in the case of myocardium analysis, if the segmentation was topologically incorrect, e.g. no longer a closed ring, there would be no distinction between the inner and outer boundary, which would lead to entirely misleading measures of the perimeter, as shown in Figure 1.1. Therefore, there is a need for segmentation methods that ensure topologically accurate results [11].

Incorrect topology in automated medical segmentation methods is not a rare occurrence but has gone largely unnoticed due to typical evaluation metrics used to assess a network's performance. In fact, the most popular metrics often only assess at the pixel-wise level, ignoring the higher-level structure of a shape. Therefore, to accurately evaluate segmentation performance, a variety of metrics are required.

One common example of problematic topological errors is found when segmenting the outer layer of the brain, known as grey matter in adults or the *cortical plate* in fetuses. In clinical research, the aim of segmenting this structure is to analyse certain 3D properties (such as thickness, curvature, surface area and volume) as a way to detect abnormalities. However, as the structure is thin and heavily convoluted, it is prone to topological errors, such as holes across the surface, making further surface analysis particularly challenging. When studying fetal development, ultrasound is the modality of choice and is routinely used for fetal examinations. However, in ultrasound volumes, there is poor contrast between the cortical plate

and underlying tissues. Moreover, they also contain large shadows, which obstruct key regions of anatomy (further discussed in Chapter 3), which can result in the creation of topological errors within the cortical plate segmentation. Because of these challenges, there is a lack of research quantifying the developing fetal brain in 3D from ultrasound.

In this thesis, I propose a novel deep-learning segmentation method that utilises prior knowledge of a structure's topology to generate topologically-accurate segmentations. I demonstrate the method across a range of natural and medical segmentation tasks where the structures have known (or expected) topologies.

1.2 How This Thesis Came To Be

A brief note on how this thesis took its structure.

At the start of my DPhil, my aim was to study the dynamic changes of the developing fetal brain, and more specifically how its surface develops over gestation. As ultrasound is the modality of choice in prenatal care, I wanted our research to use this modality to facilitate clinical translation in future work. To study the 3D properties of the developing surface, the cortical plate must first be extracted, which was the first step of my DPhil. However, after trying a plethora of machine learning segmentation methods, I was haunted by topological holes across the extracted surfaces, which made further analysis of its properties extremely challenging. This lead me to research topological-aware segmentation methods, where I found I was not alone in my haunting and that such errors were present in a wide variety of medical segmentation tasks [17–19]. The layout of this thesis is therefore not in the same order as my DPhil journey. I first propose and thoroughly examine our novel topology-preserving segmentation framework, before applying it to a range of medical imaging tasks, and finally, use it to segment the cortical plate from ultrasound in order to measure the morphological changes across gestation.

1.3 Contributions

Machine learning algorithms have revolutionised the field of medical segmentation, with an increasing number of papers being published on the topic each year [20, 21]. However, the reported performance of such methods is beginning to plateau, with new methods reporting marginally better results than the previous state of that art (SOTA) method [21]. Further, as the goal of new segmentation methods is often to outperform SOTA methods, little work has been performed to ensure that segmentations are fundamentally useful to the medical community and the subsequent analysis tasks [5, 21]. One source of such problems has been topological errors. Across this body of work, I try to address clinically relevant medical questions using deep learning, whilst ensuring the final outputs from the methods are fundamentally fit for the question at hand.

In Chapter 4, I propose a machine learning segmentation network, Topology Enforcing Segmentation Network (TEDS-Net), that incorporates prior knowledge of the structure of interest's topology, in order to encourage the segmentation to have the correct shape features. This method is first applied to the task of cardiac segmentation, a task where previous SOTA methods often result in topological errors [17, 21], and achieves a 100% topological accuracy, whilst remaining competitive to SOTA methods on other metrics (such as Dice and binary cross entropy).

The backbone of TEDS-Net is built upon the continuous guarantees of topology-preserving fields. However, in Chapter 4, I show how these guarantees begin to break down when working in the discrete domain, and propose additional modifications to more strictly enforce topology-preservation, a step often overlooked in previous methods.

Supervised machine learning segmentation models are extremely data-hungry, requiring a large number of labelled scans. Acquiring such a large amount of labelled data is a major limitation of current deep-learning segmentation methods. In Chapter 5, I address this problem by proposing a set of minimal labelling protocols, where TEDS-Net can be used to perform accurate volume segmentation

by only using supervision from a few labelled slices, drastically reducing the time required labelling new volumes.

Across this thesis, I apply TEDS-Net to a range of datasets and tasks. The choice of these datasets was deliberate, to show TEDS-Net's ability to segment structures from a range of imaging modalities (MRI, CT, ultrasound and natural images), across different dimensions (2D and 3D), from a variety of known topologies profiles, and across different segmentation scenarios (multi-structure, multiple topologies within a dataset and with minimal labels).

Having successfully established an automated topology-preserving segmentation method, in Chapter 6, I apply it to the challenging task of cortical plate segmentation from in-utero ultrasound volumes. In this chapter, I show how TEDS-Net can anatomically guide the segmentation in regions where large acoustic shadows obstruct key regions of the cortical plate. From the topologically-correct segmentations, I propose methods to analyse the morphological properties of individual fetuses, before characterising the expected development across the healthy population through growth charts and cortical morphological atlases. These results are then compared to previous MRI and 2D ultrasound studies and found to have a high overlap.

1.4 Thesis Layout

This thesis is structured as follows:

In Chapter 2, I review the past and present methods used to perform medical image segmentation, focusing on deep learning methods, which will be used throughout this thesis. I overview how such methods are traditionally assessed and show that a range of metrics is required to accurately evaluate the performance of a method. Further, I introduce how current models and optimisation strategies can lead to topological errors and review how current deep-learning methods have tried to overcome such errors. Throughout this discussion, I describe how deformation fields can be used to segment medical

structures. Further, I show how such fields can enforce topology preservation in the continuous domain, and that these guarantees begin to break down when working in the discrete domain.

In Chapter 3, I describe the clinical motivation of studying the developing fetal cortical plate and review how it develops across gestation. Further, I discuss how fetal brain development is currently assessed in routine clinical practice, as well as in the research domain.

Chapter 4 is the first of the contribution chapters, where I propose our topology-preserving deep learning method: TEDS-Net. In this chapter, I further discuss how to more strictly enforce topology-preserving fields whilst working in the discrete domain, and perform thorough hyper-parameter exploration and ablation studies to show the limitations of such guarantees. In this chapter, I show TEDS-Net's performance on cardiac segmentation and illustrate how topological errors can lead to misleading perimeter measures. The work presented here is central to experiments performed in Chapter 5 and Chapter 6.

In Chapter 5, TEDS-Net is extended for a series of different segmentation tasks, which all offer unique challenges. First, I perform digit segmentation, to show TEDS-Net's ability to segment multiple different topologies within the same training cycle and trained model. Next, I apply TEDS-Net to multi-class segmentation to show how group topology can be preserved as well as individual structure. Finally, I extend the algorithm for 3D liver segmentation. In this task, I also show how TEDS-Net can learn whole volume segmentation through weak supervision, i.e. only one labelled slice per volume, reducing the time required to generate ground truth labels.

In the final chapter, Chapter 6, I apply TEDS-Net to the task of cortical plate segmentation from ultrasound scans. Further, I design an automated pipeline to characterise the properties of the developing cortical plate, which is applied

to the predicted volumes. From the individual properties, I develop a series of expected growth charts measuring both local and global changes across gestation.

Due to the interdisciplinary nature of this thesis, the literature review and theoretical explanations are distributed across the relevant chapters. Chapter 2 offers an introduction to segmentations, topology and the choice of evaluation metrics. Whereas, Chapter 4 begins with an extended explanation of topology and topology-preserving fields when applied in the discrete space. Further, in Chapter 4 I provide a condensed summary of topology in deep learning segmentation methods. Chapter 3 introduces the cortical plate and explains why characterising the developing cortical plate is important, before reviewing previous literature that has characterised this development. I begin Chapter 3 by summarising this literature and motivating why a new analysis pipeline to study cortical plate development from ultrasound is required

1.5 Originality

I declare that I am the sole author of this thesis document, and I produced all of the tables and figures included herein unless explicitly stated otherwise. Critiques and comments were provided by my supervisors Prof. Ana Namburete and Prof. Mark Jenkinson, and members of my research group Oxford Machine Learning NeuroImaging lab (OMNI). Unless otherwise specified, the pre-processing, code development and analysis were my own individual work. The ultrasound images used throughout were acquired by clinical collaborators as part of the INTERGROWTH-21st consortium.

Here I overview the originality of each chapter:

Chapter 4 utilises data from the Automated Cardiac Diagnosis Challenge (ACDC) [22], an open access dataset, generated for MICCAI 2017, which is publically available online[1]. In this chapter, I employ four SOTA baselines,

[1]ACDC Dataset: https://www.creatis.insa-lyon.fr/Challenge/acdc/databases.html

namely; U-Net [23], Swin-Unet [24], VoxelMorph [25], and the topological loss function: Topoloss [26]. The Swin-Unet[2], VoxelMorph[3] and TopoLoss[4] methods were downloaded from the original author's public repositories, with the codes modified for our task. Whereas, I wrote the code for the U-Net architecture. The implementation of these models was kept to be as consistent as possible with that reported in the original papers, however, key parameters were tuned on our new dataset, as reported in the text. I designed and implemented TEDS-Net architecture, with the Scaling and Squaring layers adapted from VoxelMorph's implementation.

In **Chapter 5** two more open-source, publicly available datasets were used. The first: MNIST [27], a large multi-digit dataset, which was downloaded directly through PyTorch[5]. The second dataset was for liver segmentation, which was part of the large Decathlon dataset[6]. I performed all the pre-processing for the datasets as described in the text. The Sli2Vol baseline was implemented by my colleague Pak Hei (Hugo) Yeung.

For **Chapter 6**, the INTERGROWTH-21st dataset was used, which was obtained through collaboration between Prof. Ana Namburete and Prof. Aris T. Papageorghiou and the INTERGROWTH consortium. Prof. Namburete generated the fetal brain atlas, which I used in atlas-based label segmentation. I manually segmented the cortical plate from the atlases and adapted Prof Namburete's code to propagate the labels back to the individuals. Further, Prof. Namburete selected the optimal subset of INTERGROWTH-21st data used throughout this group, which she had aligned and normalised in previous work.

[2]Swin U-Net: https://github.com/HuCaoFighting/Swin-Unet, accessed March 2022
[3]VoxelMorph: https://github.com/voxelmorph/voxelmorph, accessed Jan 2021
[4]TopoLoss: https://github.com/HuXiaoling/TopoLoss, accessed March 2022
[5]MNIST: https://pytorch.org/vision/stable/datasets.html
[6]Liver Decathlon: http://medicaldecathlon.com/

1.6 List of Publications

Some of the work presented in this thesis has been published during my DPhil. Below I list my first author publications that have been referenced and adapted for this thesis. Unless otherwise explicitly stated, the work performed in the papers listed below and presented in this thesis was my own individual work.

Peer-Reviewed Conference Proceedings

M. K. Wyburd, N. K. Dinsdale, A. I. L. Namburete and M. Jenkinson *TEDS-Net: Enforcing Diffeomorphisms in Spatial Transformers to Guarantee Topology Preservation in Segmentations.* International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2021.

M. K. Wyburd[7], L. Hesse, M. Aliasi, M. Jenkinson, A. T. Papageorghiou, M. C. Haak and A. I. L. Namburete *Assessment of Regional Cortical Development Through Fissure Based Gestational Age Estimation in 3D Fetal Ultrasound.* Perinatal Imaging, Placental and Preterm Image Analysis. Springer, Cham, 2021. 242-252.

M. K. Wyburd, M. Jenkinson and A. I. L. Namburete *Cortical plate segmentation using CNNs in 3D fetal ultrasound.* Annual Conference on Medical Image Understanding and Analysis. Springer, Cham, 2020

Journal Publications (Under Review)

M. K. Wyburd, N. K. Dinsdale, A. I. L. Namburete and M. Jenkinson *Anatomically Plausible Segmentations: Explicitly Preserving Topology Through Prior Deformations. In Revision*: submitted to Medical Image Analysis (MedIA) (2022)

[7]This work was joint first authorship with L. Hesse. The work presented in this paper is not presented in this thesis, however, some of the literature review is relevant to Chapter 3

Conference Presentations

M. K. Wyburd, M. Jenkinson and A. I. L. Namburete., *Anatomically Plausible Segmentations: Explicitly Preserving Topology Through Prior Deformations.* in The Annual Oxford University Imaging Festival, *September 2022, Oxford*

M. K. Wyburd, M. Jenkinson and A. I. L. Namburete., *Artificial Intelligence can Measure Fetal Brain Development from 3D Ultrasound* in STEM for Britain, *March 2022, London* - Bronze Award for Engineering

M. K. Wyburd, M. Jenkinson and A. I. L. Namburete., *Using Deep Learning to Segment the Developing Cortical Plate from 3D Fetal Ultrasound* in International Society of Ultrasound in Obstetrics and Gynecology (ISUOG), *September 2021, virtual* - Oral Presentation

M. K. Wyburd, L. Hesse, M. Aliasi, M. Jenkinson, A. T. Papageorghiou, M. C. Haak and A. I. L., *Deep learning-based assessment of second-trimester cortical plate development in three-dimensional ultrasound* in International Society of Ultrasound in Obstetrics and Gynecology (ISUOG), *September 2021, virtual* - Oral Presentation

M. K. Wyburd, M. Jenkinson and A. I. L. Namburete., *Using Deep Learning to Segment the Developing Cortical Plate from 3D Fetal Ultrasound* in Radiological Society of North America (RSNA), *December 2020, virtual*

2

Review of Medical Segmentations and Topology

This chapter provides a summary of the current landscape of the task of medical segmentation. It starts by reviewing popular segmentation methods, with a spotlight on deep learning methods and more specifically convolutional neural networks (CNNs), which are predominantly used in this thesis (Section 2.1). In Section 2.2, I introduce how segmentations are quantitatively assessed, and show why a range of metrics are required to accurately describe a structure. Further, I introduce how CNNs and typical optimisation strategies can lead to topological errors within segmentations, before finally reviewing recent work that has worked to overcome such errors (Section 2.3).

2.1 Segmentation Methods

In recent years, the task of automated segmentation has been transformed by deep learning methods and more specially CNNs, which now form the basis of most newly developed algorithms [20]. In this section, I provide an overview of some past segmentation methods, highlighting atlas-based and deformable models, before focusing on deep learning algorithms, which will be predominately used throughout this work.

2.1.1 Semantic Segmentation

Throughout this work, when discussing segmentations, I am referring to *semantic* segmentation, rather than *instance* or *panoptic*. Semantic segmentation assigns every pixel of an image with a class label (e.g. tumour or not tumour) and labels every instance of that class with the same label. Whereas, instance segmentation treats each instance of the same class as a distinct instance. Panoptic segmentation combines both segmentation and instance and assigns each pixel with two labels, a semantic and an instance ID.

In digital analysis, an image $\mathbf{X} \in \mathbb{R}^{H \times W \times C}$ is stored as a matrix, where $H \times W$ are the image dimensions and C the number of channels. Greyscales images have a single channel $C = 1$, whereas colour images have $C = 3$: a red, green, and blue channel. Each pixel $\mathbf{p}_{i,j} = \mathbf{X}[i, j]$ within an image contains a value between either 0 and 1 or 0 and 255, describing the intensity or shade of $\mathbf{p}_{i,j}$. In semantic segmentation, every $\mathbf{p}_{i,j}$ within an image is assigned a class out of a set of c labels, resulting in a segmentation map $\hat{\mathbf{Y}} \in \mathbb{R}^{H \times W \times c}$.

2.1.2 Traditional Segmentation Methods

Traditional segmentation methods can be broadly characterised into two groups: identifying discontinuities and identifying similarities [28]. Methods that identify discontinuities do so by detecting distinct changes in intensities, such as edge detection methods. On the other hand, methods that identify similarities, are methods that perform segmentation by grouping together pixels with similar

properties, such as thresholding and region growing [29, 30]. These initial methods provided a good starting point for segmentations but were quickly outperformed by *atlas-based techniques* and *deformable models*.

Atlas-Based Atlas-based segmentation methods are currently one of the most popular medical imaging segmentation methods and are often integrated into deep learning networks. Atlas-based segmentation is performed by *registering* a group of images to a labelled atlas image[1], and then propagating (e.g. reversing the registering) the labels back to the individual images [31–33]. Registration is the process of geometrically aligning one image to another or to a common coordinate space. In neuroimaging, registration is commonly used to align a set of brain volumes to a common space (e.g. a template image), to facilitate comparisons across time or subjects [34]. Deformable and nonlinear medical image registration has been widely studied [35–38], and usually involves learning an alignment deformation field, which can be done in a variety of ways, from feature-based deformation [39] to deep learning techniques [40]. Methods for generating different deformation fields will be discussed in Section 2.3.2.

Central to this method is the availability of a template to which each image can be registered. In early work, the best or most representative image from a dataset was chosen to be the template, and the transformation between the individuals and the chosen template was calculated [31]. However, it is unlikely that a single subject will have a sufficient variation to represent the population. Therefore, in more recent approaches, templates have been constructed from the entire dataset or other reference populations, to ensure the whole population is well-represented [41, 42]. Templates can be created in a variety of ways, but usually involve registering the population to a common space whilst minimising a certain energy function, so that each individual is equally represented in the final template image, followed by a voxel-wise averaging.

[1] A note on the wording. A *template* is a representation of the anatomy of interest (e.g. an average brain), whereas an *atlas* is a labelled version of the template (e.g. a brain atlas with the grey matter annotated)

The popularity of atlas-based segmentation is likely due to the minimal manual labelling required, the semi-automated approach and the abundance of available labelled atlases [41, 43]. However, it is not without limitations. Firstly, the initial labelling must be performed to a high standard, otherwise, all the propagated labels will be comprised. Secondly, when labelling a new set of subjects, the atlas may not be representative, leading to poor registration and the need for a new atlas, requiring more manual labelling. Lastly, not all tasks can be represented with an atlas, such as tumour or white matter hyperintensities, as the variation across the population is far too vast.

Deformable Models Deformable models are another popular segmentation technique, which works by iteratively deforming curves to minimise some energy function in order to delineate a structure [44]. Initially, these models used snakes (energy-minimising spline-guided curves that are pulled towards lines and edges) to iteratively outline a structure [44]. In this, the curves are made up of a series of points that are controlled by different energy terms to keep the points evenly spaced and smooth, whilst encouraging them to be on top of high-gradient pixels within an image (e.g. an edge). These models are initialised either by a user picking the starting position or through automatic algorithms. Each point along the curve is then subsequently updated using gradient descent until there is no further change in point position [45, 46].

The initial snake algorithms were found to be slow to converge, especially when using an estimated starting position, and are only able to detect gradients when the points along the snake are directly over them. To overcome these limitations, the Gradient Vector Flow method was implemented [47]. In this, the image gradients are extended over a large distance, by diffusing the gradients outward to create a field of forces, better guiding the points along the curve and decreasing the convergence time. Despite performance improvements, snakes still struggled with complex shapes. To overcome this, geometric deformable models implemented using

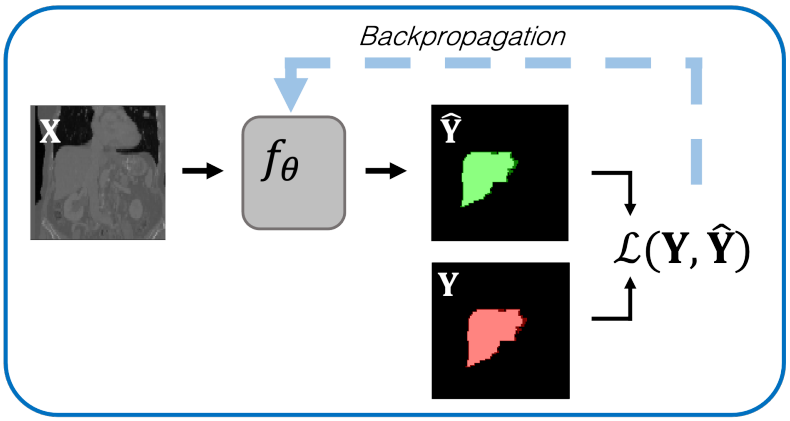


Figure 2.1: An schematic of supervised deep learning segmentation networks. In this, a network f_θ described by parameters θ is used to extract relevant features from an input image $\mathbf{X}$ in order to predict a probability map $\hat{\mathbf{Y}}$. The prediction is then compared to the ground truth $\mathbf{Y}$ using a loss function $\mathcal{L}$, in order to update θ through backpropagation.

the level set method were proposed [48, 49]. Although these methods have improved accuracy, they are still heavily dependent on initialisation and image quality [50].

2.1.3 Deep learning Segmentation

Machine learning algorithms have changed the landscape of automated segmentation, offering almost instantaneous segmentations for new images[2]. There is a wide variety of machine learning segmentation methods, from random forests to transformers, however, they generally follow the same format: from an input image $\mathbf{X}$ relevant features are learnt in order to perform a voxelwise segmentation $\hat{\mathbf{Y}}$. In this thesis, I focus on the use of supervised, deep learning algorithms to perform segmentation.

In supervised learning, it is assumed that for every input image $\mathbf{X}$ there is also a gold standard label, referred to as the ground truth, describing the structures of interest: $\mathbf{Y} \in \mathbb{R}^{H \times W \times c}$, which is used to compare the predicted segmentation to. In this, training data $D = \{X = \{\mathbf{X}_1, \mathbf{X}_2 \cdots \mathbf{X}_N\}, Y = \{\mathbf{Y}_1, \mathbf{Y}_2 \cdots \mathbf{Y}_N\}\}$, consisting of pairs of images and labels, is fed into a network ($f_\theta(\mathbf{X})$), which aims to learn a mapping from $\mathbf{X}$ to $\mathbf{Y}$.

Deep learning refers to networks (f_θ) that are made up of a series of fully-connected or convolutional layers, with learnable weights (θ) which extract infor-

[2]In this section, I am exclusively talking about 2D images for simplicity, however, these methods extend trivially to 3D volumes.

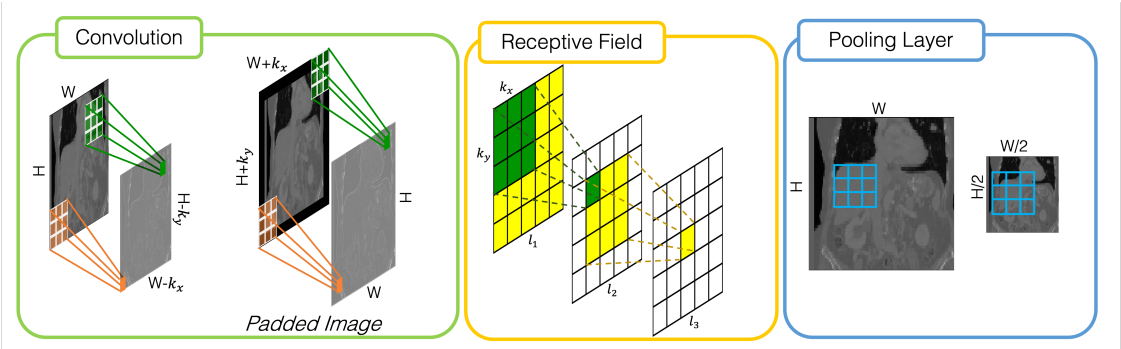


Figure 2.2: The building blocks of CNNs. Schematics of convolutions applied to an image. The kernels (shown in orange and green) are applied across the whole image to build a feature map. A schematic of the receptive field is shown in the middle, the feature at each location on l_3 is based upon every pixel value from the input image l_1. The importance of the pooling layer is illustrated on the RHS, where after pooling the kernel can learn higher-level features.

mation from the input data, in order to produce a probability map ($\hat{\mathbf{Y}}$). These weights are optimised through backpropagation [51] by comparing the predicted segmentation ($\hat{\mathbf{Y}}_i$) to the known label ($\mathbf{Y}_i$), using a chosen loss metric ($\mathcal{L}(\hat{\mathbf{Y}}, \mathbf{Y})$). This process is summarised in Figure 2.1. The two techniques used to optimise f_θ vary across methods. In this section, I summarise three leading deep learning methods, namely; CNNs and transformers.

2.1.3.1 Convolutional Neural Network

CNNs are arguably the most popular family of networks for medical image analysis [20]. They extract features from an input image $\mathbf{X}$ by performing a series of convolutions [13, 14, 51]. CNNs are usually built using the same blocks, namely convolutional layers, activation functions, and pooling.

Convolutional Layers In a convolution, denoted $\mathbf{w} * \mathbf{X}$, a weight matrix, usually termed *kernel* or *filter*, made up of learnable weights $\mathbf{w} \in \mathbb{R}^{k_x \times k_y}$ is applied to a whole image $\mathbf{X} \in \mathbb{R}^{H \times W}$ to obtain a filtered image. The kernel, typically of size $[3 \times 3]$ in 2D implementations, is applied to each image patch centred on i, j,

$$\mathbf{X}_{\text{patch}} = \mathbf{X}[i - k_x : i + k_x, j - k_y : j + k_y],$$

for $i \in k_x/2 : N_x - k_x/2$, $j \in k_y/2 : N_y - k_y/2$, within an image, producing a single value, which is assigned to pixel location i, j in the filtered image, as shown in Figure 2.2.

The filtered image is found to have a smaller dimension than the input, due to the size of the kernel. To keep the dimensions the same after a convolution, zero padding is often applied around the input, so that the centre of the kernel is placed over the entirety of the input, as shown in Figure 2.2.

Typically in CNNs, stacks of convolutional filters are applied at different network depths (as discussed later in this section). The outputs of these convolutions are referred to as feature maps, as they extract features corresponding to the filter weights $\mathbf{w}$. Across training, the network learns which features are required to produce an accurate segmentation, by learning $\mathbf{w}$.

Convolutional kernels were used frequently in digital image processing before machine learning, with certain weights obtaining certain features from an image. For example, a kernel with weights:

$$\mathbf{w} = \begin{bmatrix} -1 & 2 & -1 \\ -1 & 2 & -1 \\ -1 & 2 & -1 \end{bmatrix}$$

obtains the vertical lines within an image [52]. This inspired the creation of CNNs, to try and intuitively analyse the volumes by extracting the features that the network deemed important for the current task, by learning the $\mathbf{w}$.

Activation Function Between convolutions, activation functions, always known as *nonlinearities*, are often applied. Without such functions, the network would only learn a linear function of its input, which is too restrictive, as the relationship between $\mathbf{X}$ and $\mathbf{Y}$ is unlikely to be linear [53]. Thus, including activations between convolutions, complex functions can be learnt. This nonlinearity has led to much of the success of CNNs. Popular activation functions include: sigmoid

$$a(\mathbf{X}) = \frac{1}{1 + e^{-\mathbf{X}}},$$

tanh

$$a(\mathbf{X}) = \frac{e^{\mathbf{X}} - e^{-\mathbf{X}}}{e^{\mathbf{X}} + e^{-\mathbf{X}}},$$

and ReLU

$$a(\mathbf{X}) = \max(0, \mathbf{X}).$$

The most appropriate choice depends on the task and the location in the network. For instance, as a sigmoid function keeps the output between 0 and 1, it is often used at the end of a neural network to normalise the probabilities. However, they are rarely used between convolutional layers, as they often lead to *vanishing gradients*. This effect occurs when the gradients within the network become too small, preventing backpropagation from effectively updating the weights. If the input to a sigmoid activation function becomes too large (or small), the derivative becomes close to zero. A similar effect happens when using tanh, which squishes the output between -1 and 1. As the ReLU activation function has an infinite range, it reduces this effect, which is found to result in faster convergence [53].

Pooling Layer A convolutional layer extract features from a local neighbourhood, defined by the dimensions of the kernel. As the size of the kernel is usually much smaller than the input dimensions, it does not learn large-scale features or dependencies across the image. To overcome this, pooling layers are introduced, which reduce the dimensions of the feature maps, meaning that the kernel (whose dimensions do not change) can now extract larger-scale information. This effect is shown in Figure 2.2. Therefore, pooling layers increase the size of the final *receptive field*. The receptive field describes the area of the input that a network has "seen" and uses to base its final prediction on, as shown in Figure 2.2. After a series of convolutional layers l, the receptive field increases and, by downsampling the feature maps, the area is further increased. Pooling layers also help to reduce the computation demands in CNNs, as fewer convolutions are required. Pooling is often performed using Max Pooling, where the highest pixel value is extracted from a pooling window of size $p_x \times p_y$, which is usually 2×2 in 2D tasks.

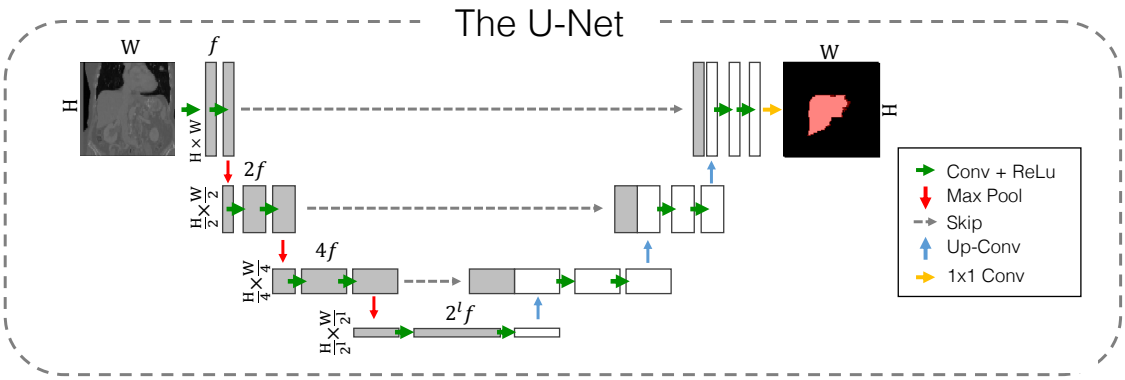


Figure 2.3: An overview of the popular U-Net architecture, as presented in [23].

The U-Net Most CNN architectures are designed around these three main building blocks: convolutions, activation functions, and pooling layers [14, 54]. In this section, I introduce the U-Net [23], a seminal CNN architecture, which has transformed the field of medical segmentation. In what follows I outline the main features of the original U-Net and U-Net style networks.

The U-Net is a fully convolutional network made up of two branches, the *contracting* and *expanding* branches, with *skip connections* connecting the two. An overview of the architecture is shown in Figure 2.3. Both branches are made up of a series of convolutional blocks. In this, f convolutions are applied to an input (resulting in f features maps), followed by a ReLU activation function, before f new convolutions are applied to the new feature maps, followed by a second ReLU activation. The contracting or *encoder* branch is built upon l layers of convolutional blocks, with Max Pooling applied between each layer, to reduce the dimensions of the features maps by half: hence the name *contracting*. Between each layer the number of feature maps is doubled, by doubling the number of convolutional kernels used, maintaining representation. The expanding (or *decoder*) branch joins the end of the encoder, where the feature maps are upsampled to twice their resolution, usually using linear interpolation. The last set of feature maps from the corresponding layer in the expansion branch is then concatenated to the upsampled features, using skip connections, which are then followed by a convolutional block. Skip connections help the network reconstruct the segmentation by using the fine-grained details

learnt in the encoder with the more abstract features learnt in the decoder. The upsampling, skip-connection and convolutional block, are then repeated for l layers until the final feature maps have the same resolution as the input image. Finally, a 1×1 convolution is applied to the output of the final convolutional block, to reduce the feature maps to the required number of output channels (c), where each structure is assigned its own channel. Finally, a soft-max activation (or sigmoid, depending on c) is applied, to enforce that the probability of each pixel (i, j) over all classes is equal to 1: $\sum_{ch}^{c} \hat{\mathbf{Y}}[i, j, ch] = 1$.

The core U-Net has been found to excel across a wide range of medical segmentation problems, with little or no adaption of the model required, and often found to outperform the previous state of the art (SOTA) methods [55–58]. Since the publication of the U-Net, additional components have been found to further enhance its performance [59–64]. For example, spatial dropout is now very commonly found in U-Net implementations [59, 60]. Dropout works by probabilistically removing parts of the input image or feature maps located deeper within the CNN, to stop the network from overfitting to the training set. The performance was further advanced by using batch normalisation between each convolution [65], which is now regularly implemented. Larger modifications of the U-Net have also been produced, including 3D U-Nets [66, 67] and the attention U-Net [68].

The U-Net as it was proposed in the original 2015 paper, with the additional modifications described above, is still found to produce gold-standard segmentation performance, with many new methods simply adapting the architecture to improve a certain task. However, even for out-of-the-box implementation, the U-Net style architectures are still SOTA.

2.1.3.2 Transformers

CNNs revolutionised segmentation methods, however, in the last few years transformers have been proposed, which have outperformed CNNs for natural image segmentation and more recently some medical imaging problems [24, 69]. Transformers [70] were first proposed for natural language processing but have now been adapted

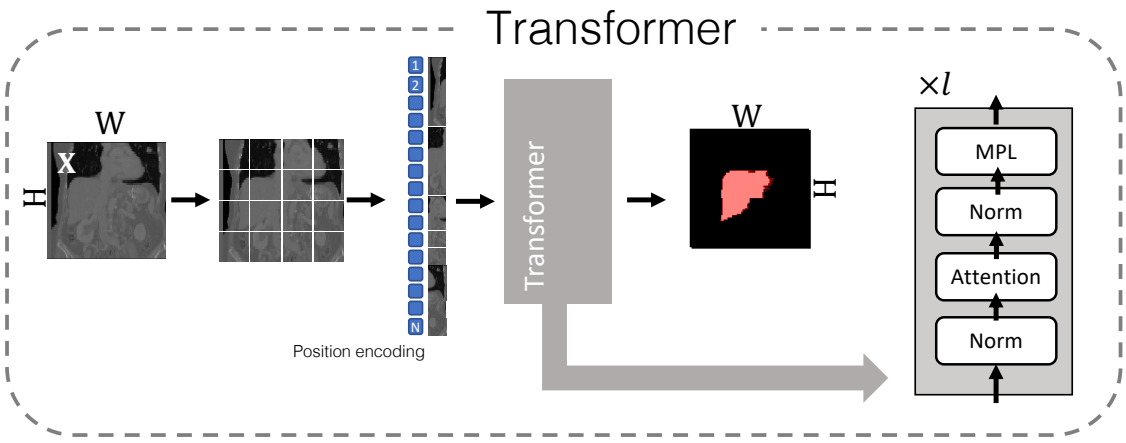


Figure 2.4: An schematic showing the overview of transformer networks used for segmentation.

for segmentation tasks [69]. Unlike CNNs, transformers rely on self-attention mechanisms to capture long-range dependencies between patches of an image. The networks take in position embedded image patches, which are then flatted and then input to a transformer [70], as shown in Figure 2.4. Transformer blocks are made up of normalisation, followed by an attention module and finally linear layers.

2.1.3.3 CNNs vs Transformers

Global Features One limitation of solely convolutional-based CNNs, such as the U-Net, is that they predominantly learn local features and not the higher-level structure within an image. As the convolutional layers base each pixel's prediction on the features seen within the receptive field, the convolutions never see the whole image (unless the receptive field is the same size as the input image, which is uncommon). Therefore, such locally-acting networks, risk not learning the topology of the desired structures, which can lead to segmentation being anatomically implausible. Transformer networks learn long-range dependencies better than CNN, as each pixel can be connected through the attention modules.

Inductive Bias Inductive biases refer to the set of assumptions on which algorithms are based. CNNs are built upon two main biases. Firstly, pixels that are near each other are related, hence why local convolutions can extract relevant

features. Secondly, that all regions within an image should be processed identically, i.e. the same convolutional filters are applied to each region of an image. Conversely, transformers have no strong biases, and the network is tasked with learning the important pixel relationships. There are pros and cons to both approaches. Due partially to these biases, CNNs can perform to a high standard whilst training on minimal data, whereas, transformers are data-hungry. However, biases can also constrain the CNNs models. As discussed previously, CNNs only learn local features, whereas, transformers can learn specific long-range dependencies if they are deemed relevant to the task. In recent work, models have been developed to take the best features of CNNs and Transformer, such as the ConVit, which is initialised with convolutional biases that the network can choose to ignore when necessary [71].

Implementation As discussed above, transformers usually require more training data than CNNs. Moreover, typically they are also usually made up of more parameters, such as the Switch transformer, which is made up of 1.6 trillion parameters [72]. However, due to the patch-based approach, transformers are less computationally expensive, and therefore, require less training time per parameter. For example, a U-Net with 5.2 million parameters takes 0.22 minutes per epoch, compared to the standard Swin-Unet (a fully transformer-based network) which has 413 million parameters and trains in 1.53 minutes per epoch[3].

2.2 Segmentation Assessment

In order to train and evaluate a segmentation network, loss functions $\mathcal{L}$ and evaluation metrics[4] are required. Broadly, these metrics compare features between a ground truth label ($\mathbf{Y}$) and a prediction ($\hat{\mathbf{Y}}$). To fully quantify a method's performance, a range of metrics are required [73, 74]. Here I introduce and justify the selection of metrics that are used throughout this thesis.

[3] as trained on an Nvidia GeForce GTX 1080 GPU, with a memory of 10GB

[4] A note on terminology: although *metric* and *measures* are widely used interchangeably in the imaging community, by definition a *metric* is asymmetric to its operand which is not true for all the *measures* described below. However, following convention, throughout this thesis I equate the two.

2.2.0.1 Binary Cross Entropy (BCE)

Cross entropy is a measure of the difference between two probability distributions, and is commonly used to train deep-learning models. Binary cross entropy (BCE) measures the difference between two probability maps after a threshold has been applied:

$$\mathcal{BCE}(\mathbf{Y}, \hat{\mathbf{Y}}) = -\mathbf{Y}\log(\hat{\mathbf{Y}}) - (1 - \mathbf{Y})\log(1 - \hat{\mathbf{Y}})) \tag{2.1}$$

As BCE looks at differences between each pixel, it works best for balanced classes, e.g. where the structures are of similar sizes. For highly imbalanced classes, weighted BCE can be used.

2.2.0.2 Dice

Arguably the most popular segmentation metric, Dice Score, or Dice coefficient [75], measures the overlap between the prediction and label, where 0 is no overlap and 1 is full overlap:

$$\mathcal{D}(\mathbf{Y}, \hat{\mathbf{Y}}) = 1 - \frac{2\left|\mathbf{Y} \cap \hat{\mathbf{Y}}\right|}{|\mathbf{Y}| + \left|\hat{\mathbf{Y}}\right|}. \tag{2.2}$$

Although Dice has a relatively interpretable definition and scale, it does not consider spatial context, losing an indication of the distance or size of a segmentation error. To overcome this limitation, distance-wise metrics can be used in combination with Dice.

2.2.0.3 Hausdorff Distance (HD)

The Hausdorff distance ($\mathcal{HD}_{\max}$) complements Dice, as it calculates the maximum distance between the label and segmentation [76].

$$\mathcal{HD}_{\max}(\mathbf{Y}, \hat{\mathbf{Y}}) = \max\left\{\max_{y \in \mathbf{Y}} \min_{\hat{y} \in \hat{\mathbf{Y}}} d(y, \hat{y}), \max_{\hat{y} \in \hat{\mathbf{Y}}} \min_{y \in \mathbf{Y}} d(y, \hat{y})\right\} \tag{2.3}$$

where, y and $\hat{y}$, represent the pixels within $\mathbf{Y}$ and $\hat{\mathbf{Y}}$, respectively, and d is the Euclidean distance between them. However, it can be skewed by outliers, so often

the 95% Hausdorff distance is used, which returns the 95% percentile of the set, $\mathcal{HD}_{95}$. Distance metrics can be measured in pixels or in the spacing defined by the image (e.g. mm, which is used for most medical imaging tasks). Either convention can be used, however, for both, it is important to know the corresponding size of the whole image or structure. Reporting the distance in mm provides a direct physical interpretation of the metric, which may be more relevant to clinicians. On the other hand, pixel measures are often better when comparing distance scores across different size structures, as the size of the structure may be more similar in pixels than mm. As I am not computing average distances between sets of different structures, I report all distance metrics in mm.

2.2.0.4 Average Symmetric Surface Distance (ASSD)

Similarly to $\mathcal{HD}$, average symmetric surface distance (ASSD) computes the average distance from the points on the boundary of $\mathbf{Y}$ to the boundary of $\hat{\mathbf{Y}}$, and vice versa.

2.2.0.5 Relative Error

Another way to assess segmentation performance is to compare a shape property ρ (such as volume, surface area or perimeter) between the label and prediction and compute the relative error between the two results

$$\delta = \frac{\left|\rho(\hat{\mathbf{Y}}) - \rho(\mathbf{Y})\right|}{\rho(\mathbf{Y})}. \tag{2.4}$$

Although δ provides an important insight into the segmentation, it can also be misleading. For example, the segmentation might have no cross-over with the real label but by chance might have a similar volume. Therefore, it is important to use this metric in combination with the distance and overlap metrics.

2.2.0.6 Topology

The metrics described above provide a good overview of the overlap and distances between the label and ground truth, however, they do not directly assess whether the segmentation has the correct *topology*. The topology of an object describes

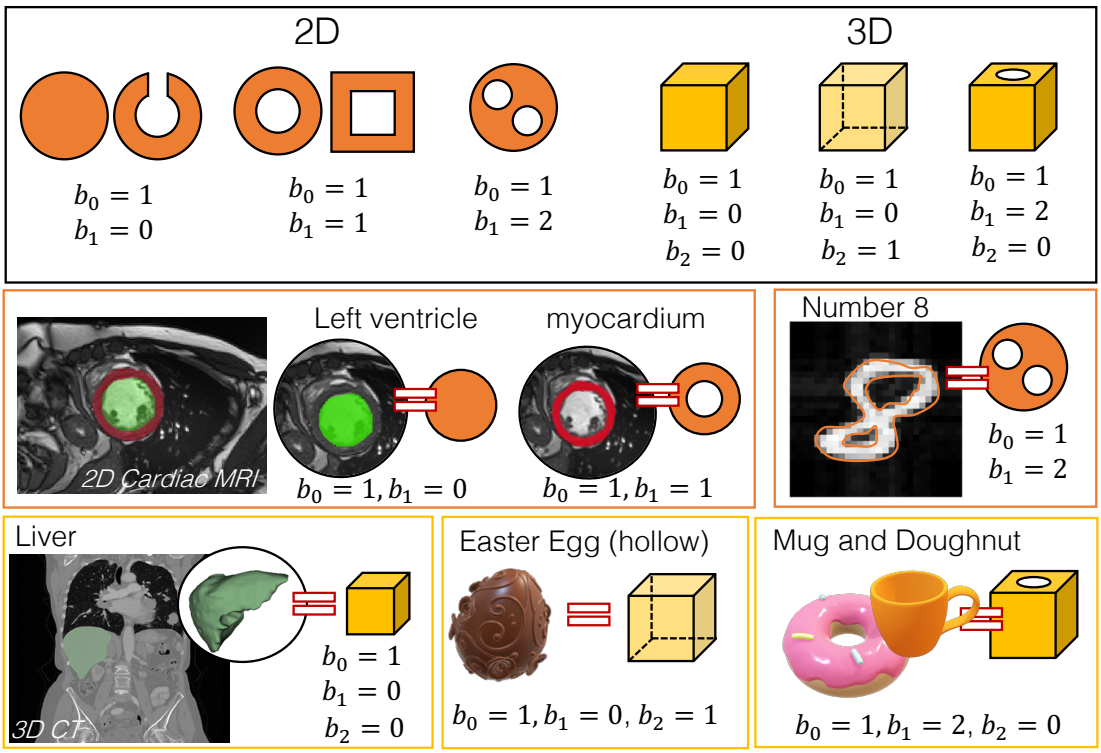


Figure 2.5: An overview of topology in 2D and 3D shapes. The top row shows a set of basic shapes and their Betti numbers, which have the equivalent topology to more complex shapes, shown in the bottom two rows.

its global shape, more specifically, it describes the properties of an object that are preserved through continuous deformations, such as twisting and stretching, without deformations such as tearing, glueing and hole filling [77]. In medical imaging, the topology of many structures is consistent across patients. Therefore, if a segmentation has the incorrect topology, it might be anatomically implausible, which in turn could lead to misleading (or just incorrect) measures in subsequent steps in the medical imaging pipelines, as previously illustrated in Figure 1.1.

Simply, topology is the measure of connected components and holes within a structure. A doughnut and a mug have the same topology, as they are both a single connected component with a hole (the handle or doughnut hole). Therefore, one can be deformed (without tearing, joining or twisting) to resemble the other.

To assess topology, the number of components and holes within a shape can be counted. Betti numbers are a sequence of numbers that report the number of 'holes' an object has in each dimension b_{dim} and the total number of connected

components b_0. Figure 2.5 shows an overview of simple shapes in both 2D and 3D, and their Betti numbers. In the case of the square with a hole, the Betti number is $b_0 = 1, b_1 = 1$, as the hole is counted as 1-dimensional. For the hollow cube, there are zero 1-dimensional holes $b_1 = 0$ but now a 2-dimensional hole within the hollow shape $b_2 = 1$.

To assess topology, the difference between each Betti number (dB) can be measured:

$$dB = \sum_{n=0}^{\dim} b_n(\mathbf{Y}) - b_n(\hat{\mathbf{Y}}), \tag{2.5}$$

and if any values differ, the segmentation is said to have the incorrect topology $\mathcal{L}_{topo} = 0$, such that:

$$\mathcal{L}_B = \begin{cases} 1 & \text{if } dB = 0 \\ 0 & \text{otherwise} \end{cases} \tag{2.6}$$

The topology of a structure can also be calculated by measuring the Euler characteristic χ. The Euler characteristic is a measure of shape, where all topologically equivalent shapes will have the same χ. It can be calculated as $\chi = V - E + F$, where V, E, and F are the number of vertices, edges and faces within a shape. Alternatively, it can be calculated using Betti numbers: $\chi = b_0 - b_1 + b_2$. Across this thesis, I choose to use Betti numbers rather than χ to measure topology, as it provides additional information about the types of topological errors (e.g. holes or number of connected components).

Topology is seldom assessed in deep learning segmentation evaluation, despite its ability to reveal fundamental mistakes within the segmentations. Moreover, as such errors can be obvious on visual inspection, clinicians are unlikely to trust segmentation algorithms that produce such mistakes.

2.2.1 Segmentation Quality

The range of the metrics described above is required to accurately assess segmentation performance. Table 2.1 shows the performance of four hypothetical segmentation methods (A-D) compared to ground truth label($\mathbf{Y} \in \mathbb{R}^{1050\times1050}$)

Table 2.1: The segmentation performance between the ground truth and example segmentations in Figure 2.6.

	Dice (↑)	HD (↓)	HD 95^{th}(↓)	ASSD (↓)	BCE (↓)	SA Error (↓)	Topology (↑)
A	81.8	**86.0**	**53.0**	**5.6**	1.0	0.30	**1**
B	83.1	322.4	210.9	21.7	1.1	**0.01**	**1**
C	**95.8**	311.0	198.0	15.5	**0.27**	0.05	0
D	90.3	311.0	198.0	18.4	0.70	0.19	0

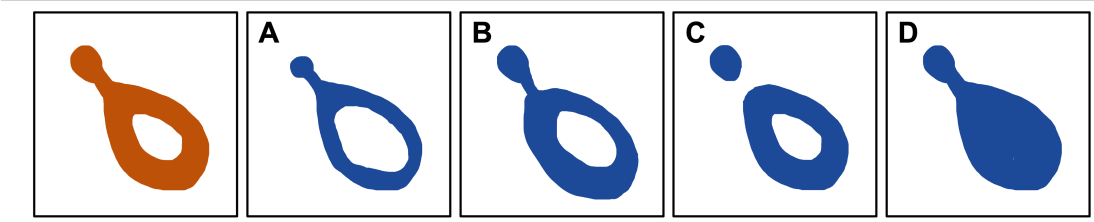


Figure 2.6: A toy example of a ground truth label (shown in orange) and example segmentations (A-D).

From Table 2.1, we get an idea of what the underlying segmentations must look like. Segmentation A has the correct shape (based on topology), however, is likely too thin or thick, judging by the drop in Dice and BCE and low distance metrics. Segmentation C has an incorrect shape, likely leading to poor distance metrics, but given the good Dice and BCE performance, it is likely the segmentation is well delineated despite the topological error. Segmentation A, B, and C all perform best on one or more of the metrics, but none unanimously outperform the others. So which method performed the best? If the evaluation had only been performed using Dice, Segmentation C would likely be chosen, despite resulting in the incorrect shape. Conversely, if the algorithms were purely assessed using SA error, Segmentation B would have been chosen, despite it having the worse distance metrics. This simple example was created to show that choosing the best segmentation network is not always trivial, and a wide range of metrics are required to describe the segmentation.

Now visually inspecting the segmentations shown in Figure 2.6, can the best method be chosen? The answer is likely to change depending on why segmentations were required in the first place. In medical imaging, segmentations are usually used as an intermediate step, to get an automated measure of shape or size, or to track movements. If a segmentation was required to provide an automated

measure of the area of the background enclosed, or the perimeter of a structure, then Segmentation C or D, respectively, would be the most suitable. However, if the segmentation was required to get an assessment of shape, then Segmentation A or B would probably be the most suitable.

This example illustrates that a range of segmentation metrics is required to get an accurate assessment of segmentation quality. Often in deep learning segmentation tasks, only one metric (usually Dice) is assessed, which does not provide a full picture of the segmentation performance. Further, topology is almost always neglected. Moreover, the medical question we are trying to answer using the segmentation often dictates which measure should be more weighted than others. Across this thesis, I explicitly try to enforce segmentations to have the correct topology, to ensure they are anatomically plausible, whilst maintaining a high overlap and low distance measures.

Throughout this thesis, to quantify segmentation quality I use the set of measures described across this section. As shown in Table 2.1, a range of measures are required to accurately describe the behaviour of the segmentations in Figure 2.6. Each metric provides a different perspective of the quality of the segmentation, whilst combining them helps provide a more comprehensive evaluation. The Dice score provides a strong measure of overlap, however, it does not assess overall shape or spatial context and therefore should be used with complimentary measures of topology and distance metrics (e.g. HD and ASSD). Moreover, measures such as topology should not be used on their own, as it is possible for segmentation to have the correct topology but zero overlap with the structure of interest. Thus throughout this thesis, Dice and BCE are used to assess the overlap of our segmentation, whereas HD and ASSD provide measures of distance, whilst topology assesses the overall shape of the structure.

2.2.2 Loss Functions

Loss functions are used to train a network by evaluating its current state and updating the weights within the network accordingly to minimise the loss. This process is often done through stochastic gradient descent with backpropagation.

Gradient descent is an optimisation algorithm that finds the best set of model parameters θ that result in the smallest loss value $\mathcal{L}$. Whereas, backpropagation is the algorithm used for calculating the gradient of a $\mathcal{L}$ with respect to θ. Therefore, to use these algorithms, $\mathcal{L}$ must be differentiable.

Soft Dice (1-Dice) and cross-entropy are commonly used as loss functions for segmentation. However, as these functions perform pixel-wise evaluation, they do not learn the importance of the higher-level structures within the segmentation, which can often result in topological errors with the segmentations. To overcome this, pixel-wise loss functions can be used with topology-aware functions.

2.3 Topology in Medical Segmentation

So far in this chapter, I have discussed how typical deep learning segmentation methods, e.g. transformers and CNNs, can lead to topology mistakes, due to pixel-wise optimisation or inductive biases. Moreover, I have shown how often these mistakes go unnoticed, due to typical evaluation metrics. In recent years, there has become an increasing interest in ensuring topologically accurate segmentations. In this section, I overview such methods and divide them into two groups: *pixel labelling* and *segmentation through deformations*.

2.3.1 Pixel labelling

In recent years, a plethora of methods have been proposed to encourage semantic segmentation networks to learn topological features, overcoming the limitations of pixel-wise optimisation. For instance, a series of topological loss functions have been designed to work in combination with pixel-wise functions, to encourage networks to focus on the higher-level shape information [17–19, 26, 78]. The majority of these functions have been based on *persistent homology* and have been demonstrated on a range of computer vision and medical imaging tasks. In short, persistent homology measures the topological features of a structure across of range of spatial resolutions and concludes that persistent features across this range are most likely to represent the true features of a structure [79].

In Hu et al. [26], the authors proposed one of the first methods, referred to as *TopoLoss*, which compares the persistent homology and Betti numbers (previously discussed in Section 2.2) between the ground-truth labels and the predicted probability maps. As this method was found to be extremely computationally expensive, TopoLoss was only applied to small patches of the images. This procedure, however, was also found to create small disconnected components throughout images, which have to be removed through post-processing operations, such as morphological opening. Moreover, loss functions can only implicitly encourage topology preservation, with the best-reported method achieving 96.7% topological accuracy [79].

Higher-level shape information of desired structures has also been incorporated into CNNs using additional networks, tasked with learning the topology. In Oktay et al. [61] an auto-encoder, trained to extract shape information from a set of anatomical labels, was used to regularise a separate segmentation network. Alternatively, in Mosinska et al. [80] the feature maps from a pre-trained VGG-Net were used as descriptors of higher-level shape information and were incorporated into a final segmentation network through a topology-aware loss function. Although these methods report promising pixel-wise performance and provide prior knowledge of higher-level structure, it is unclear which shape features the networks focus on, and how relevant they are to the structure's overall topology. In Cao et al. [81], the authors used multi-task learning to learn the topology of stacked retina layers, where a U-Net style architecture with an additional decoder path was used to predict signed distance maps of the boundaries between individual retina layers. This soft constraint gave the network some topology guidance, however, as the topology performance was unreported, it is unclear how strict the topology preservation is. Moreover, this method is designed for a very specific segmentation problem and its applicability to new datasets is unknown.

Although the methods described above provide additional shape information, they are still ultimately limited by the method of labelling individual pixels rather

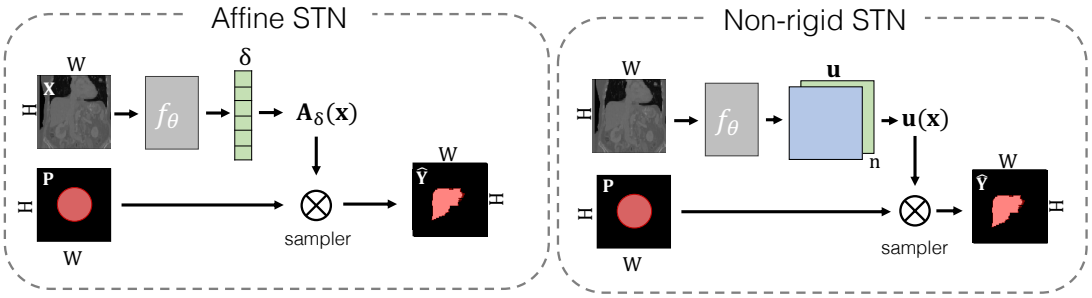


Figure 2.7: An overview of spatial transformer networks (STN) originally presented in [82], shown for both affine and non-rigid transformations. STNs can be integrated into different networks (f_θ), from CNNs to transformers. In this the sampler applies the deformation to the prior $\mathbf{P}$, to find the prediction $\hat{\mathbf{Y}}$.

than a whole structure. Further, the proposed methods only softly constrain the networks to focus on topology.

2.3.2 Segmenting through Deformation

When performing atlas-based segmentation, atlas labels are deformed in order to delineate a new image. A similar approach has been integrated into CNNs, through the use of spatial transformer networks (STN) [82]. STNs are flexible modules that can be integrated into any network (f_θ) and in general, are tasked with learning a deformation field ($\mathbf{\Phi}$), whose type can vary, in order to transform a target image or prior.

STNs were first proposed to improve classification tasks [82]. In this, a CNN was used to extract relevant features from an input image ($\mathbf{X}$) in order to create an affine transformation ($\mathbf{A}$) to align $\mathbf{X}$ to improve the classifier's predictions. They have since been adapted for segmentations [83–86]. In such methods, in 2D[5], a learnt deformation field $\mathbf{\Phi} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ is used to transform a prior shape ($\mathbf{P} \in \mathbb{R}^{H \times W}$) in order to complete a segmentation task $\hat{\mathbf{Y}} = \mathbf{\Phi}(\mathbf{P})$, where $\hat{\mathbf{Y}} \in \mathbb{R}^{H \times W}$, as shown in Figure 2.7. The networks are optimised by comparing the warped shape ($\hat{\mathbf{Y}}$) to the ground truth label ($\mathbf{Y}$), often using BCE or Dice loss. In the rest of this thesis, when I refer to STNs, I am referring to STNs adapted for segmentation.

[5]In this section I mainly refer to 2D images for ease, however, the concepts also apply across different dimensions.

When using STNs, the true topology is often explicitly incorporated into the network through $\mathbf{P}$. However, the topology of $\mathbf{P}$ is not guaranteed to stay the same after a deformation. As discussed in Section 2.2.0.6, the topology of an object describes properties that are preserved through continuous deformations. Therefore, to enforce the preservation of topology, continuous, also referred to as *topology-preserving fields* must be used [87].

In this section, I outline previous STN methods and how different types of $\mathbf{\Phi}$ can be produced and constrained. Further, I describe how both linear and nonlinear transforms can be topology-preserving, and review the previous methods that have utilised this property.

2.3.2.1 Affine Transformation

An affine transformation restricts the transforms to be in the family of affine matrices $\mathbf{A}$, which can include: rotation, scaling, shearing, translation or a combination of the above. In this, the same matrix $\mathbf{A}$ is applied globally. These transformations are always topology-preserving, as they can not tear or connect parts of the structure [88].

In Wickramasinghe, Knott, and Fua [85], the authors generated probabilistic prior shapes constrained by the known topology of each structure of interest. They extracted features from an input image using a U-Net style architecture that had additional linear layers branching off at the bottleneck, which were used to predict an affine transformation vector δ, as shown in Figure 2.7. The affine transformation was then used to deform a set of priors. The deformed shapes were then concatenated onto each layer of the decoder branch. This method was evaluated for the segmentation of a mouse's synaptic junction and optic disk from histology and retinal images, respectively. Although this method achieved promising segmentation results, topological errors were still present, which was likely due to image resampling: this limitation will be discussed later in this section.

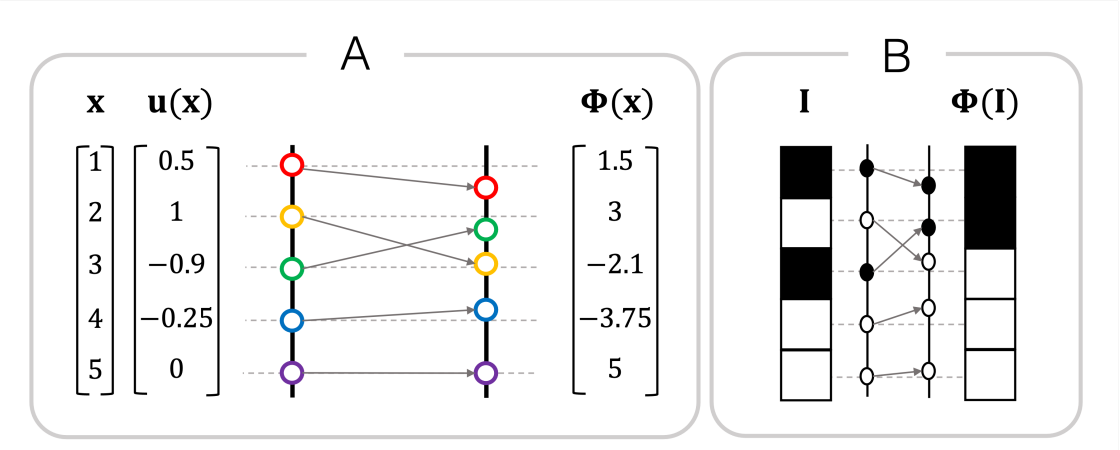


Figure 2.8: Panel A, shows a simple 1D example of displacement ($\mathbf{u}(\mathbf{x})$) and deformation fields $\mathbf{\Phi}(\mathbf{x})$, as shown as 1D vectors defined over a grid $\mathbf{x} : (1, 2, 3, 4, 5)$. Panel B shows how $\mathbf{\Phi}(\mathbf{x})$ can be used to deform an image $\mathbf{I}$, using methods such as linear interpolation or nearest neighbour sampling. After deformation, the two images have different topologies. In this simple example, I have shown the transformation in the *push* convention, e.g. showing where $\mathbf{x}$ ends up. However, the *pull* convention, e.g. showing where the information of the warped space has come from, is more common, albeit less intuitive.

2.3.2.2 Nonlinear Transformation

Nonlinear transformations can be far more flexible than affine transforms, as instead of a single matrix describing the transformation at each position, the field can now vary as a function of position $\mathbf{x}$. An intuitive way to model such functions is using a displacement matrix $\mathbf{u}(\mathbf{x}) \in \mathbb{R}^n$ that is a function of $\mathbf{x}$ and describes how much each grid point $\mathbf{x}(i, j)$ is displaced in each direction[6]:

$$\mathbf{\Phi}(\mathbf{x}) = \mathbf{x} + \mathbf{u}(\mathbf{x}). \tag{2.7}$$

In other words, if $\mathbf{x}$ is a 1D grid, $\mathbf{x} : (1, 2, 3, 4, 5)$, $\mathbf{u}(\mathbf{x})$ would be a 1D vector that describes how much each position in $\mathbf{x}$ is displaced, as shown in Figure 2.8. $\mathbf{\Phi}(\mathbf{x})$ then describes the new deformation positions as a reference of the old positions $\mathbf{x}$.

When a deformation is applied to a digital image $\mathbf{I}(\mathbf{x})$, which is described over the same domain $\mathbf{x}$, the resulting warped positions are often no longer on the discrete grid points, as shown in Figure 2.8B. Therefore, a method is required to resample

[6]Although conventions vary across the literature, throughout this thesis I will use $\mathbf{u}$ to refer to displacement fields and $\mathbf{\Phi}$ for deformation fields, where $\mathbf{\Phi}(\mathbf{x}) = \mathbf{u}(\mathbf{x}) + \mathbf{x}$, and $\mathbf{x}$ describes the reference grid where the displacements occur.

(or digitise) the warped image ($\mathbf{\Phi}(\mathbf{I})$) back into the discrete positions. This is usually performed through interpolation, which refers to methods that estimate a pixel value, based on a set of known surrounding values. Popular interpolations methods include *nearest neighbours*, which bases its estimations on the closest known data points; *linear interpolation*, which plots the known pixel values over position and fits a straight line between neighbouring points, intensities for unknown positions can then be read directly from the plot; and, finally, *cubic interpolation*, which similarly to linear interpolation plots values over position but instead of fitting straight lines it fits splines. Interpolation is also performed after Affine transformations. This simple example is just in 1D but the methods generalise to higher dimensions.

In Figure 2.8B, after image deformation and resampling, the new image ($\mathbf{\Phi}(\mathbf{I})$) has a different topology, illustrating how nonlinear transformations can change the topology of a warped image. In this example, there is a large gradient between the displacement at $\mathbf{u}(\mathbf{x} = 2)$ and $\mathbf{u}(\mathbf{x} = 3)$, which results in a change of the image topology.

Diffeomorphism Diffeomorphisms are a popular transformation that guarantees topology preservation. If a transform is not diffeomorphic, then topology is not necessarily preserved [89]. Diffeomorphisms are continuous, preventing the structure from being ripped (creating holes or disconnected components) or joined together (closing holes and joining disconnected components). That is, they have a continuous inverse and bijective (e.g. resulting in a one-to-one mapping) at all locations [89, 90]. Often transformations are encouraged to have diffeomorphic behaviour, without strictly being diffeomorphic, by smoothing over large gradients within the field that risk creating holes or joining structures, as seen in Figure 2.8B.

Unconstrained Deformations To build a deformation field, $\mathbf{u}(\mathbf{x})$ can be estimated directly, known as a free-form deformation [89]. In its simplest form, $\mathbf{u}(\mathbf{x})$ is defined independently at each grid point ($\mathbf{x}(i, j)$). In other words, if $\mathbf{u}(\mathbf{x})$ was modelled as a matrix, a displacement in each direction would be assigned to each pixel within $\mathbf{u}(\mathbf{x})$. This results in very flexible but poorly constrained deformations.

This approach has been commonly incorporated into segmentation STNs [83, 84, 86]. In these approaches, at the end of the CNN, a 1×1 convolution is usually applied to predict an n-dimension displacement field $\mathbf{u}(\mathbf{x}) \in \mathbb{R}^n$, rather than predicting a c class segmentation map, as shown in Figure 2.7. The deformation is then used to transform $\mathbf{P}$, denoted $\mathbf{\Phi}(\mathbf{P})$, where $\mathbf{P}$ is chosen by the user. However, such methods have no topological guarantees.

As the fields are unconstrained, it is common to use a regularisation term in the loss function to encourage smooth[7] or small deformations, to try and encourage diffeomorphic behaviour. Two commonly used functions are $\mathcal{L}_{\text{smooth}}$ and $\mathcal{L}_{\text{L2}}$. $\mathcal{L}_{\text{smooth}}$ penalises the sum of squared first or second-order derivatives, to encourage locally smooth deformations, whereas $\mathcal{L}_{\text{L2}}$ penalises the $L2$-norm of the deformation vectors to encourage minimum displacements [86]. Although these loss functions encourage the fields to have some diffeomorphic properties the fields are fundamentally unconstrained and therefore unlikely to be topology-preserving.

Constrained Deformations Constrained deformations refer to transformations that are restricted to behave in a certain way, e.g. topology-preserving. Although diffeomorphisms can be enforced in an unconstrained manner, it is often simpler to generate them using constraints. An elegant property of diffeomorphisms is that when one diffeomorphism is resampled (also known as a *composition* and denoted $\circ$) by another, the resulting field will also be diffeomorphic [89]. In other words, if $\mathbf{\Phi}^1$ is diffeomorphic and is applied to another diffeomorphic transform $\mathbf{\Phi}^2$: $\mathbf{\Phi}^1 \circ \mathbf{\Phi}^2 = \mathbf{\Phi}^3$ (or $\mathbf{\Phi}^1(\mathbf{\Phi}^2) = \mathbf{\Phi}^3$), then the new field $\mathbf{\Phi}^3$ will also be diffeomorphic. Therefore, by initialising with a diffeomorphism and performing a composition, the new field is *constrained* to be diffeomorphic, whilst the magnitude of the displacement vectors can be amplified.

Deformation fields containing negligible displacement vectors are by definition diffeomorphic. In the discrete domain, this can be thought of as the displacement

[7] In mathematics, smoothness refers to continuous derivatives across all orders. However, in this thesis, I use this term more loosely, referring to fields without large gradients between neighbouring regions that risk creating holes or joining structures.

being so small that neighbouring grid points can not cross over each other. Thus, by initialising with a field containing negligible displacements and performing a series of compositions, a large displacement diffeomorphism can be produced. One efficient method to generate such fields, is the *Scaling and Squaring* approach [89, 91, 92]. In this, a negligible field ($\mathbf{\Phi}^1$) is first applied to itself $\mathbf{\Phi}^1 \circ \mathbf{\Phi}^1 = \mathbf{\Phi}^2$, before the process is repeated using the new field h times:

$$\begin{aligned} \mathbf{\Phi}^1 &= \mathbf{u}(\mathbf{x}) + \mathbf{x} \\ \mathbf{\Phi}^2 &= \mathbf{\Phi}^1 \circ \mathbf{\Phi}^1 \\ \mathbf{\Phi}^4 &= \mathbf{\Phi}^2 \circ \mathbf{\Phi}^2 \\ &\vdots \\ \mathbf{\Phi}^{2^h} &= \mathbf{\Phi}^{2^{h-1}} \circ \mathbf{\Phi}^{2^{h-1}}. \end{aligned} \tag{2.8}$$

In the continuous setting, this method is equivalent to the evolution of a flow field that remains constant over time, with each composition representing a discrete unit of time [89].

Diffeomorphisms have commonly been used in image registration [89, 93–95], where the displacement fields are iteratively updated until acceptable registration is observed, a process found to be extremely time-consuming. With recent advances in computational hardware, registration methods have begun to be integrated into deep-learning-based registration methods, such as the popular VoxelMorph, which registers two MRI brain volumes together [25, 40, 96]. In this, a CNN learns the deformation required to warp a *moving* volume onto a *fixed* volume [25, 40, 96]. In recent versions of their approach, methods have been incorporated to try and guarantee topology preservation by generating diffeomorphic transformations, through the use of Scaling and Squaring layers, pioneering the integration of diffeomorphism into CNNs [25, 40, 97]. Recent work has adapted popular diffeomorphic-registration methods, to propose a joint registration and segmentation framework [98], with a focus on topology preservation in the predicted segmentation. However, the final topology performance is often unreported [98].

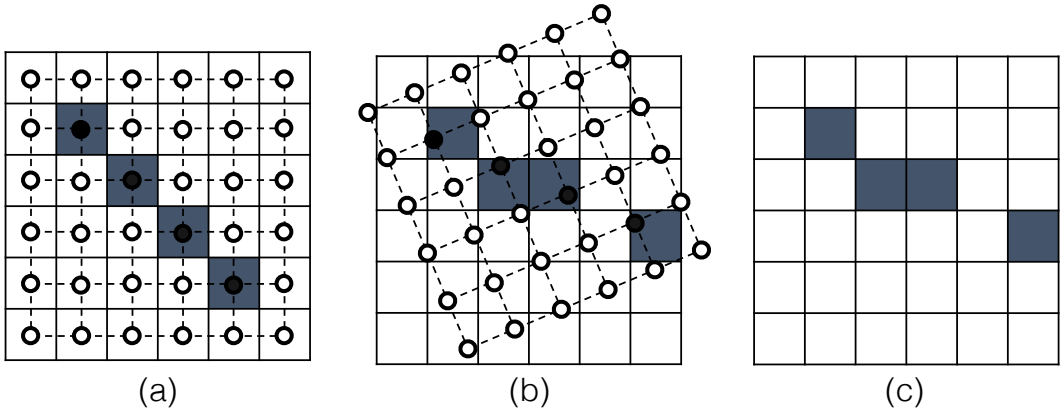


Figure 2.9: An example of how deforming a discrete image using a discretely represented field can result in topological violations. (a) shows a binary image of a line, which is connected at the corners and has a Betti number of $b_0 = 1, b_1 = 0$. (b) shows a rigid, topology-preserving transform applied to the grid, resulting in (c). The topology within (c) has now changed from (a) to $b_0 = 2, b_1 = 0$ as a result of the resampling. The inspiration for this figure was taken from an example in [99].

2.3.2.3 Implications of Discrete Representations

Diffeomorphisms guarantee topology preservation and are often utilised for this property. They are inherently continuous but in practice, they are often represented discretely with a finite number of parameters, where their guarantees can begin to break down [89]. This is predominantly due to the resampling and interpolation required when storing and applying such transformations. For example, when working in a discrete setting, fields are often stored over discrete grid points in an n-dimensional matrix, which risks losing valuable information between points. Then, in order to use the field, interpolation must be applied to compute the deformation between the discrete grid points. However, after interpolation, it is unlikely that the transformation will behave the same as the original transformation. Further, interpolation and resampling can create *folding-voxels* within the deformations, which can lead to topological violations. Folding-voxels occur when two (or more) grid points cross over each other during the deformation, as shown in Figure 2.8, resulting in discontinuities within the field that can lead to a change in topology.

In the case of the discretely implemented Scaling and Squaring layers, the diffeomorphism generated after each composition must be discretely stored before

being interpolated again in order to perform the next composition, which likely changes the nature of the field. Further, during this process, if folding-voxels are created, they can be amplified in subsequent composition layers. Therefore, due to these violations, when working in the discrete domain transformations can no longer be classified as diffeomorphic. Across this thesis, I try to generate a discrete approximation of diffeomorphism, which I refer to as topology-preserving fields. These effects will be further explored in Chapter 4.3.2, but for now, it is important to note that the methods used to generate and store topology-preserving fields in the continuous domain do not directly translate to the discrete domain.

Representing fields discretely can change their behaviour. However, even if a field was truly topology-preserving, topology can still be violated when resampling a deformed image back into a discrete space. Figure 2.9 shows an example of such violations when storing an image at discrete grid points and applying a rigid-body (e.g. only rotation and translation) transformation, which by definition is topology-preserving, before resampling the image back into the discrete domain. In this, the act of resampling changes the original topology and is therefore another effect that must be accounted for.

2.3.2.4 Discrete Topological Modifications

The limitations of resampling images and fields, and composing discretely represented fields are well-known across the registration community. To overcome the possible violations, additional modifications have been proposed to more strictly encourage topology preservation.

Field Smoothing In addition to field regularisation through loss functions discussed above, field smoothing can be applied directly to the deformation fields, limiting large gradients between the discrete grid points. One method of doing this is by applying Gaussian smoothing, of sigma σ, directly to the fields. This technique is commonly used in Demon registration algorithms [100], where registration is performed iteratively, by computing the similarity between a fixed and moving target at each step, with Gaussian smoothing applied at each iteration, leading

to a smoother registration field [93, 100]. An alternative method is to instead apply Gaussian smoothing between each intermediate composition layer (e.g. in the Scaling and Squaring algorithm, shown in Equation 2.8), known as *fluid-like* registration [101]. This method is more commonly used when large magnitude displacements are required, as the smoothing is only applied at intermediate steps, rather than to the final fields, allowing the learned fields to remain relatively flexible.

Measuring Folding-Voxels To assess whether a generated field is topology-preserving, the *Jacobian determinant* can be measured at each position, to detect folding voxels within the field [87]. The Jacobian determinant is based on gradients around a grid point $\mathbf{x}$, which describes how the transformation is changing the volume around a point. To calculate the Jacobian determinant, the Jacobian matrix is first measured. The Jacobian matrix of a field ($\mathbf{\Phi} : \mathbb{R}^n \rightarrow \mathbb{R}^n$) measures the partial derivatives ($\frac{\partial \mathbf{\Phi}(\mathbf{x})}{\partial x_n}$) across each dimension to assess the gradient in the n-dimension:

$$\mathbf{J} = \begin{bmatrix} \dfrac{\partial \mathbf{\Phi}(\mathbf{x})}{\partial x_i} & \cdots & \dfrac{\partial \mathbf{\Phi}(\mathbf{x})}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \frac{\partial \mathbf{\Phi}_1(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial \mathbf{\Phi}_1(\mathbf{x})}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial \mathbf{\Phi}_n(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial \mathbf{\Phi}_n(\mathbf{x})}{\partial x_n} \end{bmatrix}. \tag{2.9}$$

The determinant of the $\mathbf{J}$ gives the factor by which an area or volume changes when a transformation is performed and is denoted: $|\mathbf{J}|$. In 2D, where $\mathbf{x} \in \mathbb{R}^2$, such that $\mathbf{x} = [x, y]$, the Jacobian matrix would be: $\mathbf{J} = \begin{bmatrix} \frac{\partial \mathbf{\Phi}_x(\mathbf{x})}{\partial x} & \frac{\partial \mathbf{\Phi}_x(\mathbf{x})}{\partial y} \\ \frac{\partial \mathbf{\Phi}_y(\mathbf{x})}{\partial x} & \frac{\partial \mathbf{\Phi}_y(\mathbf{x})}{\partial y} \end{bmatrix}$

The Jacobian determinant of a field $\mathbf{\Phi}$ is a scalar function of the coordinates such that it describes the behaviour around each point (x) as being either expanded: $|J_{\mathbf{\Phi}}(\mathrm{x})| > 1$, contracted: $0 < |J_{\mathbf{\Phi}}(\mathrm{x})| < 1$, or folded: $|J_{\mathbf{\Phi}}(\mathrm{x})| \leq 0$. Topology-preserving fields must have positive Jacobian determinants at every grid point.

In recent work, topology preservation in an STN has been encouraged through a novel loss term that penalises folding voxels by measuring the Jacobian determinants within the predicted deformation field [102]. However, both the final topology performance and Jacobian determinant values are left unreported, and as topology preservation is only implicitly encouraged, it is very possible that folding voxels are still present.

2.4 Conclusion

Across this chapter, I have shown how deep-learning-based methods have rapidly transformed the field of medical segmentations, producing automated and *accurate* delineations. I have discussed how segmentation quality can be defined in a variety of ways and how a range of quantitative metrics are required to truly assess a segmentation network's performance. Further, I have demonstrated that optimising networks solely on pixel-wise loss functions can lead to the networks overlooking the importance of the higher-level structural information, e.g. topology. Additionally, CNNs, the most commonly used segmentation architecture, often fail to learn a structure's topology due to the local convolutions used to extract features, which further risks topological errors.

Producing a topologically-accurate segmentation is important to ensure that they are fundamentally usable for downstream analysis tasks. In recent years, the importance of topology-aware deep-learning methods has been highlighted with new methods proposing topology-aware loss functions or integrating shape priors into the networks [18, 19, 26, 79]. However, such methods have been shown to help decrease, but not eliminate, the frequency of topological errors. To enforce the preservation of topology, topology-preserving fields are commonly utilised, with methods incorporating such fields becoming increasingly popular in both the machine learning registration and segmentation communities [40, 82, 103]. However, the guarantees afforded by continuous topology-preservation begin to break down when working with fields in the discrete domain [103].

Across this thesis, I try to tackle the task of automated medical image segmentation whilst trying to ensure our methods result in topologically accurate results. More specifically, I design a network to produce topology-preserving deformations, in order to deform a prior that has the correct topological characteristics. To produce topology-preserving fields in the discrete setting, I include additional smoothing modifications discussed in this chapter to try and overcome the limitations of discrete sampling and storing, such as field regularisation and Gaussian smoothing.

3

Review of the Fetal Cortical Plate

This chapter contextualises the importance of characterising the development of the fetal cortical plate from in-utero ultrasound scans and why it requires topologically-correct segmentations. In this, I introduce what the cortical plate is, why it should be studied and how it is currently monitored in routine clinical care, before reviewing previous work that has characterised its complex growth.

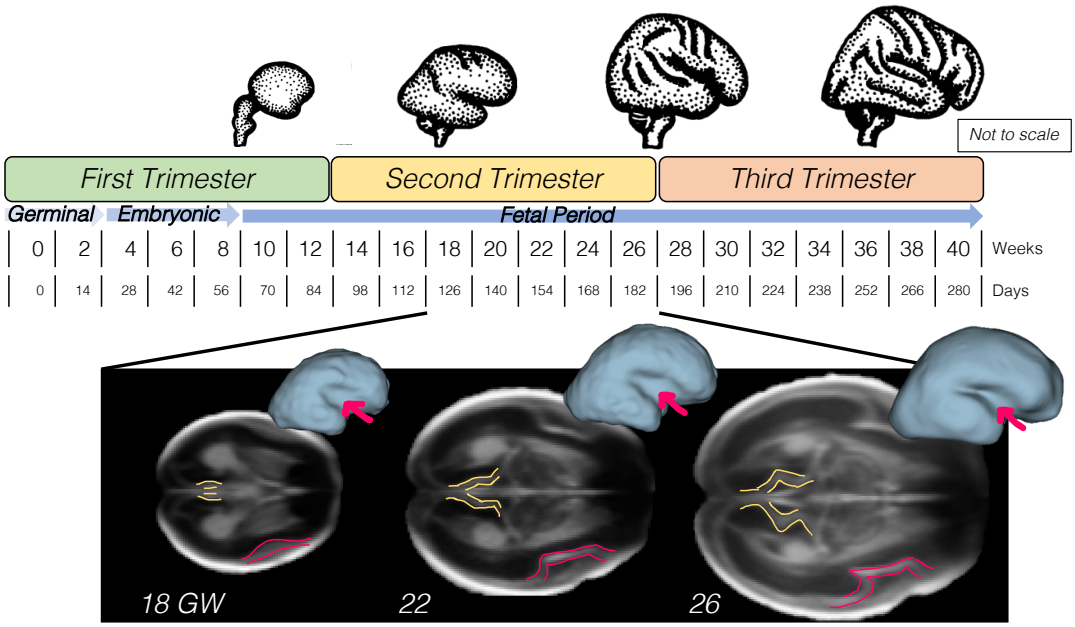


Figure 3.1: The appearance of the fetal brain across gestation. The timeline of fetal growth is given in both gestational weeks (GW) and days, with the former predominantly used throughout this thesis. The development across the second trimester is shown using an ultrasound brain atlas for three ages. In these volumes, the Sylvian Fissure (SF) and the Parieto-occipital Fissure (POF) are highlighted in pink and yellow, respectively. The schematic drawings on the top row were adapted from [104]

3.1 Fetal Brain Development

Throughout the course of pregnancy, the fetal brain undergoes dynamic morphological changes. Figure 3.1 provides an overview of fetal brain development across *gestation*; the period of development inside the womb. As the fetal brain develops, the once smooth outer layer of the cerebrum starts to become heavily folded, a process known as *gyrification* [105–108]. The development of the peaks and troughs, known as *gyri* and *sulci*, across the cortex means the cortical surface area scales almost linearly with volume [109], which is associated with increased cognitive abilities [110, 111]. In healthy development, the timetable of folding is found to be temporally consistent across gestation, with variations from the norm often found in at-risk or diseased pregnancies [112–115]. Therefore, monitoring the developing surface can provide a key insight into fetal health, with the potential to detect biomarkers.

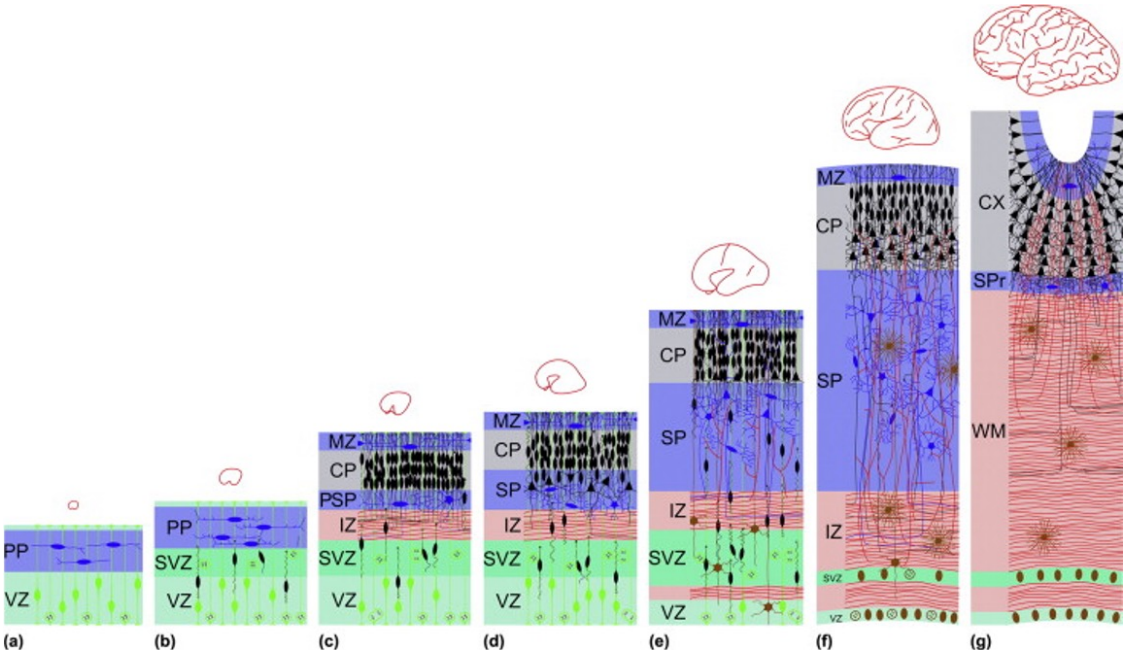


Figure 3.2: Figure taken from [117]. The organisation of the migrating neurones throughout fetal development, to show what the cortical plate (CP) is. The schematic shows the growth from the early embryonic period (a) to the late fetal period (g), where the brain begins to resemble that of an adult. In the embryonic period, the brain is made up of the preplate (PP) and ventricular zone (VZ), before VZ separates to form the subventricular zone (SVZ) and from PP the presubplate (PSP), cortical plate (CP) and marginal zone (MZ) emerge. The beginnings of the white matter (WM) are seen from around the eighth GW in the form of the intermediate zone (IZ). PSP matures to become the distinctive subplate (SP), which continues to grow across gestation.

In this section, I first review what the cortical plate is and how it develops across gestation in healthy pregnancies (Section 3.1.1 and Section 3.1.2) before describing, how this development varies in at-risk or diseased pregnancies (Section 3.1.4).

3.1.1 The Cortical Plate

The brain begins to develop thirteen days after conception and continues to change until death [109]. However, fetal and childhood development is the most rapid and dramatic throughout the lifespan [116]. In adults, cortical folding is commonly studied by segmenting the *grey matter*, but in fetuses this region is immature so the temporary *cortical plate* is studied instead, as shown in Figure 3.2.

The formation of the first well-defined neural structure, *the neural tube*, occurs three weeks after conception during the embryonic period [109]. By the sixth

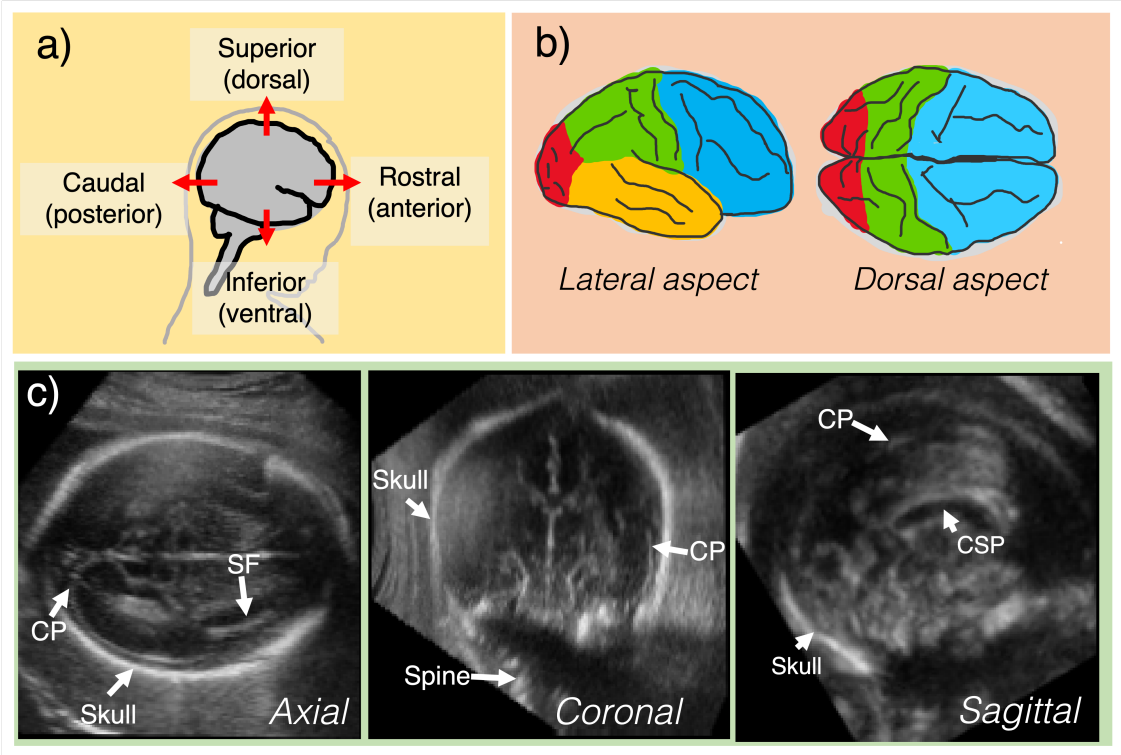


Figure 3.3: An overview of the terminology used when describing fetal brain imaging. Panel a) introduces the language used to describe the relative position in the brain. Panel b) shows the four main lobes, *frontal* in blue, *parietal* in green, *temporal* in yellow, and the *occipital* in red. Panel c) shows how the fetal brain looks in ultrasound imaging across the three different planes: axial, coronal, and sagittal. Across this thesis, the brain will be shown in this orientation.

gestational week (GW), the neural tube expands to make 3 distinct brain vesicles, which further subdivide into 5 secondary vesicles. These are the beginnings of the cerebrum, cerebellum, brain stem, pituitary gland and hypothalamus.

At the beginning of the seventh GW, the cerebrum vesicle has an abundance of neurons in the ventricular zone (VZ), which begin to migrate to form the pre-plate (PP). This later splits creating the cortical plate (CP) and subplate (SP), which contain different neuron types, as shown in the first three panels of Figure 3.2. The neurons continue to migrate to the cortical plate forming a layered structure, with the oldest neurons deepest and the newest neurons forming superficial layers. By the eighteenth GW the cortical plate is made up of six distinct layers, which begin to differentiate forming vast networks by extending into the white matter (WM) [118]. This is the beginning of the white and grey matter layers seen in adults [117].

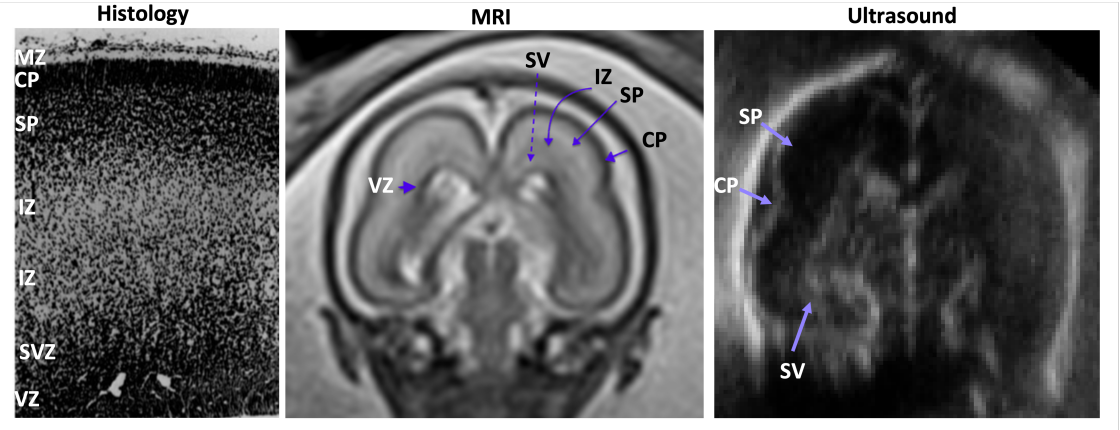


Figure 3.4: The distinct brain layers across three modalities: histology, MRI, and ultrasound at 22 GW. The histology scale bar is $100\mu m$. The MRI image was obtained from [120], the ultrasound volume from the INTERGROWTH-21st dataset, and the histology from [121].

The layers within the cerebral hemispheres can be seen with varying contrast across three different modalities, as shown in Figure 3.4 at 22 GW. In both the histology and MRI image, the cortical plate is dark, due to the densely packed neurons, and there is a large contrast between the cortical plate and surrounding layers. Conversely, in the ultrasound image, the cortical plate appears light, with little contrast between it and the skull or subplate. This is due to how ultrasound signals are acquired, which will be further discussed in Section 3.2.2.

Thickness Neuronal migration continues throughout gestation, which increases the cell density resulting in a thicker cortical plate. Further, the cortex continues to thicken until puberty. After puberty, the grey matter begins to gradually decrease due to neuron pruning [109]. As cell density is tightly correlated to brain function, cortical thickness is found to be inconsistent across the cortical plate in children, with the dorsal regions of frontal and parietal lobes often found to be the thickest and the occipital lobe the thinnest [116] (see Figure 3.3 for terminology). Cortical thickness in the temporal areas has commonly been linked to reading and language function in children, with decreased thickness linked to language impairment and dyslexia [119].

3.1.2 Cortical Folding

At the beginning of the fetal period, the cortical plate is a smooth surface, which by term (e.g. the end of pregnancy ∼40 GW) has transformed into the recognisable heavily folded grey matter, commonly seen in adults, as shown in Figure 3.1. One role of cortical folding is to increase the cortical surface area and volume, whilst being contained within the skull. A larger cortex means more neurones can fit inside it, and thus is found to increase cognitive function. Moreover, folds bring neighbouring neurones closer together, which is thought to increase the speed of brain communication [111]. The cause of cortical folding is still debated, however, the literature suggests it is due to a combination of mechanical and molecular mechanisms [111, 122–124].

3.1.2.1 Folding Timetable

The emergence of individual gyri and sulci has been extensively studied and found to follow a rigid timetable over gestation. An overview of such studies are shown in Table 3.1. The process begins at the tenth GW with the creation of the interhemispheric fissure (IHF), which separates the two hemispheres of the brain and continues growing until 24 GWs [125, 126], as shown in Figure 3.5. A *fissure* indicates a deeper grove within the cortex and is often interchangeable with *sulcus* [127]. The emergence of sulci can be grouped by size and trimester. The first emerging sulci are the largest and develop prior to the 26 GW. Towards the end of the second and the beginning of the third trimester, the cortical surface undergoes rapid transformations with the development of numerous smaller secondary and tertiary sulci [112].

The establish the healthy cortical folding timetable, a series of observation studies have been performed by studying postmortem fetuses [104, 126, 128], in-utero (e.g. whilst fetus is still in the womb) MRI scans [125, 129, 130] and in-utero ultrasound scans [131–133].

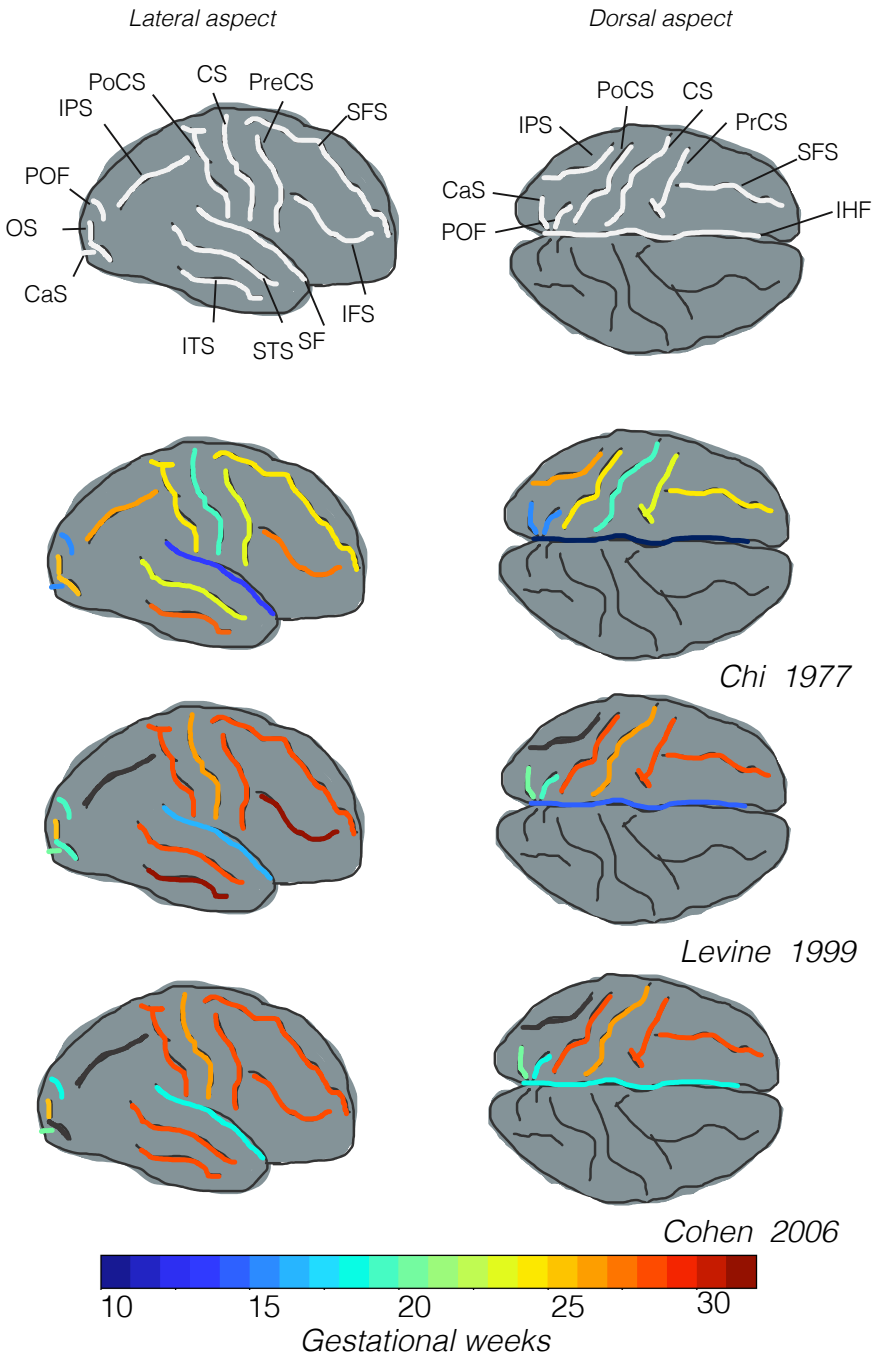


Figure 3.5: The emergence of primary and secondary sulci, gyri and fissures across the developing cortex, identified over three modalities; post-mortem (Chi, Dooling, and Gilles (1977) [126]), MRI (Levine and Barnes (1999) [129]) and ultrasound (Cohen-Sacher et al. (2006) [133]). Note that, from Cohen-Sacher et al. [133] the assessment only started from 18 GW and therefore it is possible that the SF and IHF (both visible at 18 GW) were visible before this date. The fissures described are: central sulcus (CS), superior frontal sulcus (SFS), inferior frontal sulcus (IFS), precentral sulcus (PreCS), superior temporal sulcus (STS), inferior temporal sulcus (ITS), Sylvian fissure (SF), postcentral sulcus (PoCS), Interparietal sulcus (IPS), Parieto-occipital fissure (POF), occipital Sulci (OS), calcarine sulcus (CaS), interhemispheric fissure (IHF)

Table 3.1: Studies describing sulcal, gyri and fissure emergence across different modalities (PM: postmortem and US: Ultrasound). **N** describes the number of postmortem brains or scans analysed. **Method** describes which technique was used to detect the growth of folds, *Ob* indicates observation and *Curv* curve fitting algorithms.
† - months after birth.
‡- number of pregnancies (not scans), which were followed up every 2 weeks.

Authors	Age Range (GW)	N	Modality	Method
Chi, Dooling, and Gilles (1977) [126]	10 - 44	507	PM	Ob
Dorovini-Zis and Dolman (1977) [128]	22 - 1†	80	PM	Ob
Hansen et al. (1993) [104]	6 - 28	NA	PM	Ob
Garel et al. (2001) [125]	22-38	173	MRI	Ob
Levine and Barnes (1999) [129]	14 - 38	53	MRI	Ob
Habas et al. (2012) [134]	20 - 28	40	MRI	Curv
Yun et al. (2022) [130]	19 - 36	106	MRI	Curv
Toi, Lister, and Fong (2004) [132]	15 - 29	50	US	Ob
Monteagudo and Timor-Tritsch (1997) [131]	14 - 40	262	US	Ob
Bernard et al. (1988) [22]	10 - 37	70	US	Ob
Cohen-Sacher et al. (2006) [133]	18 - 40	21‡	US	Ob

Postmortem Evaluating the morphological development of the cortical plate was first studied using postmortem histology [126, 128]. Chi, Dooling, and Gilles [126] examined 507 fetal brains between 10 to 44 GW[1] and recorded the sequential development of sulci and gyri across both hemispheres [126]. They performed schematic drawings to demonstrate the expected changes over this wide gestational period. This work was the first in-depth guide of sulcal development, reporting the Sylvian fissure (SF) emergence from 14 GW and describing left-to-right asymmetries in gyri and sulci timings, with the right cerebral hemispheres reaching gyral complexity earlier than the left. An overview of their findings for specific fissures and sulci is shown in Figure 3.5.

In Dorovini-Zis and Dolman [128], the authors described the appearance of the cortex between 22 GW to one month of postnatal life for 80 fetuses, which was found to support the previous findings [126]. It was reported that between 28 to 30 GW, the fetal cortex underwent the largest change, which became more challenging to assess using 2D histology slices alone. With advanced in medical imaging, Hansen et al. [104] performed a similar qualitative investigation using

[1]Throughout this work I will refer to gestational age ranges like this, where the first value starts at the beginning of that week, e.g. $10^{+0\text{days}}$, and the second, goes up until the end of that week e.g. $44^{+6\text{days}}$

postmortem embryos and fetuses between 6 to 28 GW using magnetic resonance (MR) instead of histology. Their timetable of sulci events was found to have a high correspondence with the two previous histology studies [104, 126].

Although these postmortem studies provided the first timetables of folding across the developing surface, they may not accurately describe healthy development. Due to the nature of postmortem subjects, there is a high uncertainty of brain normality, as samples are often obtained from stillbirths and abortions, where cognitive health can not be confirmed. Therefore, the timetable of gyrification events described by the studies above is not guaranteed to be representative of the healthy population. For example, in Dorovini-Zis and Dolman [128], 72% of the brains studied were from preterm fetuses, which have been associated with abnormal folding, further discussed in Section 3.1.4. Moreover, the process of extracting, fixing and chemically preserving the brain tissues risks damaging or disturbing the underlying structure and the environment does not mimic the structures as seen in-utero.

MRI As imaging techniques improved, it became possible to scan the fetal brain in-utero, which allows follow-up cognitive assessment to ensure healthy development is being studied. In Garel et al. [125] and Levine and Barnes [129], authors used MR imaging to study the fetal development between 22 to 38 GW and 14 to 38 GW, respectively. Studying a large series of normal fetal brains, Garel et al. [125] found in-utero MR observations followed the same order of events as the post-mortem neuroanatomic results reported in [126], however, a temporal lag was found. They concluded this was a consequence of the MR slice thickness being greater than the slices used by the neuropathologists. A similar lag in healthy fetuses was found by Levine and Barnes [129], shown in Figure 3.5, which was further increased in fetuses with central nervous system (CNS) abnormalities [129], indicating that fissures can be used as a landmark for healthy development. In more recent work [130, 134], the sulcal timetable has been quantitatively measured by modelling surface curvature and estimating sulci emergence by significant curvature change. These studies were found to be consistent with the previous MRI observation reports [125]

for some fissures, however, large discrepancies (over 2 weeks) were also found. In the majority of the cases, these techniques detected folds earlier than previously reported in MRI, fitting better with the postmortem findings, which may imply that curve fitting is more sensitive to subtle folds than observational methods.

Ultrasound Fissure emergence timetables have also been found from ultrasound scans [22, 131, 132]. Initial work reported a consistent lag between fissure timings between ultrasound and the postmortem studies, with the timetable found to be more similar to that reported from MRI [131]. Cohen-Sacher et al. [133] described the longitudinal fissure emergence from 21 pregnancies that were scanned every 2 weeks, from 18 weeks to term. This study found small differences (usually less than 1 GW) compared to the MRI studies [125], with the majority of the structures emerging earlier than reported in MRI. Moreover, this study identified some structures 4-6 GW earlier than the previous ultrasound reports [22, 132]. Figure 3.5 compares the emergence timetables between postmortem [126], MRI [129], and ultrasound [133] studies. The postmortem identifies the SFS, PreCS, IFS, PoCS, CS, POF and CaS, earlier than either of the in-utero scanning methods. The OS is identified at the same time across all three modalities. Both the IFS and ITS are reported earlier in ultrasound than in MRI. Although the reported weeks have been found to vary across the modalities, the order of events is found to be consistent across the reports.

3.1.3 The Sylvian Fissure

One of the most prominent features of the human brain is the SF, a major sulcus observed on both hemispheres, separating the frontal and parietal lobe from the temporal lobe [126, 128]. The SF goes through dynamic changes from when it is first seen at 14 GW until term, as shown in pink in Figure 3.1, and therefore, has the potential to be used as a landmark to characterise development. At 14 GW the sulci is a smooth depression [135], which becomes more defined and raised between 18 and 22 weeks GA [132]. As the boundaries rise, they enclose the

triangularly shaped *insula*, a brain region associated with consciousness [136]. As the surrounding lobes grow, the insula becomes slowly engulfed, a process known as *operculization* [132], which begins at the eighteenth GW but full closure is not seen until the postnatal age of 2 years [135, 136]. Operculization of the SF follows a zip-like pattern starting at the posterior end of the brain, bringing the parietal lobe and temporal lobe together, and continuing in the anterior direction [137] (see Figure 3.3 for terminology). Between 28 and 35 GW, the majority of the insula is covered, which itself becomes convoluted, with the remaining SF appearing as a deep horizontal groove. Delays or defects in SF operculization can cause serious impairments to development as it encompasses areas important for language and speech, auditory function, sensory and motor function and consciousness [137].

3.1.4 Abnormal Gyrification

Significant changes in cortical folding have commonly been associated with severe intellectual disabilities [138], neurological disorders [139], psychosis [140] and birth defects [114, 115]. Extreme variations from healthy cortical folding can broadly be categorised into two types of brain disorders: *lissencephaly* and *polymicrogyria*. Lissencephaly is a set of disorders, which are characterised by a smooth brain surface (e.g. missing folds). This is most commonly caused by a genetic mutation resulting in stunted neuron migration, seen from the twelfth GW [141]. Polymicrogyria is a condition characterised by an abundance of small cortical folds. Many different forms of polymicrogyria have been identified with a broad scale of severity [142].

Polymicrogyria and lissencephaly are severe, and observable gyrification abnormalities, however, more subtle changes have also been detected in at-risk and diseased pregnancies. In what follows, I describe two case studies of regional gyrification abnormalities, which will be referred back to throughout this Chapter.

PreTerm Cortical folding abnormalities have commonly been associated with adults born preterm, although the connection between them is still unknown [113, 115, 143]. Preterm babies are those that are born, alive, before the completion

of 37 weeks of pregnancy, and is the single biggest cause of mortality in infants aged < 5 years [144–146]. An estimate of 11.1% of worldwide live births is preterm, with the rates increasing yearly [146]. Preterm birth can also result in long-term complications, such as cerebral palsy, sight and hearing problems, impaired learning and chronic health issues [146].

The majority of preterm births happen spontaneously with common causes such as multiple pregnancies, infections, diabetes or high blood pressure. However, often no cause is identified. Predictive tests of preterm birth are incredibly important due to the high impact prematurity has on maternal and neonatal health, as well as the economic cost [145]. Further, early detection is key so any possible medical intervention can be offered. However, the two main current detection methods can only be used for imminent birth (within a week), and therefore are often too late for medical intervention or maternal reassurance [145, 147–149]. Therefore, there is a need for a non-invasive, early detection method that can be integrated into routine care.

Preterm newborns have been found to have regional advanced gyrification (e.g. more heavily folded) compared to in-utero age-matched fetuses from as early as the 21 GW [115]. It is now widely accepted that premature birth leads to complex regional aberrant gyrification across the lifetime, however, it is unknown if these changes can be seen in-utero and therefore used as a potential biomarker. Using functional MRI (fMRI), a recent study found variation in functional connectivity between fetuses which go on to be preterm and term newborns [150] and concluded this could be used as an early detection biomarker for preterm birth. However, fMRI is expensive and not commonly used in routine antenatal care. Therefore, investigating whether fetuses exhibit abnormal cortical folding prior to preterm birth from routine pregnancy scans, has the potential to be used as an early detection, screening method.

Heart Disease Abnormal cortical folding has been linked to fetuses and infants with congenital heart disease [114, 151, 152]. Congenital heart disease is an

umbrella term describing a range of heart defects present from birth. From in-utero MRI scans, local folding delays were found between fetuses with hypoplastic left heart syndrome (HLHS) from as early as 25 GW across the developing surface, with some sulci found to have greater lag times than others, compared to the healthy population [114]. In this condition, the left side of the heart is severely underdeveloped [153, 154], affecting the blood flow through the heart, body and brain. The condition is rare, occurring in about 2 in every 10000 live births [154]. However, if left untreated it is fatal [153, 154].

The process of cortical folding is energy-intensive, requiring plenty of oxygen [109, 112]. It is therefore hypothesised, that conditions such as HLHS can delay gyrification, as it reduces the amount of oxygen reaching the brain. Thus, monitoring cortical folding has the potential to detect such conditions.

3.2 Clinical Assessment

In routine care, fetal development is monitored across gestation, including assessment of the brain. This section introduces the current clinical guidelines followed during routine ultrasound scans and introduces why ultrasound is the modality of choice in prenatal care. Further, I review some additional screening methods that have been proposed to specifically assess the cortical plate.

3.2.1 Routine Care

Throughout gestation, pregnant women are offered a series of screening tests and ultrasound scans to monitor fetal development. The ultrasound scans are routinely performed transabdominally and are typically scheduled between 8 to 14 GW for pregnancy dating; 18 to 21 GW for anomaly screening; and at 36 GW for growth assessment. The International Society of Ultrasound in Obstetrics and Gynecology (ISUOG) has developed best-practice guidelines for routine scans to ensure consistent and reliable assessments [133, 155–157]. The ISUOG guidelines include an optimised approach to evaluate fetal CNS, known as the *basic examination* [156]. The basic examination is usually performed during the second, also referred to as the *anomaly*,

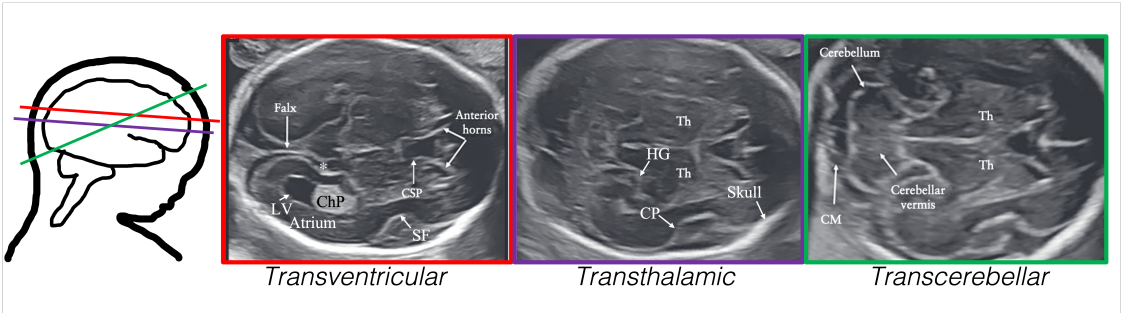


Figure 3.6: The three main fetal brain planes obtained during the routine *basic examination*. The structures highlighted are thalami (Th), hippocampal gyri (HG), lateral ventricles (LV), cavum septi pellucidid (CSP), choroid plexus (ChP) Sylvian Fissure (SF), Cortical plate (CP), cisterna magna (CM). Figure adapted from [159]

ultrasound scan, however, guidelines are in place if the examination takes place earlier or later [133, 157]. If there is an indication of increased risk of CNS anomalies, this examination may be performed using a three-dimensional (3D) ultrasound probe and/or the sonographers might refer the subject to have a fetal MRI scan [158].

Basic Examination The basic examination requires the sonographer to perform both a qualitative and quantitative evaluation of the fetal brain. During the qualitative examination, the sonographers visually assess the shape, texture and presence of key brain structures. This information is assessed across three main axial planes that allow for clear visualisation of the structures of interest, as shown in Figure 3.6.

The first view is the transventricular plane where the sonographers check the appearance of the lateral ventricles (LV), cavum septi pellucidid (CSP) and atrium. In this view, sonographers also check for dilation of the ventricles, also known as *ventriculomegaly*, which is identified in 1 in every 100 fetuses at 20 weeks. This condition can be benign, however, it can cause severe neurodevelopmental delays [160]. It is recommended for the atrium diameter to be measured as a biomarker for ventriculomegaly and is considered normal if it is less than 10mm in the second and third trimesters [156].

The second plane identified is the transthalamic plane, which is inferior to the transventricular plane, where the sonographer checks that the thalami (Th)

and hippocampal gyri (HG) are present. The biparietal diameter (the widest distance across the head) and head circumference are usually measured from this plane, as it is found to produce more reproducible results than measures from the transventricular plane [161]

The final plane obtained is the transcerebellar plane. In this view, the sonographer analyses the cerebellum and cisterna magna (CM) shape. The CM is fluid-filled space posterior to the cerebellum [156]. After the twelfth GW, the CM should be a stable depth between $2 - 10$mm [156]. An enlarged CM has been associated with mild developmental and significant walking delays in children [162].

It should be noted that throughout this basic examination sonographers are not instructed to qualitatively or quantitatively assess the cortical plate, with the exception of detecting the presence of the HG. Considering the importance of the cortex in healthy development, this is a missed opportunity, as it could provide crucial insight into fetal health or associated risks. Patients with increased risk of CNS anomalies are referred to neurosonographic assessment, where the SF and IHF are observed [156, 163, 164]. Severe cortex diseases, such as lissencephaly and polymicrogyria (introduced in Section 3.1.4) can be identified in these screenings, however, more subtle changes, like that described for HPLS, are more challenging to identify by eye.

In summary, the expertise and knowledge required to perform the basic examination are vast. Moreover, during this scan, sonographers are also required to assess the face, neck, chest, heart, abdomen, skeletal, placenta, and umbilical cord [157]. Therefore, new measures or assessments of development should aim to be automatic to decrease the sonographer's workload.

3.2.2 Ultrasound Imaging

Ultrasound is the modality of choice for prenatal care, largely due to its relative low cost, availability worldwide, short acquisition times (in the scale of seconds) and real-time visualisation [156]. Typically, ultrasound is performed in 2D, imaging a fan-shaped area, however, 3D imaging is now becoming more commonplace [165].

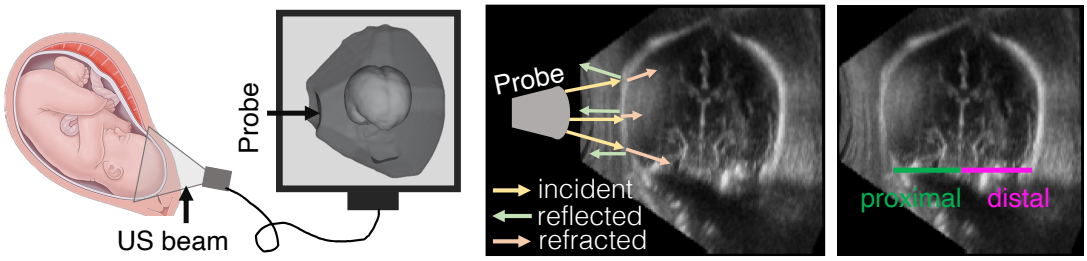


Figure 3.7: Using ultrasound to image the fetal brain. The first panel shows a schematic of what the ultrasound beam is looking at during a scan. The second panel shows a simplified representation of how the incident ultrasound beams are reflected or refracted by structure within the brain. The third panel shows large shadows obstructing the proximal hemisphere, due to the large acoustic impedance of the skull.

In short, ultrasound imaging works by emitting sound waves from multiple angles and measuring the magnitude of the partially reflected signals to reconstruct an image, as shown in Figure 3.7. However, this technique can cause acoustic shadows, shown as dark areas in the image. Shadows are created when the ultrasound waves hit bone or air, which reflect or absorb the waves [166]. For example, as the fetal skull undergoes *calcification*, the build-up of calcium in the bone, the waves are almost entirely reflected, occluding the other tissues behind it, resulting in shadows, as shown in Figure 3.7. These shadows can cause large artefacts across the volume, often occluding the hemisphere closest to the probe (referred to as the *proximal* hemisphere). This has limited sonographic studies to only analyse the *distal* hemisphere [165]. Moreover, as calcification increases with gestational age, the shadows become more prominent and thus, the cortical plate becomes more challenging to study with age.

3.2.3 Fissure Screening

In current prenatal care, the cortical plate is not routinely assessed, despite its potential to detect neurological abnormalities. To overcome this, a series of methods have been proposed to monitor the growth of individual fissures from routinely obtained ultrasound scans. In such methods, a property (such as depth or shape) of a fissure is measured and compared to a reference chart of healthy fissure

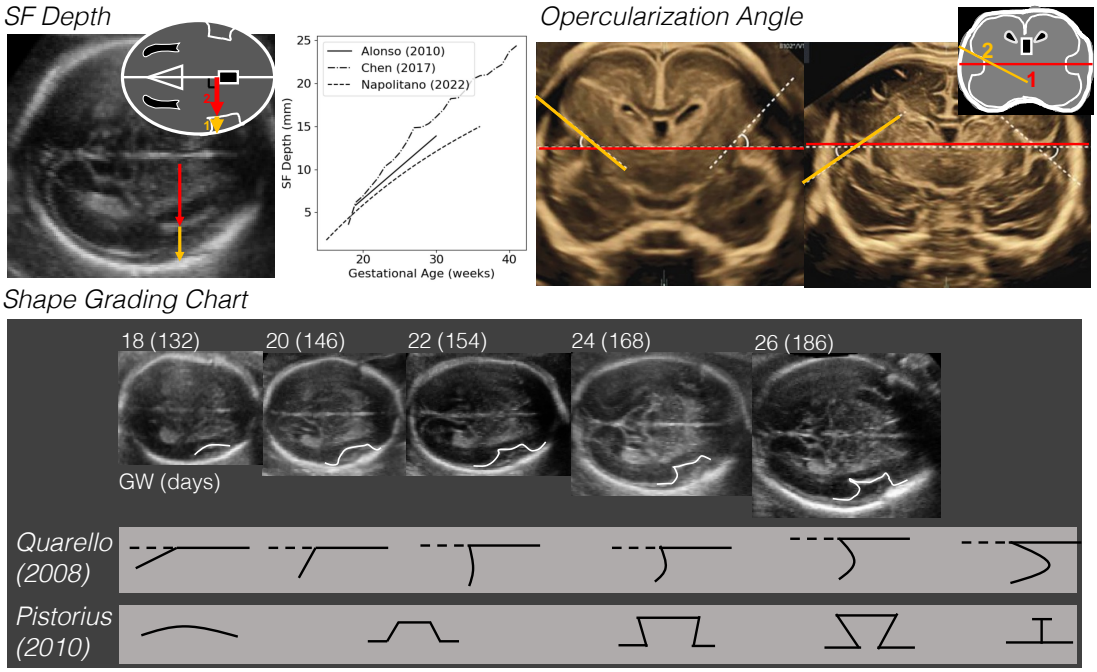


Figure 3.8: An overview of the different methods used to assess SF growth. In the top row, the protocol for measuring depth (LHS) with the projections over age [169–172], and opercularization angle (RHS) adapted from [173] is shown. In the bottom row, the two visual grading charts proposed in Pistorius et al. [168] and Quarello et al. [137] are outlined.

growth, to see if the fetus is developing at the expected rate. Although these methods are currently not used in standard care, they are commonly used in clinical trials and clinical research [167].

Fissure grading charts have been designed for a range of fissures [168, 169], however, this section focuses on those designed for the SF. As previously described, the SF undergoes a rapid transformation after its first emergence at 14 GW, with the structure deepened before undergoing opercularization. Moreover, if the fetus is at risk of CNS abnormalities, the SF is assessed. SF grading charts have been proposed to assess depth, opercularization angle, and shape, as outlined below.

3.2.3.1 Depth

As the SF undergoes opercularization, it begins to deepen. Thus, initial screening methods are proposed to assess development by measuring SF depth across gestation. For consistency, a standard protocol has been established to measure SF depth

from ultrasound images. First, the standard transventricular axial plane is found, as shown in Figure 3.6 [156]. The insular depth, which is defined as the distance between the midline and the inside of the cortical plate, perpendicular to the midline, is then delineated, as shown with the red arrow in Figure 3.8. Finally, the SF depth is found by extending the insular depth line until it reaches the inner parietal bone, whilst remaining perpendicular to the midline indicated with the yellow arrow.

A range of studies have quantified SF depth from ultrasound scans and produced expected growth trajectories across gestation [169–172]. In the majority of these studies, the SF depth was measured using the protocol described above from 2D ultrasound probes [169, 170, 172], with the exception of one study that extracted the transventricular plane from 3D ultrasound scans [171]. Across these studies, the SF depth was found to deepen with age, which followed an almost linear projection during the first and second trimesters, as shown in Figure 3.8. However, despite following the same labelling protocol, the growth projections were found to vary between the studies. Each work analysed a large cohort ($N = 227$ [171],$N = 746$ [172], $N = 442$ [169],$N = 180$ [170], where N is the number of pregnancies included, which were often each studied longitudinally), so it is unlikely that these growth differences are due to normal anatomical variability across the population. The discrepancy is more likely due to the large intra- and inter-user variability found when performing such measurements [174].

3.2.3.2 Opercularization Angle

As the SF develops, it begins to fold over itself, making the cortical plate angles more acute. To assess this change, Poon et al. [173] proposed an SF grading chart based on opercularization angle, measured from transvaginal ultrasound scans in the coronal plane. To assess this angle, first, a horizontal line perpendicular to the midline that crosses the insula and below the ventricles is drawn. Next, a second line along the distal edge of SF is drawn and the angle between the two is measured. In this, positive angles show obtuse shapes, linked to early gestation, whereas, negative angles are acute, which were found after 25 GWs. The SF angle

reference chart was produced by measuring the opercularization angle, using the protocol described above, for a group of $N = 422$ healthy fetuses between 18 and 30 GWs. The angle was then reported over gestational age, which was found to have a significant negative polynomial association [173].

The authors concluded that assessing the SF angle from transvaginal scans could serve as an additional screen tool. However, as transvaginal ultrasound scans are not currently used in routine care, it would be more impactful if such methods were described from typically used transabdominal scans.

3.2.3.3 Shape

Performing manual measurements, such as depth or operculization angle, is found to be time-consuming as well as user dependent, which has led to the development of shape grading charts. In this, a set of grades which describe the key stages of fissure development are defined, as shown in Figure 3.8.

SF operculization was assessed and scored by Quarello et al. [137], who proposed a 10-step scoring system based on the posterior angle of the SF. However, the distinction between grades was found to be very subjective and a large age overlap was found between grades. Pistorius et al. [168] simplified this grading, by only including five grades to describe the SF growth between 20 to 40 GWs. In this, each grade resembles a stage of operculization, as shown in Figure 3.8, and the expected range of gestational ages for each grade was found by studying a cohort of 28 patients from 20 to 40 weeks of gestation.

Despite promising inter- and intra-observer correlation coefficients reported in these works, a follow-up investigation found the results were difficult to reproduce when using the same data [174]. This is likely due to the subjective nature of assigning grades, and parietal due to the requirement to find the correct plane in order to accurately assess the structure, a limitation of all the grading methods presented here.

3.2.3.4 Limitations

The methods proposed here have been designed to monitor brain development, with the aim of being integrated into clinical practice. They utilise commonly used ultrasound scans, to ease such integration, and cover a wide period of fetal development. However, they are not without their limitations. Firstly, analysing fissures individually can be extremely time-consuming and subjective [174]. Moreover, these methods are specific to a single fissure, therefore new structures require new protocols and reference charts. Further, these methods only assess the 2D nature of folds, losing valuable 3D information.

3.3 Assessing the Developing Cortex

Across the fetal period, the cortical plate undergoes complex morphological changes. As the fetal head grows, the cortical plate volume rapidly increases and it begins to fold, allowing more neurones to fit within the developing grey matter whilst contained within the skull [175]. This folding leads to a larger surface area, changes in curvature and increased sulcal depth [112]. Folding also allows regions of the brain to become closer together, thus increasing connectivity [109]. Further, as the neurons migrate to the developing cortical plate, it also begins to thicken [176]. Therefore, the changes across gestation cannot be summarised using a single metric, but instead to truly characterise the development a range of morphological measures is required.

With recent advances in 3D imaging and computing, surface analysis of the cortical plate in-utero is beginning to be performed. In order to do this, a segmentation of the developing cortical plate must first be acquired. In this section, I first introduce methods used to segment the developing cortical plate (Section 3.3.1)[43, 112, 115, 177–180], before discussing the techniques designed to characterise the complex surface across gestation (Section 3.3.2).

3.3.1 Segmenting the Cortical Plate

Fetal cortical plate extraction is an extremely difficult task [41, 181]. This is partially due to the vast variations in cortical thickness, texture, shape and contrast seen across gestation, due to the migrating neurons. Thus, given that the cortex at beginning of the second trimester looks vastly different compared to the third trimester, a single delineation technique is particularly challenging.

In this section, I discuss the previous work that has delineated the developing cortical plate, with an overview shown in Table 3.2.

3.3.1.1 Adapted Toolboxes

Delineating the cortex in adults is routinely performed in MRI studies, with numerous toolboxes available for automatic segmentation such as FreeSurfer and BrainVisa [185, 191]. However, due to the vast morphological and contrast differences between the cortex of a fetus and an adult, these stand-alone tools are incompatible with fetal and newborn brain segmentation [192].

Recent work has adapted such adult toolboxes and applied them to newborns and fetuses [115, 184, 193]. For example, Dubois et al. [184] segmented the inner cortical boundary from MRI scans of preterm newborns by adapting the Cortical Folding pipeline in BrainVisa [185]. Their adaptation included bias correction and required manual tuning of parameters for each subject, and produced a segmentation of the white matter. To extract the inner cortical surface from the white matter, a smooth triangle-based mesh was applied. This semi-automatic processing pipeline was then applied to fetuses in Lefèvre et al. [115], however, due to poor tissue contrast manual corrections were required. In addition to the manual corrections and parameter tuning, this method has only been shown for the late second and third trimesters. Lefèvre et al. [115] studied 14 fetuses and 27 preterm newborns with a mean age of 29.6 ± 3.5 and 31.4 ± 2.4 GW, respectively. By the beginning of the third trimester (28 GW), the brain surface has already undergone major gyrification events and therefore, begins to resemble the adult brain, which may be beneficial for the adapted toolboxes approach. However, such methods may

Table 3.2: A summary of the methods used to perform cortical plate segmentation in 2D and 3D reported in the literature. †- deep learning with human in loop approach.

Authors	Age Range (GW)	No. Subjects	2D/3D	Cohort	Modality	Automated	Method	Topological Constraint
Caldairou et al. (2011) [181]	27 - 34	6	2D	Fetuses	MRI	Semi	Clustering	✓
Bach Cuadra et al. (2009) [182]	29 - 32	4	2D	Fetuses	MRI	Semi	Bayesian Segmentation	✗
Habas et al. (2010) [183]	20 - 24	20	3D	Fetuses	MRI	Semi	Atlas-based	✗
Dubois et al. (2008) [184]	26 - 36	35	3D	Preterm newborns	MRI	Semi	Adapted toolbox [185]	✗
Serag et al. (2012) [186]	23 - 37	80	3D	Fetuses and neonates	MRI	Auto	Atlas-based	✗
Gholipour et al. (2017) [187]	19 - 39	81	3D	Fetuses	MRI	Auto	Atlas-based	✗
Clouchoux et al. (2012) [112]	25 - 35	12	3D	Fetuses	MRI	Auto	Atlas-based	✗
Rajagopalan et al. (2011) [188]	20 - 28	40	3D	Fetuses	MRI	Auto	Atlas-based	✗
Fetit et al. (2020) [180]	21 - 38	92	3D	Fetuses	MRI	Semi	Deep learning†	✗
Khalili et al. (2019) [179]	22 – 34	12	2D	Fetuses	MRI	Semi	Deep learning	✗
Dou et al. (2020) [178]	16 - 39	23	3D	Fetuses	MRI	Semi	Deep learning	✗
Dumast et al. (2022) [189]	20 - 34	27	3D	Fetuses	MRI	Auto	Deep learning	✓
Wyburd et al. (2020) [190]	18 - 22	36	3D	Fetuses	US	Auto	Deep learning	✗

struggle with younger ages, where the cortex looks vastly different. Moreover, these toolboxes are only compatible with images from the same modality, which to date have been exclusively MRI, and therefore, would not be applicable to ultrasound volumes: the modality of choice in prenatal care.

3.3.1.2 Atlas-based Segmentation

In recent years, numerous longitudinal fetal brain atlases have been produced, with different brain tissues labelled [41–43, 186, 194]. Registration of new fetal brain volumes to an existing labelled atlas allows for label propagation from the atlas back to an individual's brain, minimising the time-consuming task of manual labelling, as previously discussed in Section 2.1.2.

The use of atlas-based fetal tissue segmentation was first performed by Habas et al. [41]. Using 3D MRI volumes of fetal brains between 20 to 24 GW, the authors constructed an average shape and intensity reference atlas, which was then manually segmented into key regions by expert observers. Using prior knowledge from the atlas about the intensities and position of different labelled classes, the authors proposed an Expectation-Maximization (EM) framework to label unseen datasets. EM algorithms iteratively update labels for each brain structure to find the most likely distribution. Unseen volumes were non-rigidly registered to the labelled atlas and the class probabilities for each voxel were initialised from the values in the atlas. This method was evaluated on unseen manually labelled volumes and found to segment the cortical plate with a high Dice coefficient of 0.79 ± 0.02.

Serag et al. [186] and Gholipour et al. [187] constructed similar fetal MR atlases but over wider age ranges: 23 to 37 and 19 to 39 GW, respectively. The performance of cortical plate label propagation was evaluated for both atlases and found to have an average Dice of 0.84 ± 0.06 in [186], and 0.9 for fetuses younger than 34 weeks and 0.84 for fetuses older than 34 weeks in [187]. Since their construction, these atlases have been used to label a plethora of MRI volumes used throughout fetal brain development studies, saving time and expertise in fetal brain labelling [112, 114, 178].

Atlas-based label propagation obtains good-quality segmentations whilst minimising the time required for manual annotation. However, manual annotation of the structure of interest on the atlas is still required and the quality of each subsequent label is limited by the accuracy of the initial annotation. Moreover, this method is often only semi-automated as each volume must first be registered to a reference space, which can be time-consuming and prone to errors, resulting in a need for manual correction [112]. A labelled atlas is also required, which may not be available in all imaging modalities and across the gestational age range of interest. All the atlases described above have been generated using MRI volumes, before being applied to new unlabelled MRI volumes.

3.3.1.3 Deep Learning

Over the last few years, CNNs have been commonly used to segment brain structures from MRI and ultrasound scans [190, 195–200]. One such example was performed by Moeskops et al. [197], where the authors successfully segmented key brain structures in preterm neonates, children and adult brains using the same architecture. However, until recently this has not been performed for the cortical plate in fetuses.

Khalili et al. [179], were amongst the first to demonstrate CNN's ability to segment subcranial structures from in-utero MR images [179]. In this work, the authors trained a U-Net style architecture to segment 7 tissue types, including the cortical plate, from brain-extracted 2D MR slices. The training data consisted of $N = 126$ MR slices obtained using a T2-weighted scan from 12 fetuses between 22 and 34 GW. In order to train the network, the ground truths for each tissue type were manually annotated in the coronal plane by a medical student, which took 40 hours per volume. The segmentation performance for each structure was then compared to previous atlas-based label propagation methods [183, 186] and found similar performance for cortical plate segmentation, achieving a Dice score of 0.84 compared to 0.82 and 0.84 reported by Habas et al. [41] and Serag et al. [186], respectively. However, it should be noted these evaluations were calculated on different test sets, and Habas et al. [41] and Serag et al. [186] used 3D volumes

compared to 2D slices in this work. Subsequent methods, networks were extended to perform cortical plate segmentation in 3D [178, 180].

Fetit et al. [180] performed 3D cortical plate segmentation from MRI scans between 22 to 34 GW using the DeepMedic architecture with a human-in-loop approach. DeepMedic was originally proposed for adult brain lesions [196] but has since been adapted for a variety of medical segmentation tasks [180, 201, 202].

In this work, two DeepMedic networks were used, each comprising of three parallel pathways of varying resolutions. The first network was trained on the 9-class segmentation problem, before an expert annotator visually inspected the predicted cortical plate label and manually corrected 3 slices in each volume, one in the axial, sagittal, and coronal plane. The second network then only performed a single class segmentation: cortex or not cortex, which was fine-tuned with the corrected scans, in addition to a new set of fetal brain volumes, independently collected from a different strength MRI scanner. The final network was then evaluated on an unseen test set and achieved a Dice score of 0.82 ± 0.07 for cortical plate segmentation, comparable performance to the 2D Dice reported by Khalili et al. [179].

This study achieved high-quality segmentation of the cortical plate over a vast range of gestational ages, which corresponds to dynamic morphological changes in the cortex. However, during training, manual annotations were required, which are time-consuming and user-dependant. Although this pipeline was considerably more complicated than the Atlas-based Label Propagation and Adapted toolboxes method, it resulted in a trained model which can be used to generate 3D labels on unseen data without the need for manual corrections.

In Dou et al. [178], authors studied $N = 57$ 3D fetal brain MR scans between 16 and 39 GW, a much wider age window than previously studied [179, 180]. To minimise the time spent performing manual annotations, authors generated ground truths through atlas-based label propagation [187]. Similarly to Khalili et al. [179], the authors used a U-Net style architecture, but with an additional attention module at each decoder layer. In a similar way to the network used in Fetit et al. [180], the attention modules use information across a range of resolutions, to try and refine

the feature maps at each resolution, highlighting the most relevant information. The training loss was a weighed loss between the U-Net outputs, Attention refinement outputs, and the final output patch. Due to memory constraints, the input of the network was a 64^3 voxel patch instead of the whole brain volume.

The performance of this network was compared to 3 state-of-the-art semantic segmentation networks and was found to outperform them on Dice, HD and ASSD. A blinded expert then compared the delineation of the cortical plate between the proposed network and the atlas-based label propagation reported in Gholipour et al. [187]. Based on majority voting, the CNN architecture was found to outperform the atlas-based label propagation method. This paper demonstrated how a single network was able to segment the cortical plate to a high standard (an average Dice of 0.87 ± 0.06), without a human-in-loop approach as previously demonstrated and over a wide range of gestational ages. Unlike the previous methods, no manual segmentations were required and, once the model was trained, a cortical plate segmentation was obtained in less than 2 minutes, which is an order of magnitude faster than most atlas-based approaches [43, 178].

3.3.1.4 Topology Correction

Since cortex segmentation is a prerequisite for 3D cortical measures, it is important that the surface is intact. This is especially key if surfaces need extracting, as discussed later in Section 3.3.2.1. However, often cortical segmentation methods result in topological errors, more specifically holes across the surface, which can make further analysis extremely challenging [189, 203]. Cortical plate segmentations are particularly vulnerable to topology mistakes, as the structure is thin and convoluted, which can result in disconnects and holes forming within the cortical folds. In adult grey matter segmentation methods, topological constraints or post-hoc corrections are commonly applied to prevent such mistakes [204–206]. More recently, such methods have begun to be integrated into fetal cortex segmentation [181, 189, 203].

Caldairou et al. [181] employed a topological-based clustering method that incorporates prior knowledge of the fetal brain to prevent topological violations.

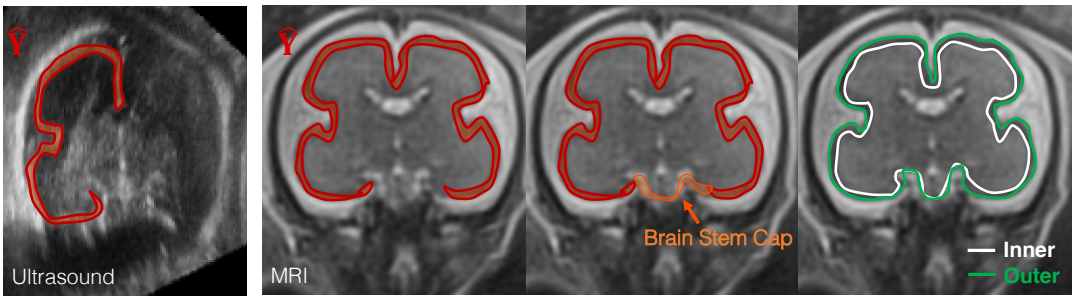


Figure 3.9: The cortical plate volume ($\hat{\mathbf{Y}}$) shown in ultrasound and MRI. This example shows how the inner and outer cortical boundary can be extracted from MRI volumes by capping the brain stem, to create a closed surface. The images are both at 24 GW and the MRI obtained from [120]

However, this method was only performed on six fetuses and topology performance was not reported. Dumast et al. [189] integrated a topological loss function (previously discussed in Section 2.3.1) [26] into a U-Net style architecture, to try and avoid disjoints within the cortex segmentation. The method was evaluated on fetal MRI scans collected from a range of studies. The networks achieved promising segmentation performance (Dice of 0.85 ± 0.01) and were found to reduce the number of topological errors, however, some holes were still present. The network was used to segment the cortex of fetuses between 20 to 38 GW, however, the segmentation and topology performance were both found to decrease across gestation.

3.3.2 Cortical Analysis

After delineating the cortical plate, an in-depth analysis of the structure can be performed using a range of metrics. From the segmentation methods, often a binary volume $\hat{\mathbf{Y}} \in \mathbb{R}^{H \times W}$ is extracted, where voxels containing the cortical plate are labelled. However, for some metrics, a surface $\mathcal{S}$ is required. In this section, I first describe how surfaces can be extracted before discussing metrics that have been used to analyse the developing cortex from both the $\hat{\mathbf{Y}}$ and $\mathcal{S}$.

3.3.2.1 Surface Mesh Generation

To generate a surface from the segmented cortical plate ($\hat{\mathbf{Y}}$), surface mesh algorithms, such as marching cubes, are often implemented [112, 115, 177, 207]. Applying such

algorithms to a topologically-correct cortical plate segmentation would produce a single mesh, encompassing the whole structure, with the equivalent topology as a solid sphere. However, often the analysis methods require the inner and outer cortical plate boundaries to be extracted separately, producing two surface meshes. To achieve this, methods have been designed to cap the brain stem from MRI images, as shown in Figure 3.9, so that the cortical plate now has the topology of a hallow sphere [191, 208]. From this representation, both the inner and outer cortical boundary can be extracted [112, 175]. If the topology is incorrect, closed surfaces can not be created preventing the extraction of the inner and outer boundaries.

3.3.2.2 Volume and Surface Area

Across gestation one of the most striking changes in the fetal brain is size, as shown for the second trimester in Figure 3.10. Cortical volume has commonly been measured across gestation and is found to follow a quadratic trend [114, 175, 209, 210]. Volume can be measured directly from $\hat{\mathbf{Y}}$, by summing the number of voxels labelled as the cortical plate. Cortical folding allows the cortical surface area to increase, whilst contained within the skull. Acquiring surface area from a $\mathcal{S}$ is a relatively simple calculation, obtained by summing all the area of the triangles within the mesh [112, 115, 176]. Surface area is usually measured from the inner cortical boundary, rather than the outer boundary, as it undergoes the larger transformation across gestation [112].

Fetuses with HLHS have been found to have decreased cortical volume and surface area compared to age-matched controls [114, 211]. For both measures, the discrepancy between the two groups was found to be larger in the third trimester than in the second. Preterm newborns have been found to have increased surface area from 28 GW compared to in-utero age-matched healthy controls, however, no changes in volume were identified[115].

3.3.2.3 Cortex Thickness

As the neurones migrate, the cortical plate begins to thicken, thus measuring cortical thickness provides an insight into the complex cortical microstructures

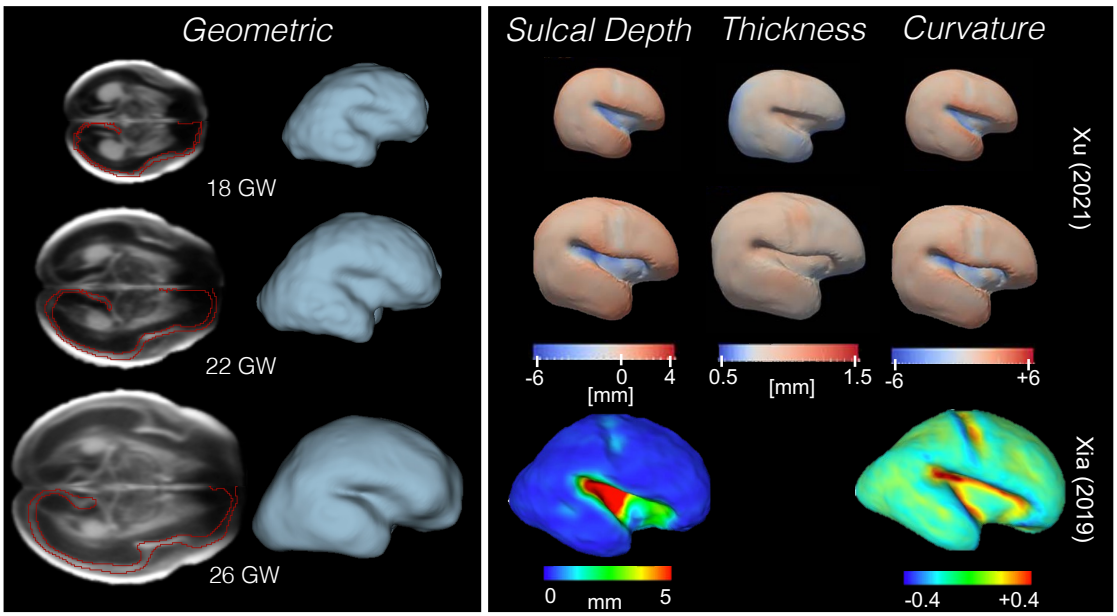


Figure 3.10: The changes in 3D cortical growth across the second trimester. The first panel shows volumetric growth with the cortical plate outlined in red. The second panel shows changes in sulcal depth, thickness and curvature, across gestation. The 3D renderings for sulcal depth, thickness and curvature were taken from Xu et al. [212] and Xia et al. [176]. There are no thickness measures at 26GW, and to the best of our knowledge, no method to date has assessed thickness in 3D past 22GW, as shown in Table 3.3.

[176]. Traditionally, cortical thickness in adults has been measured using Laplace's equation [213, 214]. In this, an opposing charge is assigned to the inner and outer cortical surface, to generate a set of equipotential surfaces, using Laplace's equation. The perpendicular vector from each location on the inner surface is then found, and the distance to the next equipotential surface is measured, this is then repeated for all intermediate surfaces. This method has been shown to be robust for highly convoluted surfaces, such as the adult brain. However, this method also requires topologically correct inner and outer boundaries.

The Laplace equation has also been used to assess fetal cortical thickness [212]. From these studies, cortical thickness has been found to vary across the structure, with the sensorimotor region (between the frontal and parietal lobes, as described in Figure 3.3) found to be the thickest region between 14 to 22 GW, before it begins to even out, as shown in Figure 3.10 [212]. Therefore, thickness is often presented across the cortex, rather than a global measure.

3.3.2.4 Sulcal Depth

As the fetal brain folds, the sulci deepen. As discussed in Section 3.2.3, fissure depth has been proposed to assess development across gestation. However, such methods only measure a single fissure, despite the sulci deepening at different rates across the surface. Therefore, using the delineated cortical plate, sulcal depth has begun to be measured across the surface [176, 212, 215, 216]. Cortical depth can be defined in two different ways: the distance between the skull and inner cortical plate boundary and between the cerebral convex hull (the surface along the gyri but without dipping into the sulci) [176] and the inner cortical boundary. Simply, these methods measure the minimum depth between either the skull or cerebral convex hull and the inner cortical plate boundary. Figure 3.10 shows how sulcal depth varies across the cortical plate, at three gestational time points. The two studies use the convex hull method, and therefore the majority of the surface has a depth of 0. It should be noted that in Xu et al. [212] and Xia et al. [176], depth is reported with different conventions (e.g. + and -).

Fetuses with HLHS have been found to have statistically meaningful regional differences in cortical depth when compared to healthy controls [114]. Further showing why regional depth, rather than a global value, should be studied.

3.3.2.5 Gyrification Index

To study the extent of cortical folding globally, the gyrification index (GI) can be measured, which has been frequently used in adult studies [107, 177]. GI was first defined in 2D as the ratio between the contour length around the cortex and the overall length of the outer contour of a 2D slice [107]. The higher the GI value the greater the folding. Hu et al. [210] performed a manual outlining of the fetal cortex from MR slices and found the GI index increases linearly with gestational age between 21 to 37 weeks [210].

The two-dimensional definition of gyrification does not take into account the inherent three-dimensional nature of the cortical surface, which was addressed by Luders et al. [217], who defined the three-dimensional GI as the ratio between

cortical surface area and the area of the cortical surface convex hull [177]. Clouchoux et al. [112] measured the three-dimensional GI, using the surface area techniques described above and found in 3D the relationship between gestational age and GI became non-linear after week 28 GA, conflicting with the 2D relationship noted in Hu et al. [210]. Fetuses born preterm have been found to have increased GI from 25 GW compared to healthy controls [115]. From 30 GW, fetuses with HLHS are found to have decreased GI compared to healthy controls, however, between 25 and 30 GW the GI is consistent between the two groups.

GI provides a good measure of global folding, however, as with thickness and sulcal depth, folding is found to vary across the cortical surface. To assess local folds, the curvature is often measured. Curvature is simply defined as the amount by which a line deviates from being straight or, in 3D, the amount a surface deviates from being a plane [218]. In fetal analysis usually, the mean curvature (H) is assessed, as shown in Figure 3.10 [219]. At each vertex ($\mathbf{v}$) within $\mathcal{S}$, local curvature can be defined by two principal curvatures ($k_{\min}, k_{\max}$), which measure how the surface bends in different directions. In practice, $k_{\min}, k_{\max}$ are obtained by locally fitting polynomials to the $\mathcal{S}$ around $\mathbf{v}$, which can be described by a matrix $\mathbf{W}$ and $k_{\min}, k_{\max}$ correspond to the two largest eigenvalues of $\mathbf{W}$ [112]. Mean curvature at $\mathbf{v}$ is then defined as $H = \frac{k_{\min}+k_{\max}}{2}$.

3.3.2.6 Analysis Limitations

Table 3.3 displays an overview of the studies used to quantify the complex morphological development of the cortical plate from extracted 3D volumes and surfaces. To date, and to the best of my knowledge, all the analysis has been performed from MRI scans, with the majority performed in-utero and just two from postmortem anatomical data. Therefore, the majority of the methods described above, have been designed specifically for MRI scans. For example, such methods often require the inner cortical boundary to be delineated separately. It is often possible to separate the inner and outer boundary from MRI, as the cortex is segmented across both hemispheres and by capping the region around the brain stem, it forms a closed

Table 3.3: An overview of the studies that have characterised the fetal brain in 3D using different morphological properties. Boundary Separation shows whether the inner and outer cortical boundaries were separated for analysis. † indicates studies performed from postmortem fetal brains

Authors	Age Range (GW)	No. Subjects	Modality	Boundary Separation	Measures				
					Volume	Surface Area	Curvature/GI	Thickness	Sulcal Depth
Scott et al. (2011) [175]	20-31	48	MRI		✓				
Hu et al. (2009) [210]	21-37	43	MRI	✓			✓		
Kyriakopoulou et al. (2017) [209]	21-38	127	MRI		✓				
Clouchoux et al. (2012) [112]	25-35	12	MRI	✓			✓		✓
Clouchoux et al. (2013) [114]	25-37	18	MRI	✓	✓	✓	✓		✓
Xia et al. (2019) [176]	26-29	25	MRI	✓			✓		✓
Lefèvre et al. (2015) [115]	21-36	14	MRI	✓	✓	✓	✓		
Huang et al. (2009) [220]	13-22	15	MRI†	✓				✓	
Shimony et al. (2016) [216]	27-42	63	MRI	✓	✓		✓		✓
Xu et al. (2021) [212]	14 -22	27	MRI†	✓		✓	✓	✓	
Studholme (2011) [221]	20-28	NA	MRI	✓			✓		

volume, as shown in Figure 3.9. From a closed volume, the inner and outer surfaces can be extracted using surface mesh algorithms discussed previously. However, due to the large acoustic impedance of the skull, in ultrasound imaging, only the distal hemisphere is clearly visible. From a single hemisphere cortical plate volume, it is extremely challenging to extract the inner and outer cortical boundary, as they cannot be turned into a closed structure. Therefore, ultrasound volumes are often incompatible with the methods designed for MRI.

The work presented in Table 3.3 all characterise healthy cortical development using a range of measures. Across this Chapter, I have shown how at-risk or diseased pregnancies are often found to deviate from the normal projections. Therefore, these characterisations have the potential to be used as a benchmark for healthy development, in order to compare at-risk groups too. For such methods to be eventually integrated into routine care, it would be most impactful to study the gestational age range around the anomaly scan, e.g. 18 to 21 GW (see Section 3.2) and from using ultrasound, the modality of choice in prenatal care. However, as shown in Table 3.3, only the postmortem MRI studies have characterised development in 3D earlier than 20 GW, with the majority of the methods studying the late second and third trimesters. Further, due to the nature of postmortem studies, healthy development is not guaranteed.

3.4 Conclusion

The cortical plate undergoes complex morphological development across gestation, with healthy development found to follow stable growth trajectories. As at-risk or diseased pregnancies are often found to deviate from these trends, monitoring the development of the cortical plate in-utero has the potential to be used as an early biomarker. In order to do this, accurate characterisation of the developing cortex is required. Further, a range of measurements, e.g. depth, thickness, surface area and volume, should be performed in 3D, to fully describe the complex structure and to capture both global and local changes.

For such benchmarks to eventually be integrated into routine care, they should be calculated from routinely obtained ultrasound scans. However, to date, the vast majority of the analysis has been performed from MRI in the late second to the third trimester. To the best of our knowledge, to date, no work has characterised the development of the cortical plate in 3D from ultrasound volumes, creating a large gap between the research and clinical domain. In this thesis, I use a novel topology-constrained deep-learning network, to overcome the limitations of large artefacts within ultrasound volumes, to accurately characterise the developing cortex in 3D in the second trimester (e.g. when the anomaly scan is performed). Further, I propose a fully automated pipeline, to analyse new ultrasound volumes.

4

Topology Preserving Segmentation Method

4.1 Overview

In this chapter, I present the Topology Encouraging Deformation Segmentation Network (TEDS-Net): a segmentation network that preserves anatomical topology whilst maintaining segmentation performance that is competitive with SOTA baselines. I show how current SOTA segmentation methods, e.g. CNNs and Transformers, can introduce problematic topological errors, whereas, TEDS-Net achieves anatomically plausible segmentation by using learnt topology-preserving fields to deform a prior. The properties of topology-preserving fields are only guaranteed in the continuous domain and can begin to break down when stored discretely. Here, I introduce additional modifications that more strictly enforce topology preservation when describing fields in the discrete domain.

Early versions of the work presented here can be found in:

Madeleine K Wyburd, Nicola Dinsdale, Mark Jenkinson, and Ana IL Namburete. "TEDS-Net: Enforcing diffeomorphisms in spatial transformers to guarantee topology preservation in segmentations." International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). Springer, Cham, 2021.

Madeleine K Wyburd, Nicola Dinsdale, Mark Jenkinson, and Ana IL Namburete. "Anatomically Plausible Segmentations: Explicitly Preserving Topology Through Prior Deformations." *In Revision*: submitted to Medical Image Analysis (MedIA) (2022)

Versions of this work have also been presented at the following:

Madeleine K Wyburd, Nicola Dinsdale, Mark Jenkinson, and Ana IL Namburete. "Anatomically Plausible Segmentations: Explicitly Preserving Topology Through Prior Deformations" at Oxford Biomedical Imaging Festival 2022

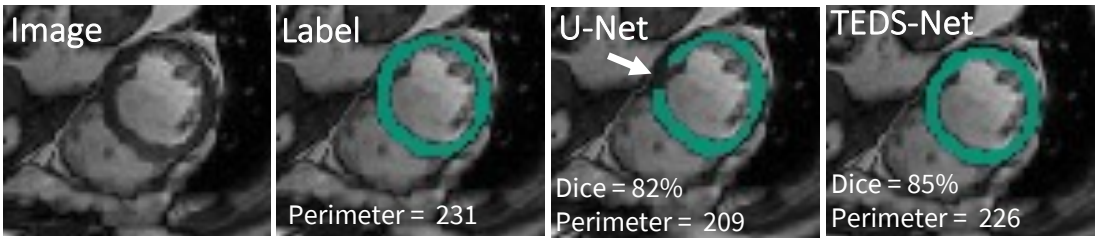


Figure 4.1: Example of topological errors found when segmenting medical images using traditional CNNs, i.e. a U-Net style architecture, trained with a pixel-wise loss function. Here I show myocardium segmentation from 2D MRI short-axis slices. The gap found within the U-Net example (indicated with the white arrow) changes the topology and interferes with automated measures of the perimeter.

4.2 Introduction

For a medical segmentation to be considered anatomically plausible its *topology* should be equivalent to that of the true anatomy. The topology, or *higher-level structure*, describes the overall shape properties of a structure that are preserved under continuous deformations, e.g. without joining or tearing, or closing or creating holes within the object, as discussed at length in Section 2.6.

The heart's myocardium label from the 2D cross-section shown in Figure 4.1, can be described as a single connected component with a hole, and thus has a Betti number of $b_0 = 1, b_1 = 1$. In such a 2D cross-section, if the myocardium segmentation had $n > 1$ components (e.g. $b_0 = n, b_1 = 1$) or $n > 1$ holes (e.g. $b_0 = 1, b_1 = n$), or no hole at all (e.g. $b_0 = 1, b_1 = 0$), its topology would not match that of the true anatomy and it would be considered indicative of a major anatomical defect. Myocardium segmentations are often required for tasks such as measuring perimeter, a biomarker for cardiac disease [222, 223]. Topological errors can make such measurements problematic, possibly by causing errors in downstream processing steps or introducing substantial errors.

CNNs are now the basis of most newly developed segmentation algorithms [20, 79]. However, as these networks are typically trained and evaluated using pixel-wise loss functions, such as Dice loss or Binary Cross Entropy (BCE), the higher-level structural information is often neglected, which can lead to anatomically

incorrect segmentation results. Further, due to the nature of convolutions, these networks only analyse local features, as discussed in Section 2.1.3. These factors can result in topological errors, as shown in Figure 4.1, which, in this example interfere with the clinically-required measure of the perimeter, defeating one of the primary purposes of the segmentation. Moreover, it should be noted that the severity of the impact of these topological errors is not reflected in the relatively high Dice score, illustrating the limitation of optimising and evaluating with such pixel-wise functions. Therefore, when performing segmentation tasks, there is a need to evaluate the topological performance in addition to pixel-wise accuracy for a more thorough assessment of the segmentation quality.

Here, I present TEDS-Net, a segmentation network used to deform a prior shape through learnt topology-preserving fields, which accounts for the discrete limitations through novel modifications and can result in 100% topology preservation. In this chapter, I first discuss how topology and topology-preserving fields are defined in a discrete setting (Section 4.3). Next, I describe TEDS-Net's architecture and our modifications to encourage topology preservation more (Section 4.4), before applying a series of ablation studies and hyper-parameter explorations (Section 4.5). Finally, I compare TEDS-Net performance to a series of SOTA segmentation methods (Section 4.6).

The main contributions of the chapter are as followed:

1. The design of a novel segmentation method, that incorporates prior knowledge of a structure's topology to produce topologically accurate segmentations.

2. Extensive hyper-parameter exploration to show where the topology performance begins to break down and the limits of the methods.

3. Comparison to SOTA to show competitive segmentation performance, whilst outperforming on topology.

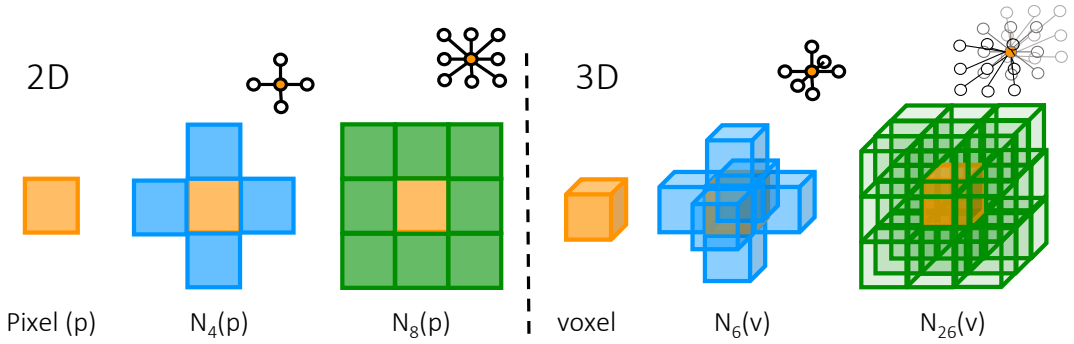


Figure 4.2: Schematic diagram showing different types of neighbourhood (N) around a 2D pixel and a 3D voxel. In the top corners, the connections to pixel and voxel in each neighbourhood are shown in graph notation.

4.3 Theory

Both *topology* and *topology-preserving fields* are traditionally defined in the continuous domain, however, throughout this work the images and fields are digitised. In this section, I describe how these properties can vary or be violated in discrete settings and introduce methods to compensate for such effects.

4.3.1 Topology

As introduced in Chapter 2.2.0.6, topology is a measure of shape, which can be calculated by counting the number of connected components and holes, as described by Betti numbers.

Topology in the discrete setting is poorly defined, which is largely due to the ambiguous nature of "connectivity" in the discrete domain. Connectivity is central in calculating Betti numbers, as b_0 measures the number of *disconnected* regions of an object (or *foreground*), whereas an n-dimensional hole b_n is defined as a region of *background disconnected* from another region of *background* and only bordering a single component of *foreground*.

The theory of topology was born from the study of continuous functions and physical objects [224, 225], where the idea of connectivity is relatively simple: *is a structure touching or not*. However, when moving into discrete grid points, there are multiple ways of defining a connection, which can change the measure of topology.

A pixel in a discrete grid is connected to all other pixels within its *neighbourhood* (N). However, there are several ways to define a neighbourhood, as shown in Figure 4.2. For example, for a pixel $p = (x, y) \in \mathbb{Z}^2$ in 2D, neighbourhoods can be made up of

$$N_4(p) = \{(x, y), (x + 1, y), (x - 1, y), (x, y + 1), (x, y - 1)\} \tag{4.1}$$

or

$$N_8(p) = N_4(p) \cup \{(x + 1, y + 1), (x + 1, y - 1), (x - 1, y + 1), (x - 1, y - 1)\}. \tag{4.2}$$

Two pixels (p, q) are said to be connected under n-connectivity if $q \in N_n(p)$ [225]. Thus depending on which neighbourhood definition is used, a pixel's connectivity changes.

Using these definitions of connectivity, image topology can be measured. Two *foreground* pixels (p, q) belong to the same connected component C, if there is a sequence of *foreground* pixels $(p_0, p_1, .., p_n)$ in which $p_0 = p$ and $p_n = q$, where p_i is a neighbour of p_{i-1} for $i \in 1, ..., n$, thus forming a continuous path of foreground between p and q [226].

It is common practice to use one type of connectedness for the foreground (e.g. 4-connectivity) and another for the background (e.g. 8-connectivity), indicated with notation $(n_{\text{foreground}}, n_{\text{background}})$-connectivity, as it allows borders to be treated as closed curves or boundaries, which is crucial for preventing topological conflicts [226, 227]. Examples of such conflicts are shown in Figure 4.3, where the background and foreground pixels have been connected to other pixels of their kind using different combinations of connectedness (as defined in Figure 4.2). When using the same connectivity for both such as (8,8)-connectivity, a region of foreground completes a closed loop, encompassing an isolated region of background, indicated by a red arrow. However, the hole is still connected to another part of the background outside the foreground, contradicting the previous finding and resulting in a topology conflict. Similar violations are shown when using (4,4)-connectivity. Conversely,

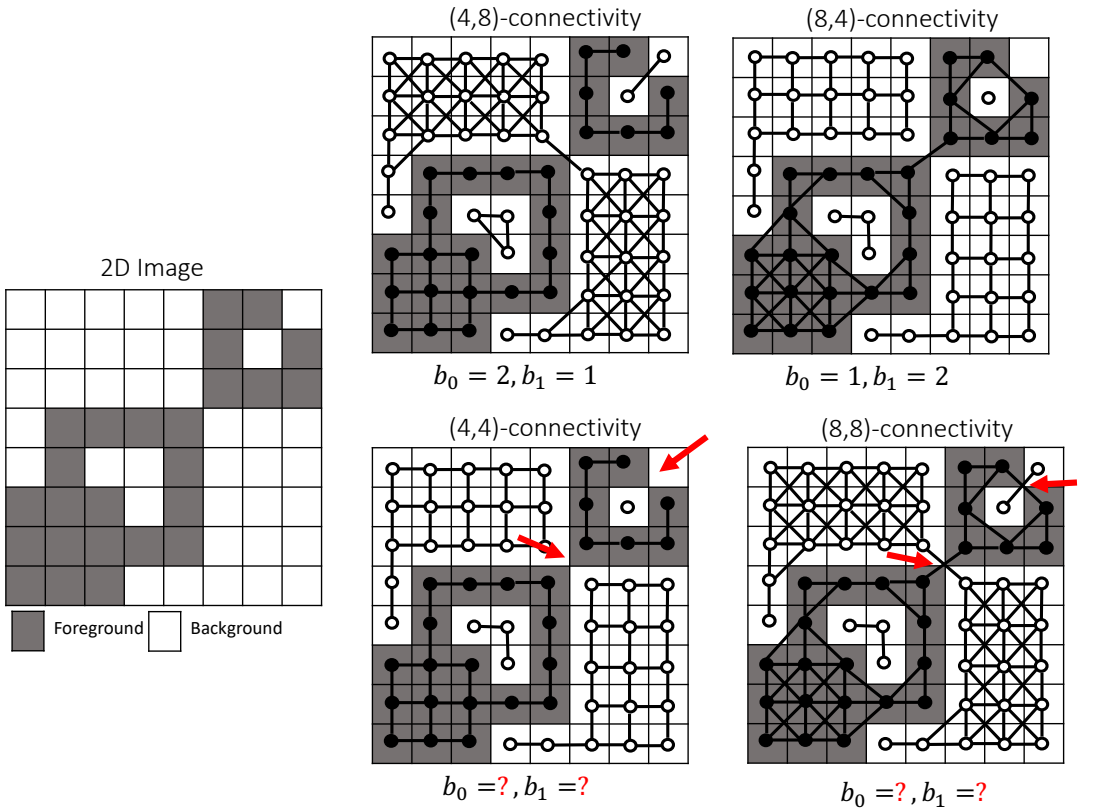


Figure 4.3: An overview of measuring topology from 2D digital images. Using different definitions of connectivity results in different Betti numbers. Moreover, opposing connectivity should be used for foreground and background pixels, to prevent topological ambiguities (indicated by red arrows) [225].

using either (4,8)- or (8,4)-connectivity, results in clear distinctions of b_0 and holes b_1. However, despite the underlying images being the same they result in different measures of topology. Thus illustrating how topology is not an intrinsic property of digital images.

As the choice of digital connectivity can alter the reported topology of a structure, it is important to understand the difference. As shown in Figure 4.3, when using (4,8)-connectivity, regions touching at the corners are no longer classed as "connected". However, these thinly connected regions are likely to be classed as "connected"by an observer. Therefore, across this work, to stay consistent, I use (8,4)-connectivity to define connected components and holes in 2D. The topological ambiguity and violations of discrete representation are also seen in 3D. To overcome this, I use (26,6)-connectivity for our 3D data.

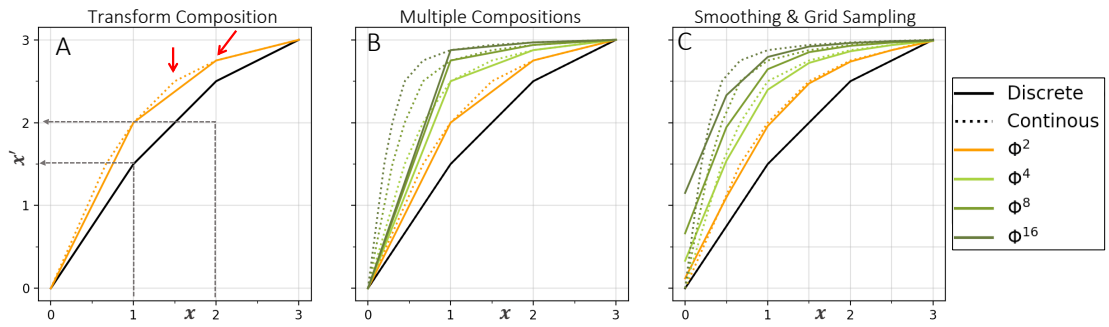


Figure 4.4: Panel **A** shows a simple 1D transform $\mathbf{\Phi}^1$ and the composition (one application of Scaling and Squaring) obtained through both discrete (solid line) and continuous (dotted line) methods. The grey arrows illustrate the two steps used to calculate $\mathbf{\Phi}^2$ at $x = 1$, $\mathbf{\Phi}^2(1) = \mathbf{\Phi}^1(\mathbf{\Phi}^1(1)) = \mathbf{\Phi}^1(1.5) = 2$. Panel **B** shows a series of composition layers and Panel **C** the same transforms after smoothing and using more finely sampled points.

4.3.2 Topology-preserving fields

For a deformation field ($\mathbf{\Phi}$) to be topology-preserving it must be continuous, that is, it must have a continuous inverse and bijective (one-to-one mapping) at all locations [89, 90], as discussed in Section 2.3. However, in practice when fields are represented on a discrete grid of points ($\mathbf{x}$) these criteria no longer apply directly [87]. In fact, when describing the properties of a discrete field ($\mathbf{\Phi}$), we are really referring to the nature of the continuous version ($\mathbf{\Phi}_c$) obtained by interpolating between grid points. Although interpolation generates a continuous field, it does not guarantee a one-to-one mapping or continuous derivatives and, therefore, cannot guarantee topology preservation [87]. Therefore, the methods used to generate topology-preserving fields in the continuous domain generally do not translate when working and storing fields in the discrete domain [87].

In the discussion that follows, I introduce how the discrete representation of such fields can change their nature and introduce the additional modifications used in TEDS-Net, to strongly encourage topology preservation in discretely represented fields.

In this work, to produce large and flexible deformations fields, I utilise the Scaling and Squaring method [91, 92] as introduced in Section 2.3.2.2 and described

in Equation 2.8. In short, this method generates topology-preserving fields, by resampling a field $\mathbf{\Phi}^1$ by itself ($\mathbf{\Phi}^2 = \mathbf{\Phi}^1(\mathbf{\Phi})^1$), h times. Figure 4.4 illustrates a simple 1D transform, $\mathbf{\Phi}$, over grid points: $\mathbf{x} = (0, 1, 2, 3)$. When performing $\mathbf{\Phi}^2 = \mathbf{\Phi}^1 \circ \mathbf{\Phi}^1 = \mathbf{\Phi}^1(\mathbf{\Phi}^1)$ in the discrete domain (e.g. at each grid point) compared to the continuous domain, differences arise, as shown in Figure 4.4A. Some differences result in gradients being created or removed within the transform (as indicated by the red arrows), changing the direction of the displacements and therefore likely to disturb the underlying topology (as seen previously in Figure 2.8). These artefacts often occur when storing the interpolated fields ($\mathbf{\Phi}_c$) in discrete locations, which risks the loss of important transform information between grid points. Moreover, these artefacts risk becoming more distinct after a series of compositions, where the fields are repeatedly interpolated, followed by discrete sampling and storing, further violating the continuous properties. This effect is shown in Figure 4.4B, where the created gradient in the discrete domain is much steeper than that in the continuous. These differences show that the methods used to generate topology-preserving fields in the continuous domain, such as scaling and squaring, may not directly translate into topology-preserving fields in the discrete domain [87].

In this work, to try and mimic the topology-preserving behaviour of the continuous representations, I introduce methods to reduce large gradients created between grid points of discrete fields, so that they more closely resemble their "continuous" counterparts. Decreasing the space between grid points can also help to better describe a transform, decreasing the abnormalities created through storing the interpolated fields, as shown in Figure 4.4C, where the spacing has been reduced by a half. Furthermore, by encouraging the fields to be smooth, steep gradients created through interpolation can be reduced. In Figure 4.4C, Gaussian smoothing has been applied between each composition layer to smooth the transform between the discrete grid points, as previously discussed in Section 2.3.2.4 [93, 100]. The modified discrete-domain fields in Figure 4.4C more closely resemble their continuous counterparts, and therefore are more likely to exhibit the same behaviour, compared to the discrete fields in Figure 4.4B. It is important to

note that, even with these additional modifications, there are still some differences between the two fields in Figure 4.4C.An example of these differences can be seen at $x = 0.5$ in panel C, where the transformed line has moved further from the continuous positions. Therefore, although these methods do not enforce behaviour identical to the continuous fields, as the artificial steep gradients have been smoothed, they can encourage continuously-defined properties such as topology-preservation. In this work, as generating topology-preserving fields is a priority, I include the additional modification, despite the differences caused.

4.4 Method

4.4.1 Problem Definition

Spatial transformer networks aim to segment a structure of interest by deforming a prior shape $\mathbf{P}$ using a learnt deformation field $\mathbf{\Phi}$ resulting in a segmentation $\hat{\mathbf{Y}}$: $\hat{\mathbf{Y}} = \mathbf{\Phi}(\mathbf{P})$. If $\mathbf{\Phi}$ is topology-preserving and $\mathbf{P}$ has the same topology as the structure of interest, then $\hat{\mathbf{Y}}$ will always have the correct topology.

4.4.2 Architectural Design

The architecture of TEDS-Net can be divided into three main components, namely: (i) the encoder-decoder; (ii) the topology-preserving fields layer; and (iii) prior sampling. A schematic of the network is shown in Figure 4.5. The encoder-decoder network is used to extract relevant features from an input image ($\mathbf{X}$) in order to predict initial displacement fields ($\mathbf{u}$). Like previous work [25, 98], these fields are then encouraged to be topology-preserving using a series of Scaling and Squaring layers but with our additional modifications to try to more strictly enforce topology preservation in the discrete domain. Finally, these fields are used to sample a prior shape with equivalent topology to the anatomy, generating a segmentation.

TEDS-Net learns two deformation fields, generated at different resolutions of the decoder. The lower resolution field, which is generated deeper in the decoder, termed the *bulk* field, is used to learn a coarse deformation in order to deform the prior shape into the correct size and location. Whereas the higher resolution

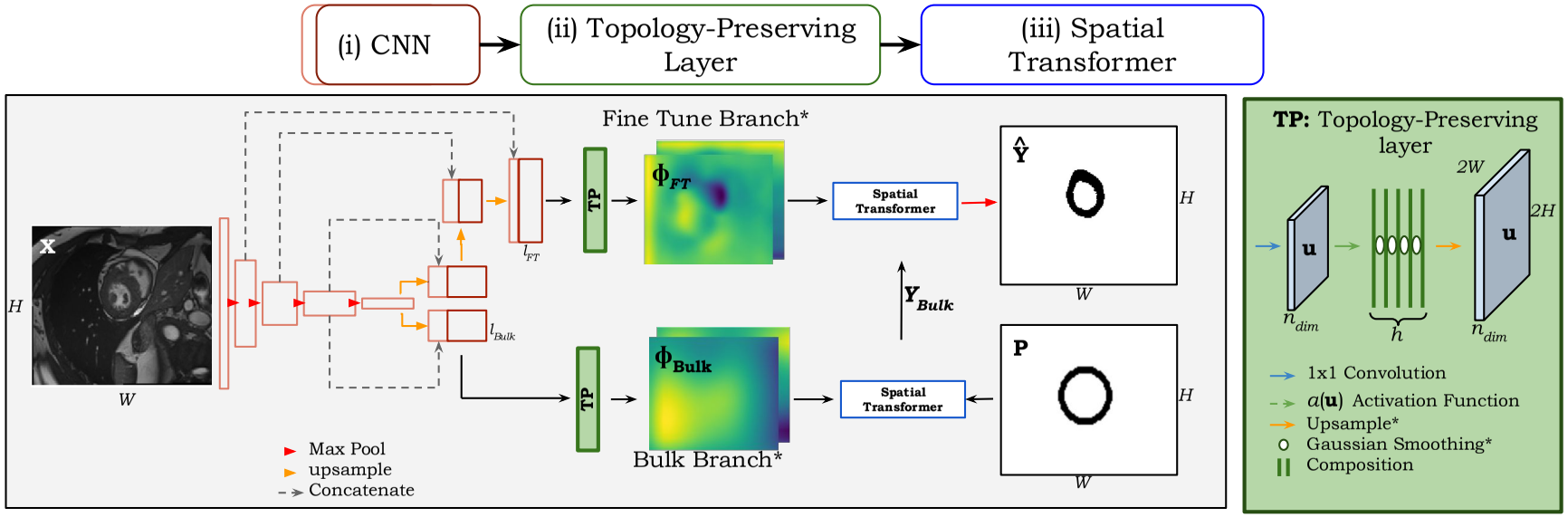


Figure 4.5: Schematic of TEDS-Net architecture. A CNN learns two initial fields, $\mathbf{u}$, at different resolutions, from an input image $\mathbf{X}$. The fields are encouraged to be topology-preserving using an activation function, amplified through composition layers and upsampled to twice the resolution of the input. The bulk displacement, $\mathbf{\Phi}_{\text{Bulk}}$, samples a binary prior, $\mathbf{P}$, generating $\mathbf{Y}_{\text{Bulk}}$, which is then sampled by the fine-tuning field, $\mathbf{\Phi}_{\text{FT}}$. The asterisks show the elements removed during ablation studies.

field, the *fine-tuning* (ft) field, learns a more detailed displacement field to more precisely delineate the anatomical structure.

4.4.2.1 (i) Encoder-decoder

An encoder-decoder network (g_θ), described by parameters θ, is used to extract relevant features from an input image of dimensions $\mathbf{X} \in [H \times W]$ to learn two initial fields: $g_\theta(\mathbf{X}) = [\mathbf{u}_{\text{bulk}}, \mathbf{u}_{\text{ft}}]$. The network resembles that of the U-Net [23], a popular segmentation network previously introduced in Section 2.1.3.1, as it is made up of a contracting encoder branch that is connected through skip connections to the expanding decoder branches. The encoder branch is made up of a series of convolutional blocks that apply two sets of convolutions with a ReLU activation function and instance normalisation between them [228]. Initially, f feature maps are used and between blocks, the number of feature maps is doubled and their spatial resolution is halved for l layers.

Two unconnected decoder streams are attached to the encoder through a bottleneck, which is formed of two convolutional blocks, with skip connections joining each layer of the encoder with corresponding decoder layers. The two streams are used to predict the bulk and fine-tuning fields at different resolutions. The bulk and fine-tuning decoder branches have $l_{\text{bulk}}, l_{\text{ft}}$ convolutional layers, respectively, where $l_{\text{bulk}} < l_{\text{ft}}$. A final 1×1 convolution is applied to both branches to reduce the number of features to two in order to generate an n-dimension displacement field: $\mathbf{u}_{\text{bulk}} \in \mathbb{R}^{n_{dim} \times (H \times W)/2^{l_{\text{bulk}}}}, \mathbf{u}_{\text{ft}} \in \mathbb{R}^{n_{dim} \times (H \times W)/2^{l_{\text{ft}}}}$.

4.4.2.2 (ii) Topology Preservation

An activation function (a) is applied to both initial displacement fields ($\mathbf{u}_{\text{bulk}}, \mathbf{u}_{\text{ft}}$) to enforce them to be topology-preserving by limiting the maximum displacements to be less than half a grid space, thus preventing grid points crossing over and violating topology. When using a grid spacing of one voxel, this activation function can be written as:

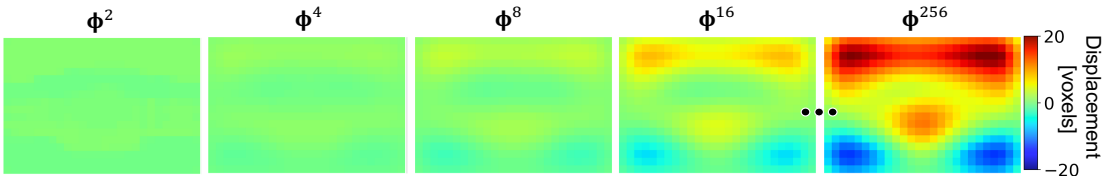


Figure 4.6: Visualisation of the scaling and squaring approach, shown in one direction. Initially, the fields have negligible displacements, but after composing it by itself $h = 8$ times ($\mathbf{\Phi}^{2^h} = \mathbf{\Phi}^{256}$) the displacements are substantially amplified.

$$a(\mathbf{u}(\mathbf{x})) = 0.5 \left(\frac{e^{\mathbf{u}(\mathbf{x})} - e^{-\mathbf{u}(\mathbf{x})}}{e^{\mathbf{u}(\mathbf{x})} + e^{-\mathbf{u}(\mathbf{x})}} \right). \tag{4.3}$$

These fields are then amplified through h scaling and squaring layers (adapted from VoxelMorph's implementation as discussed in Section 2.3.2.2 [40]), which is performed using a series of composition layers described in Equation 2.8 and shown in Figure 4.6.

Between composition steps, a Gaussian smoothing kernel with sigma σ and size $k \times k$ is applied to reduce the steep gradients created through interpolation and discrete storing, as described in Section 4.3.2. The impact of σ is evaluated in Section 4.6.2.1. The fields are then upsampled to twice the input resolution, to reduce the grid spacing describing the field: $\mathbf{\Phi}_{\text{Bulk}} \in \mathbb{R}^{n_{dim} \times 2H \times 2W}$, $\mathbf{\Phi}_{\text{FT}} \in \mathbb{R}^{n_{dim} \times 2H \times 2W}$.

It is important to note that within TEDS-Net topology preservation is incorporated in two distinct ways. Firstly, the activation function ($a(\mathbf{u}(\mathbf{x}))$) is used to strictly enforces topology preservation by restricting the initial deformations to be less than a pixel. This is required to ensure the subsequent scaling and squaring layers are also topology preserving. Whilst, additional modifications are required to overcome violations caused by storing and applying the transforms in the discrete space. Specifically, Gaussian smoothing and upsampling are used to reduce steep gradients created within the field.

4.4.2.3 (iii) Prior Sampling

The bulk displacement field is used to sample a prior ($\mathbf{P} \in [c \times H \times W]$), where c is the number of channels, $\mathbf{Y}_{\text{Bulk}} = \mathbf{\Phi}_{\text{Bulk}}(\mathbf{P})$ before the resulting image is then

resampled for a second time using the fine-tuning field: $\hat{\mathbf{Y}} = \mathbf{\Phi}_{\mathrm{FT}}(\mathbf{Y}_{\mathrm{Bulk}})$. The prediction ($\hat{\mathbf{Y}}$) is then downsampled to the original input resolution using Max Pooling before a binary threshold of T is applied.

The only prerequisite of the prior shape design is that it has the same topology as the anatomy of interest. In TEDS-Net, $\mathbf{P}$ can be a rough approximation of the anatomy, as the two-step deformation approach can first align and scale the prior before performing the more detailed deformation. In this, $\mathbf{P}$ could be a simple solid sphere to segment any structure with the same topology, e.g. a liver or brain, despite the structure's shape being drastically different. In TEDS-Net, the same $\mathbf{P}$ is used for each structure of interest, with the size investigated in Section 4.5.3.3.

4.4.3 Loss Functions

To regularise the training of TEDS-Net, both segmentation and field regularisation loss functions are used. Dice loss ($\mathcal{L}_{\mathrm{Dice}}$) is used to penalise poor overlap between the $\mathbf{Y}$ and $\hat{\mathbf{Y}}$, in order to improve segmentation performance, previously discussed in Section 2.2.0.2. A field regularisation loss term ($\mathcal{L}_{\mathrm{Grad}}$) is used to further encourage smooth deformation fields:

$$\mathcal{L}_{\mathrm{Grad}} = \sum_{i,j=1}^{2H,2W} \left\| \begin{matrix} \mathbf{\Phi}(i+1,j) - \mathbf{\Phi}(i,j) \\ \mathbf{\Phi}(i,j+1) - \mathbf{\Phi}(i,j) \end{matrix} \right\|^2 \tag{4.4}$$

where this measures and penalizes the spatial gradient at each grid point across the fields, modified from [96].

The network is trained with a weighted combination of these loss functions:

$$\mathcal{L} = \mathcal{L}_{\mathrm{Dice}}(\mathbf{Y}, \hat{\mathbf{Y}}) + \alpha \mathcal{L}_{\mathrm{Grad}}(\mathbf{\Phi}_{\mathrm{FT}}) + \beta \mathcal{L}_{\mathrm{Grad}}(\mathbf{\Phi}_{\mathrm{Bulk}}), \tag{4.5}$$

where the weighting parameters α and β are used to balance the contribution of the two field regularisation loss terms. The effects of α and β are investigated in Section 4.6.2.2.

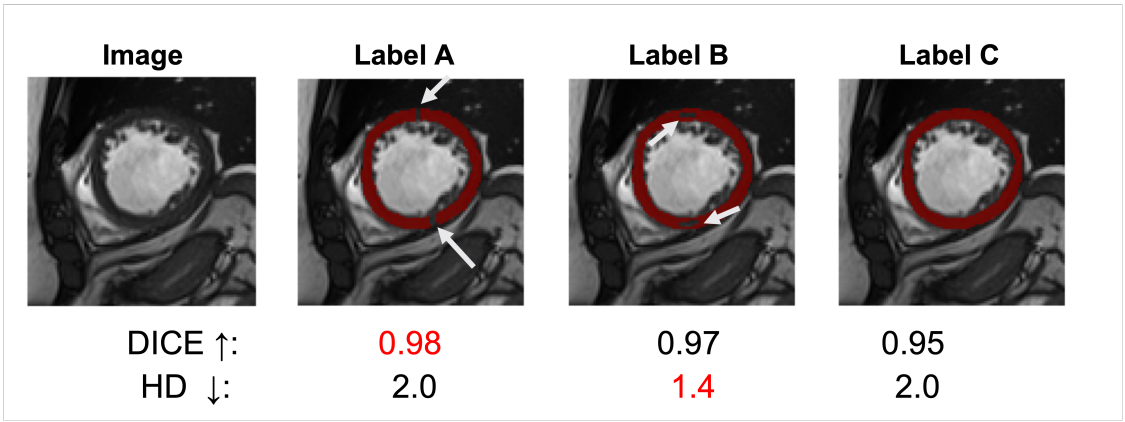


Figure 4.7: A mock example of three myocardium segmentation, displaying their corresponding Dice and Hausdorff Distance (HD). Label A outperforms the others on Dice, whereas label B outperforms on HD. However, both have topological errors (indicated by white arrows), which are not reflected in the other measure's scores.

4.4.4 Evaluation Metrics

To evaluate the segmentation performance, I used a range of metrics introduced in Chapter 2.2 to provide a thorough overview of the performance, including Dice, Binary cross entropy (BCE), Hausdorff Distance (HD), Average Surface Symmetric Distance (ASSD) and Topology. The assessment of topology was a binary score, as shown in Equation 2.6. If the topology of the prediction was equivalent to the known topology $L_B = 1$, else the topology would be deemed incorrect and $L_B = 0$. This is a binary measure and does not take into account the degree of topological mistakes. As previously discussed in Section 2.2.1, using a range of metrics, e.g. describing the segmentation overlap, distance and shape, is vital to assess the segmentation quality and to understand if and where the segmentation has gone wrong. Figure 4.7 shows a mock example of myocardium segmentation, to reiterate this point. Despite label A and label B outperforming C on Dice and HD, respectively, neither is topologically accurate, thus illustrating why a range of metrics is required.

Further, to assess whether the generated fields were topology-preserving, I measured the Jacobian determinant at each position, using Equation 2.9. The Jacobian determinant of our generated fields ($\mathbf{\Phi}$) describes the behaviour around each voxel (x) as being either expanded: $|J_{\mathbf{\Phi}}(\mathrm{x})| > 1$, squeezed: $0 < |J_{\mathbf{\Phi}}(\mathrm{x})| < 1$, or folded: $|J_{\mathbf{\Phi}}(\mathrm{x})| \leq 0$. Topology-preserving fields should thus have positive

Jacobian determinants at every grid point. Throughout this work, I record the percentage of non-positive Jacobian determinants found in each generated field, which is denoted by $\%\,|J_{\mathbf{\Phi}}| \leq 0$.

In traditional pixel-wise labelling segmentation methods, if the network predicted nothing it would likely be picked up by the evaluation measures. However, in TEDS-Net, if the network predicted no displacement field ($\mathbf{u}(\mathbf{x})$), no deformation would be performed and a segmentation identical to the prior would be 'predicted'. In both cases the networks have failed to predict anything, however, this is likely harder to detect in TEDS-Net as it still returns a segmentation with the correct topology, and generates fields containing no folding voxels ($\%\,|J_{\mathbf{\Phi}}| \leq 0 = 0$). Further, the segmentation might return believable distance metrics and a reasonable Dice score if the prior and label have some overlap. Utilising the range of metrics described above combined with qualitative inspection of the segmentations should be able to detect this potential limitation of TEDS-Net, however, it is important to note that it may be harder to detect a failed segmentation in TEDS-Net than traditional methods.

4.5 Experimental Set-up

TEDS-Net was designed to try and enforce accurate topology for segmentation tasks. However, it is built upon several mathematical theories that are only guaranteed in the continuous setting, and likely to be violated in the discrete domain. In this section, I design a range of experiments to investigate TEDS-Net performance on an open-source cardiac dataset to perform myocardium segmentation. I perform a series of ablation studies to investigate the impact of each of our topology-preserving modifications on the segmentation performance, before performing hyper-parameter exploration, to see the limits of our proposed method for this task. These experiments are to show that TEDS-Net can result in 100% preservation, but as it is dependant on the task and the hyper-parameters used and when in the discrete setting, it cannot be strictly guaranteed.

4.5.1 Dataset

I evaluated the performance of TEDS-Net on the task of myocardium segmentation using an open source dataset MRI dataset: ACDC[1] [222]. The ACDC dataset consists of two labelled MRI heart acquisitions per subject, from 100 subjects. When viewed on the short axis, the myocardium can be seen to encompass the left ventricle. In the majority of the slices, the myocardium's topology is therefore equivalent to a hollow circle ($b_0 = 1, b_1 = 1$). From each scan, I selected five short-axis slices, where the myocardium had this expected topology by assessing the topology of the label in each slice.

For the $\mathbf{P}$ a simple hollow circle was used and the size was investigated. It should be noted that the subjects' scans were not aligned and therefore the location of the myocardium was inconsistent. Additionally, different pathologies have varying myocardium thicknesses. Despite this, the same prior shape was used for each input $\mathbf{X}$, to illustrate the importance of the combination of the bulk and fine-tuning fields.

4.5.1.1 Data Augmentation

Rotations ($\pm 5°$), translations (± 5 voxels), flipping (vertical and horizontal), and resizing ($\pm 30\%$) were applied randomly to increase the dataset, resulting in 8,000 images split into 75%, 15% and 10% for training, validation and testing, respectively. Note that the dataset was split into each subset at the subject level, such that images from the same subject did not appear in more than one subset. The images were cropped around the centre to 144 by 208 pixels, to reduce the computational demand, whilst keeping the myocardium label intact. Each image was randomly cropped around the myocardium label so that all of the labels were still present whilst ensuring the images were not centred around the myocardium.

4.5.2 Implementation details

TEDS-Net was implemented in Python 3.7 using Pytorch 10.6. The network was developed on an Intel Xeon E-2146G CPU (3.50GHz, 6 cores) and an Nvidia GTX

[1]The ACDC database: `www.creatis.insa-lyon.fr/Challenge/acdc`

1080 Ti. I trained the model from scratch using the Adam optimiser with a learning rate of 0.0001. I set the batch size to 5 and trained for 200 epochs. TEDS-Net was made up of $l = 5$ layers, with $f = 12$ initial feature maps, and a 20% spatial dropout applied to each feature map.

4.5.3 Experiments

4.5.3.1 Ablation Study

An ablation study was performed to show the contribution of key elements within TEDS-Net. In TEDS-Net, I included additional modifications to try and tackle the limitations of discrete sampling and storage, namely: (i) Gaussian smoothing, (ii) upsampling, (iii) field regularisation and (iv) the activation function $a(\mathbf{u}(\mathbf{x})$. In our ablation studies, I investigate how removing these components affects the pixel-wise segmentation performance, the number of folding voxels within the fields and the final topology performance. In previous work [40, 98], field regularisation has been found to play a large role in preventing folding voxels, however, the effect of Gaussian smoothing or upsampling is unknown. Another novel element of TEDS-Net is the dual deformation branch approach, which is also investigated. The removed components are shown with an * in Figure 4.5.

4.5.3.2 Hyper-parameter Exploration

TEDS-Net is designed to preserve the topology of a prior after a deformation. This is facilitated by generating topology-preserving fields through traditional methods, e.g. Scaling and Squaring method, and further reinforced through our additional modifications. Across our network, I identified four parameters that play a crucial role in determining whether folding voxels are present; (a) the σ of the Gaussian smoothing kernels used between each composition layer, (b) the field regularisation weighting parameters (α and β), (c) the interpolation methods used in upsampling, compositions, and image resampling, and (d) the number of composition layers h used in the Scaling and Squaring approach. I further investigate these parameters

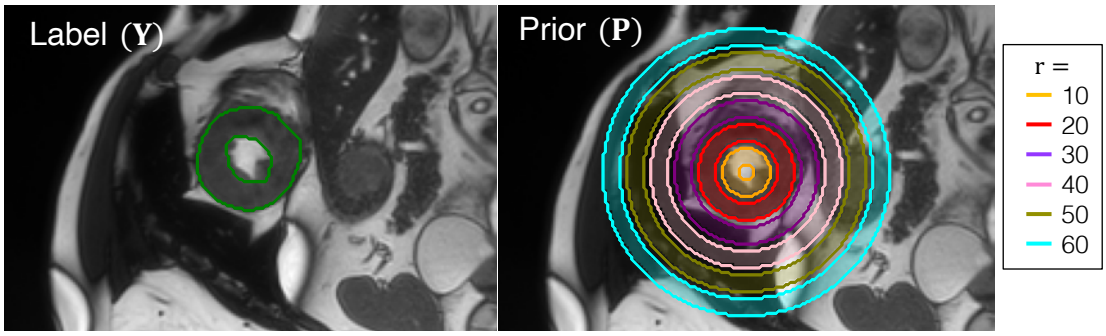


Figure 4.8: Visualisation of the range of prior shapes used to segment the myocardium in TEDS-Net.

here. Unless otherwise stated, I used $h = 8$, $\alpha = \beta = 10,000$, $\sigma = 2$ voxels and linear interpolation for all sampling.

4.5.3.3 Prior Shape

To segment the myocardium I used a simple hollow circle as the prior. In the training set, the myocardium outer radius ranged from approximately 10 to 60 voxels, after applying the minimum and maximum scaling augmentation. I investigated the impact of varying the prior's outer radius (r) on the final segmentation performance. Figure 4.8 shows the range of priors used.

4.5.3.4 Baseline Comparison

I compared the performance of TEDS-Net with the performance of several baselines, namely: the standard U-Net [23]; a transformer network Swin-Unet [24]; VoxelMorph [25]; and, using the topological loss function, TopoLoss [26]. The U-Net and Swin-Unet were primarily used as a comparison of segmentation performance, trained solely on Dice loss.

VoxelMorph was used as a baseline to compare both the segmentations and the deformation fields with respect to topological preservation. I used VoxelMorph's architecture as described in [25] and previously introduced in Section 2.3.2.2. Unlike TEDS-Net, VoxelMorph was optimised for two inputs passed into the CNN: the source and target volumes. To mimic this approach for segmentation, I passed both the MRI slice ($\mathbf{X}$) and the prior ($\mathbf{P}$) into our VoxelMorph implementation.

A combination of Dice loss and field regularisation was used for training $\mathcal{L}_{\text{vm}} = \mathcal{L}_{\text{Dice}} + \omega\mathcal{L}_{\text{Grad}}$, with the weighting factor $\omega = 3$ (chosen empirically, see Appendix A).

Similar to the TEDS-Net implementation, both the U-Net and VoxelMorph were trained for 200 epochs, with a learning rate of 0.0001 using the Adam optimiser and a batch size of 5. Both architectures had a depth of $l = 4$ and $f = 12$ initial feature maps.

As an additional segmentation baseline, I implemented Swin-Unet: a transformer-based segmentation network, which has previously shown state-of-art performance on 2D cardiac segmentation [24]. Transformer networks are often thought to better represent the global information within a task, compared to CNNs, due to the small receptive field within convolutions, as described in Section 2.1.3.3. Therefore, myocardium topology may be better learnt using Swin-Unet than traditional CNN methods, such as the U-Net. Following the implementation described in [24], Swin-Unet was initialised with ImageNet before being applied to the ACDC dataset. The architecture of Swin-Unet was kept the same as reported in [24] and the ACDC images were padded to 244×244 to keep the patch size and the number of layers consistent with the original implementation. Swin-Unet was trained for 200 epochs with a batch size of 24, the learning rate was initially set to 0.005, decreasing by 50% every 10 epochs, as implemented in the original paper.

As the final baseline comparison, I used a topology-preserving loss function: TopoLoss ($\mathcal{L}_{\text{TopoLoss}}$) [26]. This compares the topology of a predicted segmentation to the ground-truth segmentation using the concept of persistent homology, previously introduced in Section 2.3.1 [26]. As this method is extremely time-consuming, it is recommended that it be used on patches of an image and applied to a pre-trained segmentation model. To implement this method, I, therefore, used the trained U-Net (as described above) and fine-tuned all layers of the network using a combination of $\mathcal{L}_{\text{Dice}}$ and $\mathcal{L}_{\text{TopoLoss}}$: $\mathcal{L} = \mathcal{L}_{\text{Dice}} + \epsilon\mathcal{L}_{\text{TopoLoss}}$ for 100 epochs using patches of 64 by 64, where $\epsilon = 0.2$ (as found empirically - see Appendix A). Following the authors' guidelines, post-processing operations were applied to the final predictions, as TopoLoss can create small disconnected components. I applied

a binary opening operation using a kernel of size 3×3 to remove these isolated points, as recommended by the authors, and it was found that this dramatically improved the topological performance.

4.6 Results and Discussion

4.6.1 Ablation Analysis

Ablation studies were performed to show the effect that the different components within TEDS-Net had on the segmentation performance and the nature of the predicted fields, as shown in Table 4.1 (A0-7). Qualitative examples of the ablation experiment is shown in Figure 4.9

The deformation fields generated in TEDS-Net, when using all of the components proposed in Section 4.4.2, are found to be topology-preserving as they contained no folding pixels (shown with $\%\,|J_{\boldsymbol{\Phi}_{\text{Bulk}}}| \leq 0 = 0, \%\,|J_{\boldsymbol{\Phi}_{\text{FT}}}| \leq 0 = 0$). This results in 100% topological accuracy. Moreover, including all of the modifications was found to have the lowest Hausdorff distance (HD) and joint best Dice Score.

As discussed in Section 4.4.2.2, topology preservation is controlled in two distinct ways: (i) through the activation function $a(\mathbf{u}(\mathbf{x}))$ and (ii) through the modifications introduced to overcome the artefacts introduced by working in the discrete domains, namely, upsampling, $\mathcal{L}_{\text{Grad}}$, and Gaussian smoothing. Removing both sets of modifications (A0 in Table 4.1), resulted in TEDS-Net warping $\mathbf{P}$ to outside of the field of view as shown as in Figure 4.9 and by the low Dice score in Table 4.1. This is likely caused by the lack of constraints applied to the size of the deformation fields. Reintroducing $a(\mathbf{u}(\mathbf{x}))$, whilst leaving out the other modifications (A1), drastically improved the segmentation performance, achieving a Dice of 0.82. This result shows that $a(\mathbf{u}(\mathbf{x}))$ alone is enough to constrain the deformation fields in order to predict sensible segmentations. However, this ablation resulted in a large number of topological errors (17%), with examples shown in Figure 4.9 A1, which coincided with larger numbers of folding pixels ($|J_{\boldsymbol{\Phi}_{\text{FT}}}| \leq 0$) present in both generated fields. Moreover, it also performed worse

Table 4.1: Comparison of myocardium segmentation performed by TEDS-Net and ablation studies (A0-7). The mean and standard deviation are given for Dice, Hausdorff Distance (HD) and the percentage of non-positive Jacobian determinants from each field. The best result from each measure is in bold.

	TEDS-Net Components						Dice	HD [mm]	$\%\,\lvert J_{\Phi_{\text{Bulk}}}\rvert \leq 0$	$\%\,\lvert J_{\Phi_{\text{FT}}}\rvert \leq 0$	Topology
	Bulk Branch	FT Branch	Upsampling	$\mathcal{L}_{\text{Grad}}$	Gaussian Smoothing	$a(\mathbf{u}(\mathbf{x})$					
A0	✓	✓					0.00 ± 0.00	NA	60.9 ± 11.9	81.5 ± 6.8	0%
A1	✓	✓				✓	0.82 ± 0.19	4.26 ± 5.76	56.32 ± 1.97	43.8 ± 3.46	83%
A2	✓	✓	✓	✓		✓	$\mathbf{0.85 \pm 0.13}$	3.73 ± 8.34	67.10 ± 7.07	35.81 ± 4.41	97%
A3	✓	✓	✓		✓	✓	$\mathbf{0.85 \pm 0.14}$	2.77 ± 2.01	5.94 ± 2.57	22.15 ± 3.67	65%
A4	✓	✓		✓	✓	✓	$\mathbf{0.85 \pm 0.11}$	2.89 ± 1.61	$\mathbf{0.0 \pm 0.0}$	$\mathbf{0.0 \pm 0.0}$	**100%**
A5	✓		✓	✓	✓	✓	0.77 ± 0.15	3.86 ± 3.50	0.03 ± 0.17	N/A	**100%**
A6		✓	✓	✓	✓	✓	$\mathbf{0.85 \pm 0.14}$	3.01 ± 2.16	N/A	$\mathbf{0.0 \pm 0.0}$	**100%**
A7	✓	✓	✓	✓	✓		0.70 ± 0.13	3.69 ± 1.60	13.2 ± 0.6	$\mathbf{0.0 \pm 0.0}$	80%
	✓	✓	✓	✓	✓	✓	$\mathbf{0.85 \pm 0.14}$	$\mathbf{2.12 \pm 1.43}$	$\mathbf{0.0 \pm 0.0}$	$\mathbf{0.0 \pm 0.0}$	**100%**

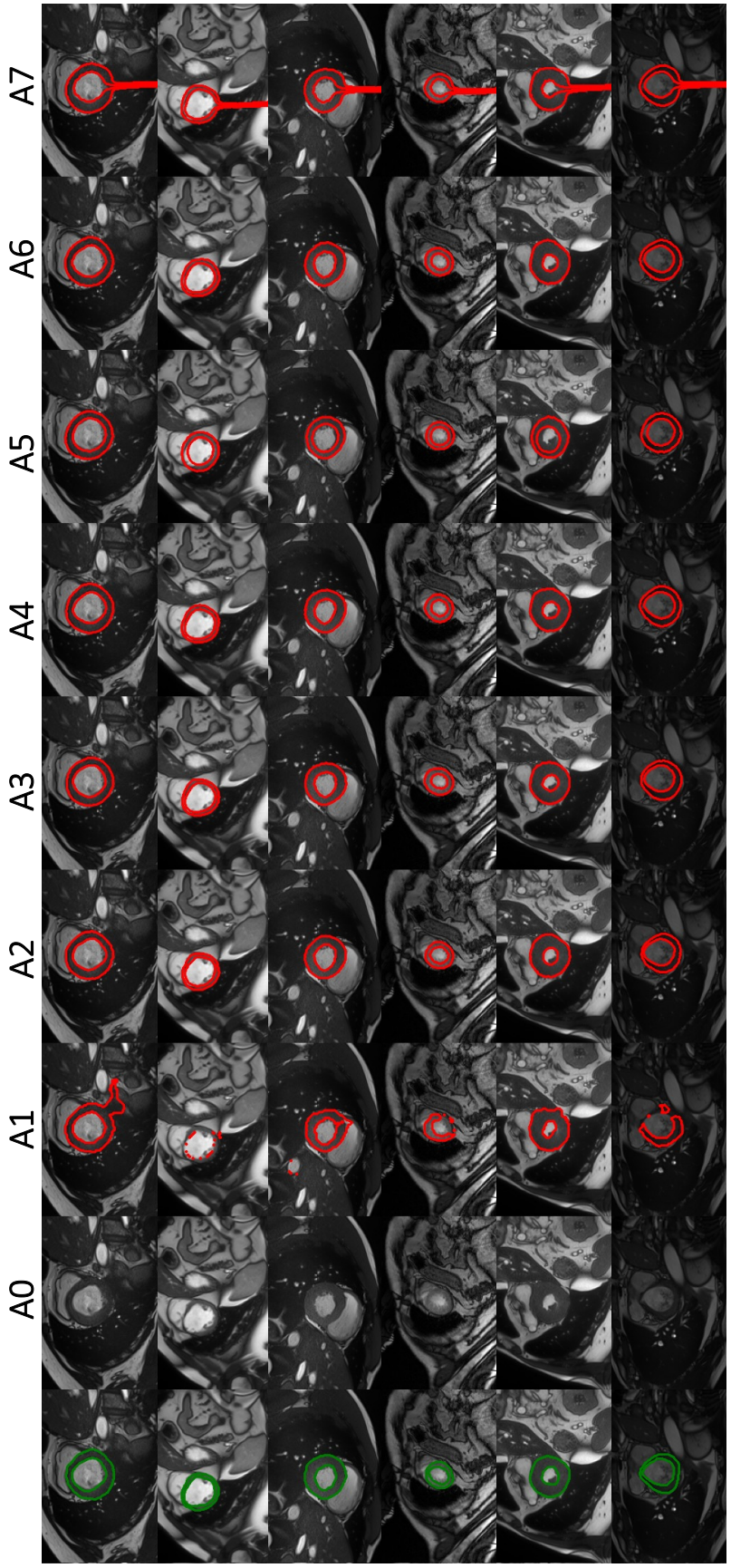


Figure 4.9: Qualitative examples of the eight ablation experiments, to visually show to the contribution of each feature within TEDS-Net. The ablation numbers correspond with that in Table 4.1. The ground truth labels are shown in green.

on Dice and HD than the full TEDS-Net architecture, implying that the topology-preserving modifications also improve the segmentation performance. Removing $a(\mathbf{u}(\mathbf{x}))$ and reintroducing the additional modifications (A7), were also found to produce myocardium segmentations, illustrating that upsampling, $\mathcal{L}_{\text{Grad}}$, and Gaussian smoothing can be used to constrain the deformation fields. However, large artefacts were found across the predictions (as shown in Figure 4.9 A7), which are likely due to the lack of constraint on the size of the initial displacement fields, allowing the network to learn smooth but large displacements that warp the prior partially out of the field of view.

To limit the numerical inaccuracies brought about by the repeated composition of two topology-preserving fields in the discrete setting, I introduced Gaussian smoothing between each layer. Without this addition, a large fraction of folding pixels were present in both fields, corresponding to a number of predictions with incorrect topology, as shown in Table 4.1 (A2). Sampling a discrete binary image with a finite warp field often results in disconnects, due to interpolations and the use of thresholds on the resultant image. To reduce this effect, I regularised the smoothness of the final deformation fields with $\mathcal{L}_{\text{Grad}}$. Without this term, folding pixels were present in both fields and 35/100 images were found to have topological mistakes, as shown in Table 4.1 (A3). It should be noted that without $\mathcal{L}_{\text{Grad}}$ the results achieved were the second best for pixel-wise performance but the worst for topology, which might suggest that this term plays the largest effect in restricting the segmentation to obey topology, which can negatively affect the Dice and HD.

A larger number of topological mistakes were found in A3 than A2, despite a smaller fraction of folding pixels. This is likely because topological mistakes only occur when the folding pixels are present near the location of the myocardium prediction $\hat{\mathbf{Y}}$. So despite a larger fraction of folding pixels, it is the location that affects whether it will lead to an incorrect hole or connection.

Despite topological mistakes being present in A1, A2 and A3, the pixel-wise performance, measured using popular segmentation evaluation metrics: Dice and HD, were found to be comparable to that with no topological errors, thus illustrating that

these traditional methods do not accurately reflect the underlying topology and additional measures of shape are required to accurately assess segmentation performance.

In A4, the fields were no longer upsampled to double the resolution of the input. This component, when included, aims to decrease the spacing between grid points, and consequently to further reduce topological violations from any interpolation steps. Removing the upsampling did not affect the topology-preserving nature of the fields, and had little effect on the segmentation performance.

Removing either the bulk or fine-tuning branch was found to reduce the performance of both Dice and HD, as shown in Table 4.1 (A5-6). Since the images are unaligned and vary in myocardium size, both branches are required to first align the prior, before fine-tuning the warped shape. TEDS-Net performs worse without the fine-tuning branch, which is likely due to the lack of flexibility in the bulk transformation, as the deformation fields are generated at a much lower resolution (before being upsampled). This is shown in Figure 4.9 A5, where the myocardium segmentation is lacking detail.

This ablation study shows that our additional modifications are required to produce fields that are topology-preserving in the discrete setting, and these are also found to improve the overall segmentation performance. In the next section, I explore the hyper-parameters that control these modifications.

4.6.2 Hyper-parameter Exploration

4.6.2.1 Gaussian Smoothing

From our ablation studies, Gaussian smoothing was found to play a key role in preventing folding pixels and ultimately stopping topological mistakes. The level of smoothing is controlled by σ within the Gaussian kernel. Here, I investigate the effect that σ (given in pixels) had on the generated fields and the segmentation performance, as shown in Figure 4.10.

The segmentation performance was relatively unaffected across different values of sigma, with no significant differences found in Dice or HD between different values of sigma (measured using ANOVA). This illustrates that increased smoothing

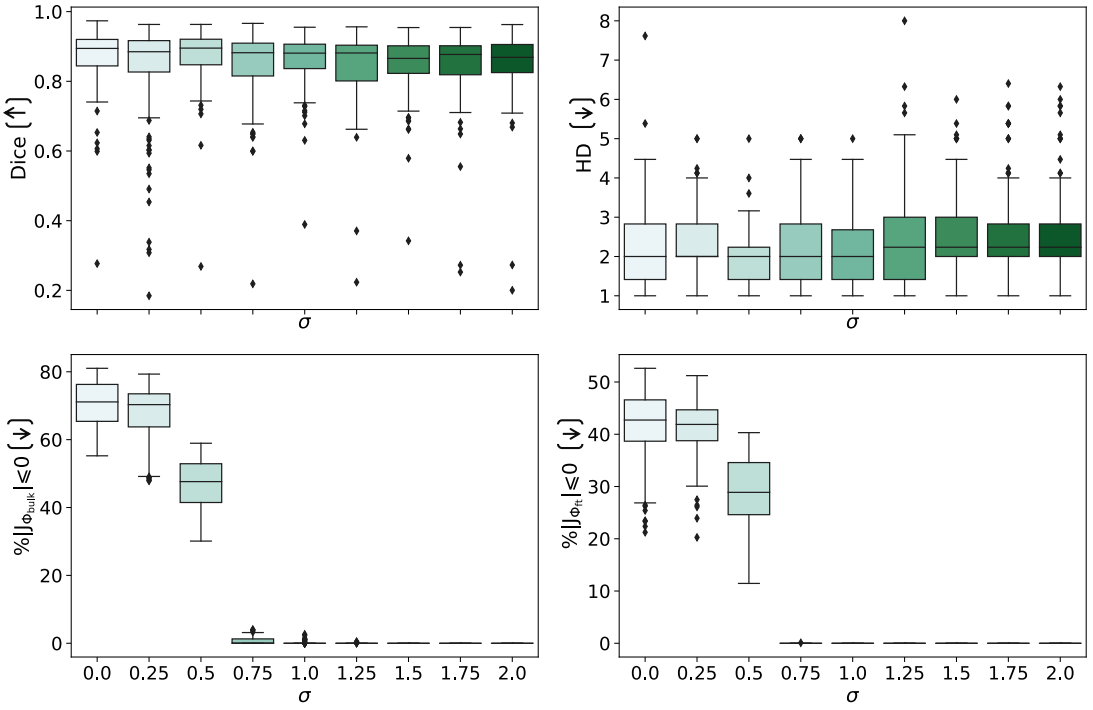


Figure 4.10: The effect that σ within the Gaussian smoothing kernel played on segmentation performance, shown with the average Dice score and Hausdorff distance [mm]. The bottom row shows σ's role in preventing folding pixels, where folding pixels are given by $\% \left|J_{\mathbf{\Phi}}\right| \leq 0$ in the bulk (left) and fine-tuning branch (right).

of the deformation fields does not negatively impact the final warped segmentation. However, a reduced sigma corresponded to the presence of folding pixels, shown by $(\% \left|J_{\mathbf{\Phi}}\right| \leq 0) > 0$. Performing smoothing with a Gaussian kernel where $\sigma \leq 1.25$ resulted in a high percentage of non-positive Jacobian determinants, which often corresponded with topological errors. For instance, when no smoothing was applied $(\sigma = 0)$, folding pixels were present in the fields generated for every input, which resulted in 5% of cases having topology mistakes. The same pattern was found when using $\sigma = 0.25$, where 11% of the segmentations had incorrect topology.

When $\sigma \leq 1.25$, a larger percentage of folding pixels were found in the bulk field compared to the fine-tuning field. The aim of the bulk field is to localise the prior, and therefore, it often contains larger displacements, which can risk large gradients being created between grid points. Therefore, it is likely that the bulk field benefits more from smoothing and thus why more folding pixels are

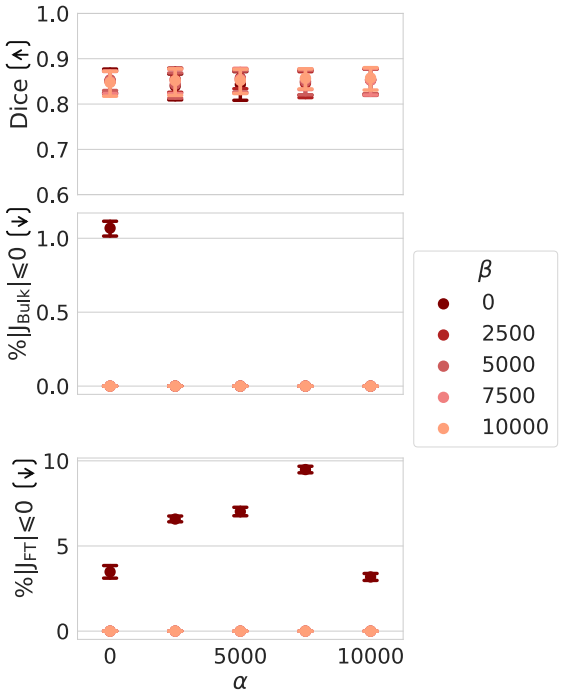


Figure 4.11: Investigating how changing the weighting factors (α and β) within the field regularisation loss function affects the topology-preserving nature and segmentation performance of TEDS-Net. Note that, only $\beta = 0$ contained folding pixels, thus values greater than 0 in the bottom two plots.

present if the smoothing is removed.

As folding pixels were present when no or low smoothing is applied in the composition layers, this shows that additional smoothing modifications are required when using methods such as the Scaling and Squaring approach in the discrete domain to more strictly encourage topology-preservation. High segmentation performance, without the presence of folding pixels, was found when using $\sigma \in [1.25, 2.0]$. *Since* $\sigma = 2.0$ *performed marginally, though not significantly, better than the other values, it was used for the rest of the experiments.*

4.6.2.2 Field Regularisation

As shown in the ablation studies, removing $\mathcal{L}_{\text{Grad}}$ resulted in the largest fraction of topological mistakes. Here, I investigate how adjusting the contribution of the two field regularisation loss weighting parameters, α and β in Equation 4.5 affects the nature of both the fields and the segmentation performance, shown in Figure 4.11. As with varying sigma (shown in Figure 4.10), the segmentation performance remains

relatively unchanged across the range of weighting parameters, indicating that field smoothness does not strongly impact the pixel-wise segmentation performance.

Removing both the field regularisation terms ($\alpha = 0$, $\beta = 0$) results in folding pixels in both the bulk and fine-tuning fields, with a larger percentage found in the fine-tuning fields. This imbalance is likely to be caused by the additional smoothing applied to the bulk field when upsampling the fields. As the bulk field is initially at a lower resolution than the fine-tuning field, to upsample both to the same resolution ($2H \times 2W$), the bulk field undergoes more smoothing from the interpolation, which may result in fewer folded pixels within the field. Removing the fine-tuning field regularisation term ($\beta = 0$) and varying α, resulted in a handful of negative Jacobian determinants within the fine-tuning field but not in the bulk field. This is also likely due to the additional smoothing in the bulk upsampling. *For the rest of this work, I used balanced field regularisation weights of* $\alpha = \beta = 10,000$.

4.6.2.3 Interpolation Methods

Interpolation of the generated fields risks violating key topology-preserving properties, as discussed in Section 4.3.2. Across TEDS-Net, there are three instances where interpolation is applied: (i) field composition layers, (ii) super upsampling, and (iii) when resampling the prior. To explore the effect that the sampling had on topology-preservation, three interpolation methods (namely: linear, bicubic, and nearest neighbour) were used at each of these three instances, as shown in Figure 4.12.

Across the combinations of interpolation methods, the worst image segmentation performance was found when sampling an image using *nearest neighbour* interpolation, regardless of the interpolation methods used for composition or upsampling. When using the nearest neighbour for the final image sampling, the network was unable to learn the myocardium segmentation, resulting in poor Dice performance. The nearest neighbour algorithm is known to lead to poor quality reconstructions, with block artefacts, which is likely the cause of these results.

Sampling the prior shapes using either *linear* or *bicubic* interpolation results in similar segmentation performance across the different composition and upsampling

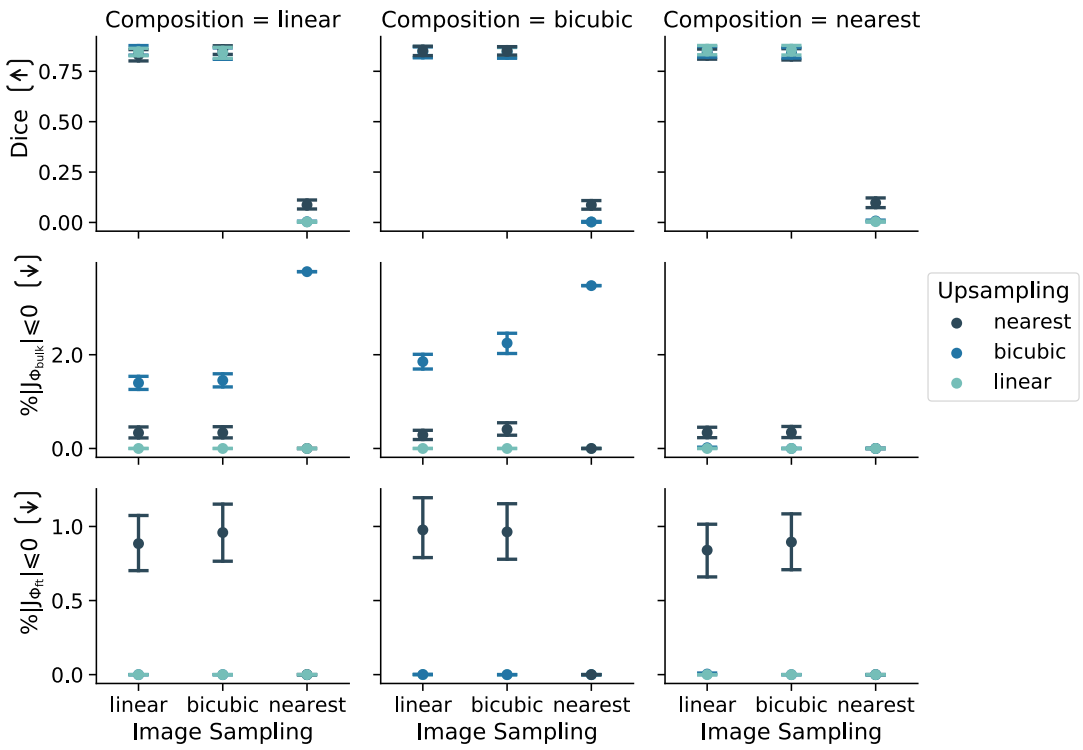


Figure 4.12: Segmentation performance and percentage of folding pixels achieved under different interpolation methods. Interpolation is implemented at three instances in TEDS-Net, namely: upsampling; image sampling; and throughout the composition layers. Only the nearest neighbour upsampling resulted in folding voxels in the fine-tuning branch (shown in the bottom row), with all other methods with $(\% \, |J_{\mathbf{\Phi}_{\text{ft}}}| \leq 0) = 0$.

methods. However, differences are seen in the number of folding pixels present in the fields. For instance, when using the nearest neighbour for upsampling, folding pixels are introduced into both fields. This may be expected as the nature of this interpolation method can create discontinuities within an image. Using bicubic interpolation for the upsampling stage also introduced a small fraction of folding pixels within the bulk displacement field $\mathbf{\Phi}_{\text{Bulk}}$. Upsampling using linear interpolation was the only upsampling technique not to introduce folding pixels into the fields.

Setting the interpolation method for both the composition layer and image sampling to either *linear* or *bicubic* resulted in strong segmentation performance and no folding pixels when used in combination with linear upsampling. *Therefore, for the rest of the experiments I used linear sampling for all interpolations.*

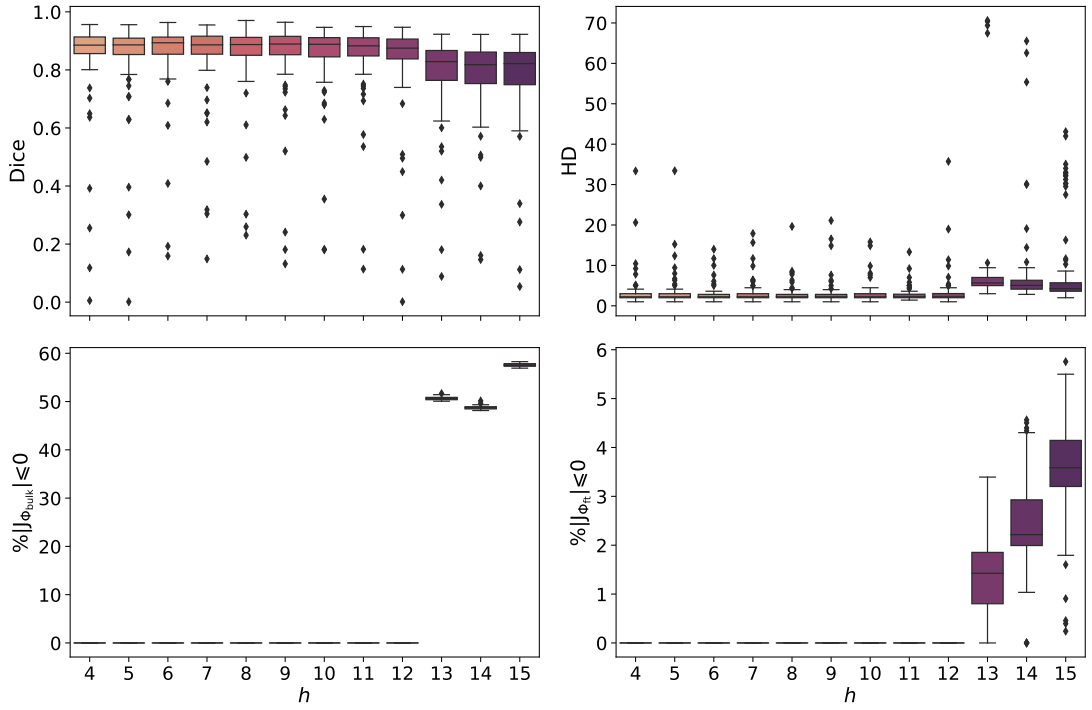


Figure 4.13: The effect that the number of integration layers had on segmentation performance and the topology-preserving nature of the generated fields. The Hausdorff distance (HD) is reported in [mm]

4.6.2.4 Number of interpolation layers

Including Gaussian smoothing between the integration layers played a key role in enforcing topology preservation, as shown in Table 4.1 (A2) and Figure 4.10. Although theoretically, accuracy should increase with the number of composition layers, this requires performing more compositions that can each bring about small violations, due to the discrete nature of the fields [89]. However, limiting the number of composition layers limits the size of deformations, which may negatively affect the segmentation performance. To investigate this compromise, I varied the number of integration layers used in TEDS-Net and measured the segmentation performance and Jacobian determinants of the resulting fields, shown by Figure 4.13.

When using $h \leq 12$ layers, the segmentation performance is stable whilst the Jacobian determinants all remain positive and all topology is preserved. The segmentation performance when using $h = 4$ was on par with the rest of the range, which may be surprising as having few integration layers limits the size of

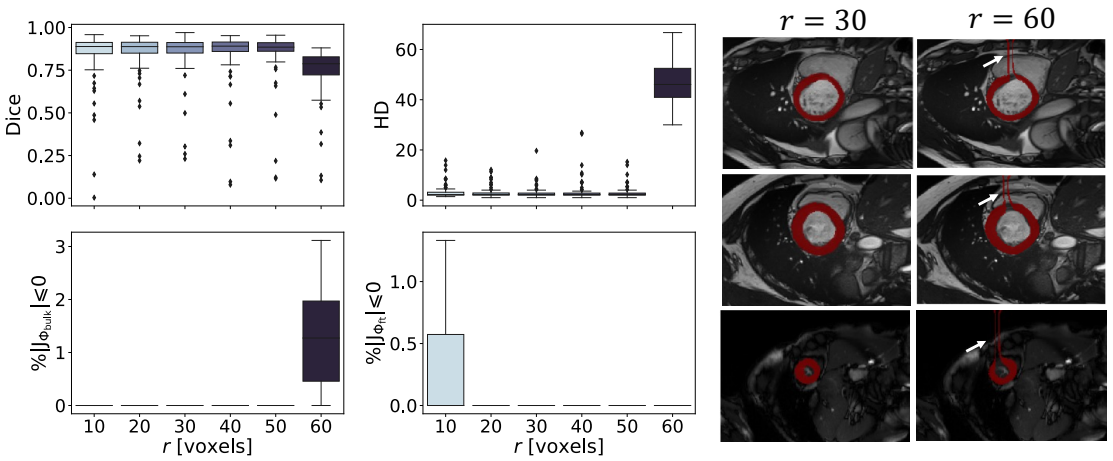


Figure 4.14: The effect that the radius of the prior (r) had on segmentation performance and the nature of the generated fields. The Hausdorff distance (HD) is reported in [mm]. Examples of the predictions from $r = 30$ and $r = 60$ are shown to illustrate the limits of TEDS-Net when $\mathbf{P}$ is much larger ($r = 60$) than the majority of the structures to segment.

the displacements. This is likely due to the dual deformation approach, where the displacements can be performed in two steps, reducing the displacement size required. When the number of layers is increased beyond $h > 12$, non-positive Jacobian determinants emerge. Therefore, whilst Gaussian smoothing has been shown to be essential in the integration layers, there is a limit to the number of compositions that can be used whilst obtaining topology-preserving sampling. *For the rest of the work, I use $h = 8$, as it was the middle of the range of h that did not produce folding voxels.*

4.6.3 Prior exploration

In TEDS-Net the only prerequisite of the prior shape is that it has the same topology of the underlying anatomical structure of interest. Using the dual deformation approach, I have shown the network can segment the myocardium to a high standard using a simple prior shape. To investigate whether the performance depends on the prior shape, I varied the prior size and measured performance, as shown in Figure 4.14. There was little variation in segmentation performance as the radius was increased between 20 and 50 pixels. However, when the radius was set to 10

pixels, a small percentage of the Jacobian determinants of the fine-tuning fields were non-positive, corresponding to 2/100 incorrect topologies. This is likely due to the decreased number of pixels in the prior representing the myocardium and the area enclosed by it, in combination with the additional smoothing modification that restricts the flexibility of the topology-preserving fields. Conversely, at the other extreme ($r = 60$), the Dice and HD dropped in performance. On visual inspection, these predictions contain large artefacts that appear to be stretching the prediction out of view, as shown in Figure 4.14. It is likely when using such a large prior the network cannot compress (or squeeze) the prior to fit the underlying structure whilst obeying our topological modifications. It, therefore, learns to segment the majority of the structure and creates long thin artefacts with the remaining prior pixels, to try and minimise the Dice Loss.

This experiment has shown that TEDS-Net can preserve topology across a range of prior sizes. However, the network begins to struggle when the chosen prior is at either end of the underlying structure size limits and therefore the choice in prior can affect the segmentation performance. *For the rest of the experiments presented here I used a prior with radius* $r = 30$ *pixels, as this is the mid-point between the upper and lower limit of myocardium size of the dataset.*

4.6.4 Baseline Comparison

Table 4.2 and Figure 4.15 display both a quantitative and qualitative comparison of the TEDS-Net myocardium segmentation performance compared to the baselines. The myocardium segmentation performance assessed with both Dice and HD was found to be extremely competitive between U-Net, Swin-Unet, TopoLoss, and TEDS-Net, and none of the five networks was found to significantly outperform the others across all measures, as measured using FSL's PALM function for classical multivariate tests, corrected for multiple subjects, multiple network and multiple metrics [229] (see Appendix B for the full statistics breakdown).

TopoLoss achieved the best segmentation performance, with an average Dice overlap of 88%. As the TopoLoss method used the pre-trained U-Net initially, it

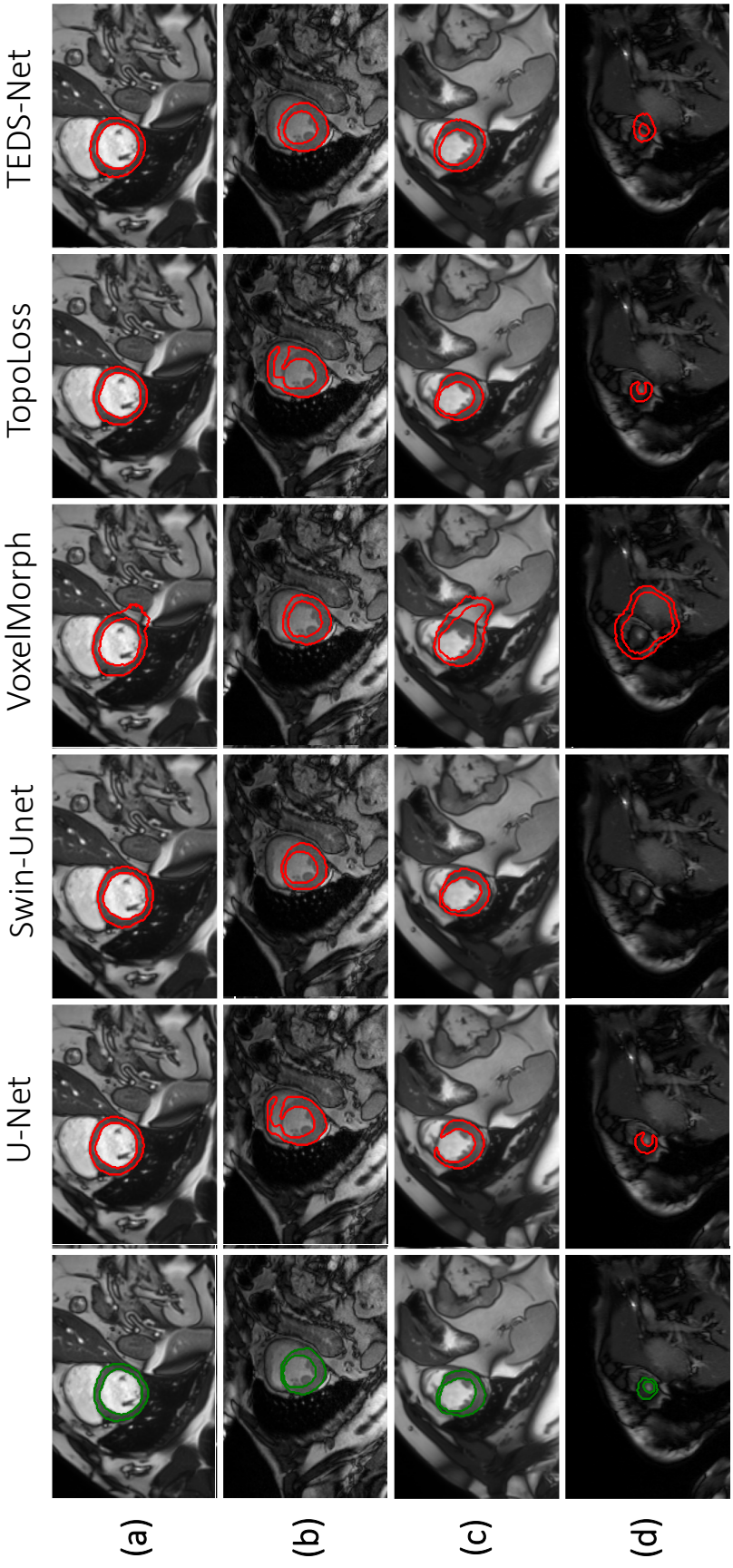


Figure 4.15: Qualitative performance of myocardium segmentation for TEDS-Net and across the baseline comparison methods for four different subjects (a-d). The ground truth labels ($\mathbf{Y}$) and predicted segmentations ($\hat{\mathbf{Y}}$) are shown in green and red, respectively.

Table 4.2: Summary of each network's performance at segmenting the myocardium from the ACDC dataset. The best result (but not significant) is highlighted in bold. Hausdorff distance [mm] (HD), Binary cross entropy (BCE) and average symmetric surface distance (ASSD) are measured in pixels.

Method	Dice ↑	HD ↓	BCE ↓	ASSD ↓	Correct Topology ↑
U-Net	0.87 ± 0.16	2.36 ± 2.36	$\mathbf{0.08 \pm 0.04}$	0.58 ± 0.04	90%
Swin-Unet	0.86 ± 0.17	2.53 ± 2.19	0.09 ± 0.04	0.67 ± 0.78	89%
VoxelMorph	0.75 ± 0.21	4.19 ± 3.59	0.18 ± 0.12	1.17 ± 1.19	98%
TopoLoss	$\mathbf{0.88 \pm 0.15}$	2.33 ± 2.65	$\mathbf{0.08 \pm 0.04}$	$\mathbf{0.53 \pm 0.41}$	95%
TEDS-Net	0.85 ± 0.14	$\mathbf{2.12 \pm 1.43}$	0.11 ± 0.04	0.72 ± 0.50	**100**%

is expected that it should perform on par or better than the U-Net. Fine-tuning with TopoLoss did improve the U-Net's topological performance from 90% to 95%, and was able to correct some of the topological errors found in U-Net (e.g. Figure 4.15c). However, in instances where large breaks in the myocardium were present (e.g. Figure 4.15b and Figure 4.15d) the topology was not fixed, which may have been particularly challenging to correct using TopoLoss's patch-based method. Moreover, TopoLoss also required additional post-hoc morphological operations, however, when the same operations were applied to the different methods the number of topological errors increased, illustrating that a single, complete method is preferred. The myocardium segmentations from Swin-Unet contained the highest number of topology errors. This may be surprising as it is thought that transformer networks learn global structures better than traditional CNNs, as they provide better long-range contextual relationships. TEDS-Net performed worse, in terms of Dice, for cardiac slices at the apex of the heart, where the myocardium is smallest as shown in Appendix C. This may be due to difficulties in reducing the size of the prior whilst retaining smooth and topology-preserving fields.

VoxelMorph's implementation was found to produce small topological errors in two of the test images. However, the corresponding predicted fields contained no folding pixels. At closer inspections, these detected "holes" within the myocardium would not be classified as a "hole" if defined using 8-connectivity instead of 4-connectivity. TEDS-Net outperformed VoxelMorph in both Dice and HD, which may be a result of the dual branch approach used in TEDS-Net, where the network first learns to localise the myocardium before fine-tuning the segmentation and/or

Table 4.3: The training time per epoch and batch, compared to the number of parameters used for each method

Method	**seconds /batch**	**minutes /epoch**	**# Parameters** $[\times 10^5]$
U-Net	0.01	0.22	5.2
Swin-Unet	0.08	1.53	413
VoxelMorph	0.01	0.25	1.0
TopoLoss	0.9	18.35	5.2
TEDS-Net	0.03	0.55	7.1

down to the increased number of parameters as shown in Table 4.3. As can be seen in Figure 4.15, it appears that VoxelMorph sometimes struggles to transform the prior to the correct location and size. Moreover, it is important to note that VoxelMorph has been optimised for image registration, not segmentation, and this comparison was mainly included to address the topology-preserving properties of the predicted fields and the flexibility of the predicted fields.

TEDS-Net was the only method to achieve 100% topology preservation. This achievement coincides with 0% non-positive Jacobian determinants in both the bulk and fine-tuning discrete fields, $\mathbf{\Phi}_{\text{bulk}}$ and $\mathbf{\Phi}_{\text{FT}}$, implying they are topology-preserving.

Although TopoLoss achieved promising segmentation performance, it also required the longest run times, as shown in Table 4.3. In comparison, U-Net, VoxelMorph, and TEDS-Net are two orders of magnitude faster, training in 0.22, 0.25 and 0.55 minutes per epoch, respectively. TEDS-Net's training time is double that of VoxelMorph and U-Net, which is likely to be due to the increased number of parameters in the network due to the two decoder branches and the time taken during the interpolation. Moreover, both TopoLoss and Swin-Unet required a pre-trained network, which was not accounted for in these run times. The Swin-Unet architecture required 415×10^5 parameters, $60\times$ greater than TEDS-Net, and this drastic difference is paired with a longer time per epoch. TEDS-Net's segmentation performance is competitive to state-of-the-art baselines whilst outperforming them on topology performance, alongside having a relatively computationally undemanding training.

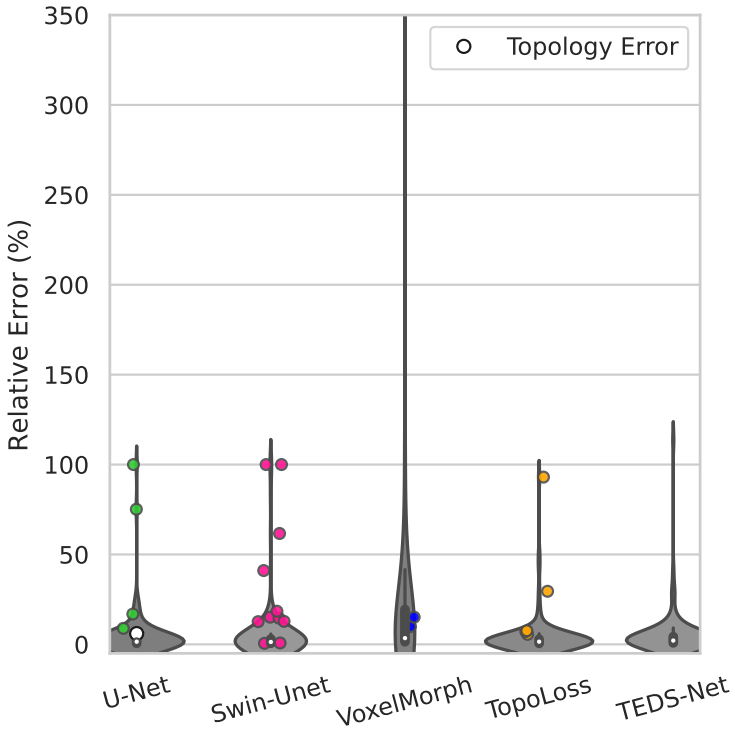


Figure 4.16: The relative error of automated myocardium perimeter measurements across the five methods. This shows the ratio between the absolute error and the known perimeter value, expressed as a percentage of the known perimeter value. The average myocardium perimeter was found to be approximately 250 voxels. The myocardium perimeter from each prediction was compared to the perimeter of the labels. The mean for each network is shown in white, and predictions with incorrect topology are highlighted.

4.6.5 Myocardium Perimeter

Segmentations of the heart's myocardium are often required to automate measures of the perimeter. Therefore, to assess whether a segmentation method is fit for the purpose, the predicted myocardium perimeter should be equivalent to that of the label. To evaluate this, I automatically measured the myocardium perimeter generated by each of the networks, using 4-neighbourhood connectivity, using code from [230]. The perimeter measure here is the total of the inner and outer boundary of the myocardium, i.e. the contours around the structure. From these measurements, I calculated the relative perimeter error for each heart slice, given by Equation 2.4.

Figure 4.16 shows the difference between the measured myocardium perimeter

and its known value across the different networks. For U-Net, TopoLoss, Swin-Unet and TEDS-Net, the majority of the measured perimeters matched closely to the paired labels, shown with absolute errors close to 0, and only a small number of outliers for each network. The segmentations with topological errors are indicated and the measured perimeter for these cases are often found to have a large relative error, showing that topological errors can hinder further measurements. Despite TEDS-Net's having achieved the correct topology across all myocardium segmentations, a small fraction of segmentations returned poor perimeter measurements. These cases were from MRI slices near the apex (e.g. the base of the heart), where the myocardium is smaller and can be more difficult to segment (shown in Figure 4.15d) and were consistently poor across the other methods, as shown in Appendix C.

4.6.6 Limitations

The aim of TEDS-Net is to segment anatomical structures by using prior knowledge of shape, to ensure the final segmentation has the correct topology. From this evaluation, I show that TEDS-Net can be tuned to achieve 100% topology accuracy whilst having a competitive segmentation performance. However, it is not without its limitations.

Firstly, TEDS-Net requires prior knowledge of the anatomy of interest, and therefore, the model is only suitable for structures with known topology. This excludes a range of medical segmentation problems, including tumour segmentation, where the number of tumours and topology is likely unknown. Moreover, care must also be taken when segmenting 3D structures from 2D slices, as the topology may vary depending on the plane and slice location. In this work, I segmented the myocardium, which in 3D has a topology that is equivalent to that of a sphere ($b_0 = 1, b_1 = 0, b_2 = 1$). When slices are taken from the short axis, in the majority of the slices the myocardium has the same topology as a ring; $b_0 = 1, b_1 = 1$, however, this may vary, and therefore a careful selection of slices must be performed.

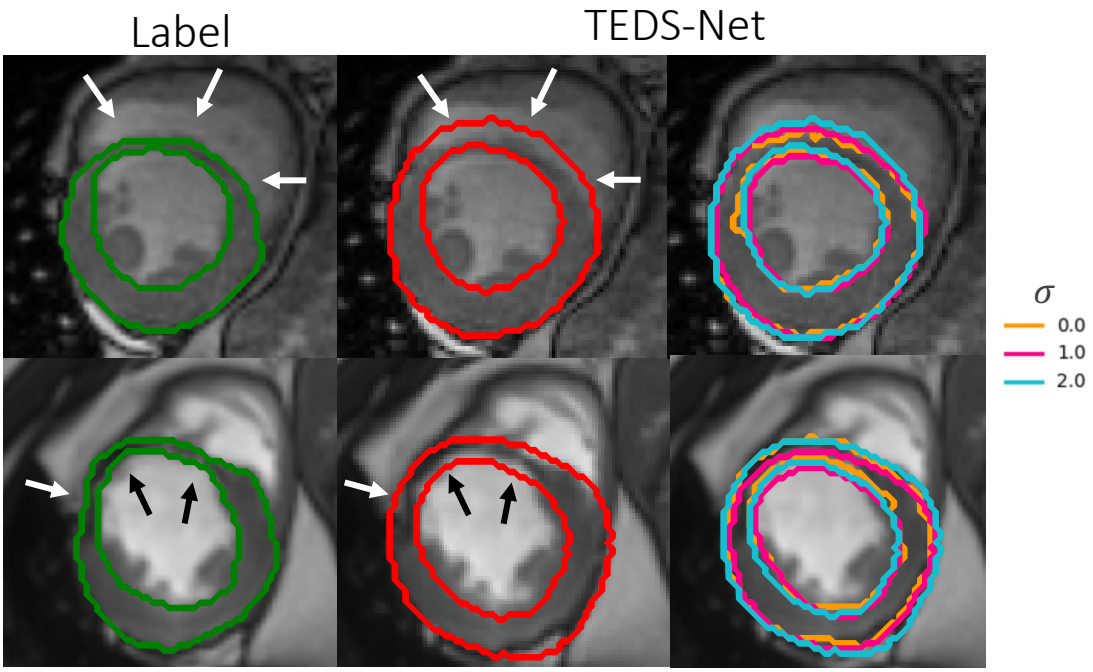


Figure 4.17: Qualitative examples to show TEDS-Net over-segmenting in regions of thin boundaries. The ground truth is shown in green, TEDS-Net predictions in red and examples at different σ's ($\sigma = 0$ is no smoothing). Arrows have been added in the same position to highlight the segmentations mistakes at the thinnest regions of the label.

Secondly, in order to achieve the generation of topology-preserving fields, additional smoothing modifications were required to help overcome the limitations of discrete sampling and storage. However, smoothing within the fields can limit the detail within the deformations, which can result in bulky segmentations. An example of this is shown in Figure 4.17, where TEDS-Net over segments the myocardium in regions where it is thin, and therefore, does not accurately represent the true thickness of the structure.

In our ablation studies, I found that Gaussian smoothing between each composition layer was crucial in removing folding pixels. However, it is likely that it also smooths over some detail in the fields, resulting in thicker segmentations. To investigate this theory, I compared the myocardium predictions at different σ. In Figure 4.17, the two qualitative examples show that applying no smoothing $\sigma = 0$ or smoothing at different σ's does vary the myocardium segmentations. At the thinnest region of the myocardium (indicated by the arrows), the prediction generated using

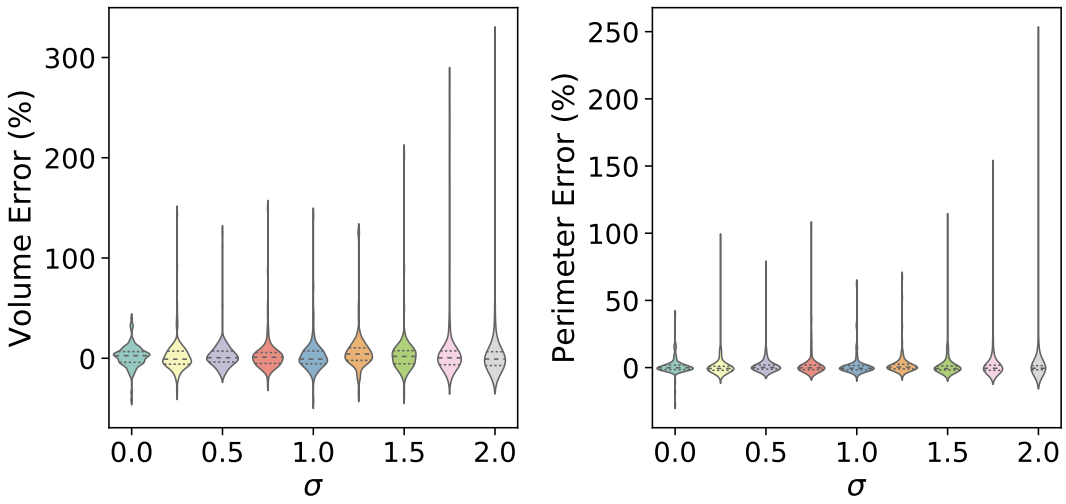


Figure 4.18: The relative error of the volume and perimeter relative errors, measured between the ground truth labels and the predictions from TEDS-Net trained and evaluated with different values of σ. The errors are expressed as a percentage of the ground truth value.

no smoothing ($\sigma = 0$) appears marginally thinner than that at $\sigma = 2$, although still not following the contours of the myocardium label. To quantitatively investigate this, I measured the relative volume and perimeter errors between the ground truth label and the predictions at each σ, as shown in Figure 4.18. If increased smoothing resulted in thicker segmentations, I would expect a larger volume error. Although the measured volume and perimeter error were greater at $\sigma = 2$ than $\sigma = 0$, there was no significant difference between the measures when using different σ values, as measured by an ANOVA statistical test.

4.7 Conclusion

In this chapter, I presented TEDS-Net, a segmentation network aimed to delineate medical structures whilst preserving the known anatomical topology. I evaluated TEDS-Net for the task for 2D myocardium segmentation and it achieved 100% topology accuracy, with pixel-wise performance competitive to the start of the art baselines. Moreover, I showed how traditional evaluation metrics, e.g. Dice Score, often do not detect problematic segmentations and therefore additional

measures, such as topology, should be routinely included to more accurately assess the segmentation quality.

To preserve topology, topology-preserving fields were used to deform a prior shape that had the correct topological characteristics. Unlike previous work, TEDS-Net field generation does not rely on continuous theorems to guarantee topology preservation. Instead, I introduced additional modifications to try and overcome the discrete limitations brought about by sampling and storage in the discrete domain. Through a series of ablation studies and hyper-parameters sweeps I showed that without these modifications, folding voxels emerge in the fields and therefore guaranteed topology preservation is lost despite following continuous mathematical principles (e.g. the Scaling and Squaring approach).

However, additional smoothing can negatively impact the segmentation performance, as some fine detail can be lost. I showed how Gaussian smoothing can over-segment the myocardium in particularly narrow parts, however, no statistically significant differences were found between volume and perimeter relative error across σ.

Through extensive hyper-parameter exploration, I investigated the range of values that key variables could take whilst achieving 0 folding voxels. In this chapter, I have shown how TEDS-Net can be used to preserve topology when segmenting the 2D myocardium. For this task, I have shown the range of hyper-parameters that can be used to prevent folding voxels. However, it is unknown whether these ranges are dataset or task-dependent. In the next chapter, I apply TEDS-Net to a range of different datasets to investigate its performance in other tasks and the appropriate hyper-parameter settings for them.

5 Application of TEDS-Net

5.1 Overview

Following the previous chapter in which I proposed TEDS-Net, this chapter applies TEDS-Net to new and more challenging segmentation tasks, whilst making minimal changes to the original architecture and hyperparameter values.

Early versions of the work presented here can be found in:

> **Madeleine K Wyburd**, Nicola Dinsdale, Mark Jenkinson, and Ana IL Namburete. "Anatomically Plausible Segmentations: Explicitly Preserving Topology Through Prior Deformations." *In Revision*: submitted to MedIA (2022)

5.2 Introduction

In the previous chapter, I proposed TEDS-Net: Topology Encouraging Deformation Segmentation Network; a network that encourages topologically correct segmentations by deforming a prior that has the correct topological properties. I showed how TEDS-Net can achieve 100% topological accuracy whilst remaining competitive to SOTA baselines on traditionally used metrics, such as Dice and Hausdorff distance, for the task of myocardium segmentation from 2D MRI slices. However, this investigation was performed on a single dataset. Moreover, the topology of the myocardium is relatively simple, the dataset is a single class, 2D, and with consistent topology across the dataset. Therefore, an investigation is required to see if TEDS-Net is suitable for other datasets and more complex segmentation tasks.

In real-life applications, segmentation networks are applied to data of different dimensions (e.g. 2D and 3D), different modalities (e.g. CT, MRI), multi-structure, and to structures with varying topology. To assess the ability of TEDS-Net to segment in such scenarios I applied it to a range of datasets and segmentation tasks.

In this section I show how TEDS-Net can successfully segment:

- Data from a range of modalities, including natural images (Section 5.3) and CT (Section 5.5)
- Multi-class, unaligned datasets (Section 5.4)
- Multiple topologies within the same training cycle (Section 5.3)
- Medical datasets in both 2D and 3D (Section 5.5)

This chapter is divided into three sections, where each new TEDS-Net application is contained within its own section, which includes dataset description, methodology, results and discussion.

5.3 Multi-Prior Segmentation

So far in this thesis, TEDS-Net has been applied to the task of 2D myocardium segmentation, where the topology of the myocardium was the same in each MRI slice. Thus the same prior was used for each slice. However, as previously discussed, the topology of the myocardium is not always the same across the heart volume, and therefore the same prior would not be suitable if the slices were randomly chosen rather than selectively picked. Moreover, segmentation networks are often required to segment different anatomy from each input, which may vary in topology.

In this section, I evaluate the performance of TEDS-Net on a dataset made up of a range of different topologies, to investigate whether it can still achieve promising performance when a variety of priors are used, each describing the different topologies within the dataset. To do this, I performed digit segmentation from the MNIST dataset [27], where the digits between 0-9 can be described by three different topologies. Moreover, as there were common topologies amongst the digits, I investigated whether TEDS-Net could successfully segment drastically varying shapes but with the same topology (e.g. "5" and "1") using the same simple prior shape.

5.3.1 Methods

5.3.1.1 Dataset

To perform digit segmentation I used the MNIST dataset [27], which is made up of 70,000 hand-drawn digits from 0 to 9. Traditionally it is used for classification tasks, so to convert it into a segmentation problem the grayscale images were used as the input $\mathbf{X} \in \mathbb{R}^{28\times28}$ and a binarised version was used for the ground-truth $\mathbf{Y} \in \mathbb{R}^{28\times28}$, by applying a threshold at 0.5. To increase the difficulty of this task, Fourier-based image augmentation was used to distort the appearance of the digits, by randomly removing lines in the Fourier space representation of the image. This augmentation technique can dramatically change the appearance of the image in very different styles depending on the locations of the lines. Therefore, it generates a large variation in image appearance, making this simple segmenting

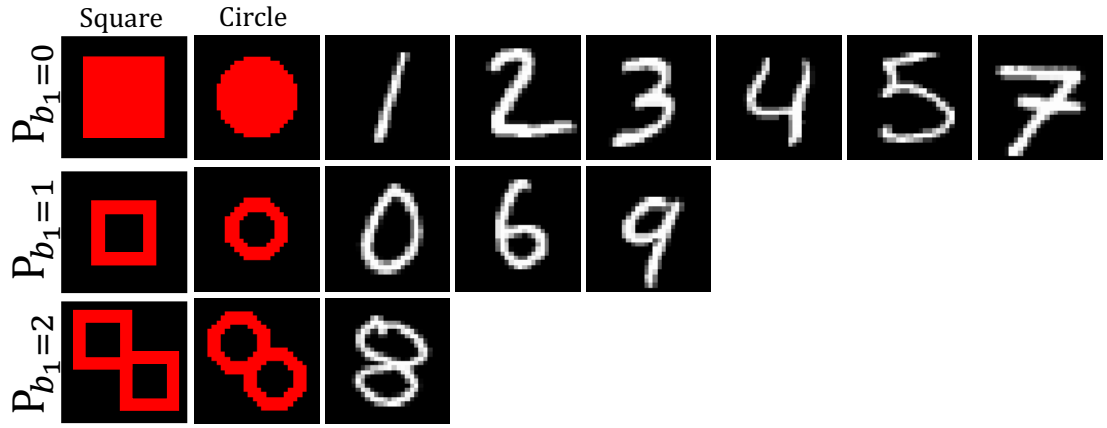


Figure 5.1: Examples of MNIST digits and, in the first two columns the two sets of priors investigated. All digits with the same topology are segmented using the same rough prior.

task more challenging. The dataset was split into 10,000 for testing and 60,000 for training, where 10% of the training set samples were used for the validation set. The dataset was split evenly across digit values.

The MNIST dataset digits can be classed as three different topologies. The topologies of digits *1,2,3,4,5* and *7*, can be defined by a single component ($b_0 = 1$) with no holes ($b_1 = 0$) and therefore have a Betti number: $B = \{b_0 = 1, b_1 = 0\}$. Whereas, *0, 6* and *9* have a single hole: $B = \{b_0 = 1, b_1 = 1\}$ and 8 has two holes: $B = \{b_0 = 1, b_1 = 2\}$, as shown in Figure 5.1. The same prior $\mathbf{P} \in \mathbb{R}^{28\times28}$ was used for each topology type, despite the digits within that topology class looking drastically different.

Two sets of prior shapes were investigated; circles and squares. For each shape, a set of priors containing zero, one, and two holes were generated and referred to as $\mathbf{P}_{b_1=0}, \mathbf{P}_{b_1=1}, \mathbf{P}_{b_1=2}$, as shown in Figure 5.1. As the MNIST digits fill the majority of the image, the priors were designed to be a similar size. The outer shape in $\mathbf{P}_{b_1=0}$ and $\mathbf{P}_{b_1=1}$ was designed to have a diameter or side length of 18, whereas $\mathbf{P}_{b_1=2}$ had a length of 14, in order to fit two shapes next to each other. The thickness of the shape for $\mathbf{P}_{b_1=1}$ and $\mathbf{P}_{b_1=2}$ was set to three, to mimic the rough line thickness of the digits.

5.3.1.2 Network Implementation

For digit segmentation, the encoder-decoder network was set up with $l = 2$ layers and with $f = 12$ initial feature maps, smaller than the original TEDS-Net architecture (described in Section 4.4.2) due to the smaller image size. As the position and size of the digits across the dataset were consistent, the bulk deformation branch was also removed and only a single deformation field was generated from the top layer ($l_{\text{ft}} = 1$) of the decoder branch.

TEDS-Net was trained for 100 epochs, using the Adam optimiser with a learning rate of 0.0001 and a batch size of 30. The number of compositions was set to $h = 8$, and Gaussian smoothing was performed with $\sigma = 2$ and $k = 3$, as previously implemented. As there was only a single field generated, for training I used a loss function: $\mathcal{L} = \mathcal{L}_{\text{Dice}}(\mathbf{Y}, \hat{\mathbf{Y}}) + \alpha\mathcal{L}_{\text{Grad}}(\mathbf{\Phi}_{\text{FT}})$, where $\alpha = 150$, which was chosen by observing the ratio between $\mathcal{L}_{\text{Grad}}$ and $\mathcal{L}_{\text{Dice}}$ within the loss curves.

To train TEDS-Net, all digits were used within the same training cycle and were paired to their corresponding prior depending on the value of the digit. Therefore, to use TEDS-Net for different topology types in this way, prior knowledge of topology for each input is required. It is important to note that, due to natural variation in handwriting, not all the images aligned with the assigned prior topology, the impact of this is investigated in the next section.

5.3.2 Results

Figure 5.2 shows the Dice overlap for each digit segmentation from both the circle and square sets of priors (shown in Figure 5.1). Both sets were able to segment each digit to a high standard, achieving an average dice overlap of 0.90 ± 0.04 and 0.88 ± 0.05 for the circle and square sets, respectively. The priors used throughout this experiment were basic shapes with the only prerequisite that they had equivalent topology to their paired digit. For instance, the same prior was used for *1,2,3,4,5,7* despite the digits looking drastically different. Despite this, TEDS-Net was able to segment each accurately, with the majority of the segmentation achieving a Dice score $> 80\%$ across all digits, illustrating the flexibility of the network and that

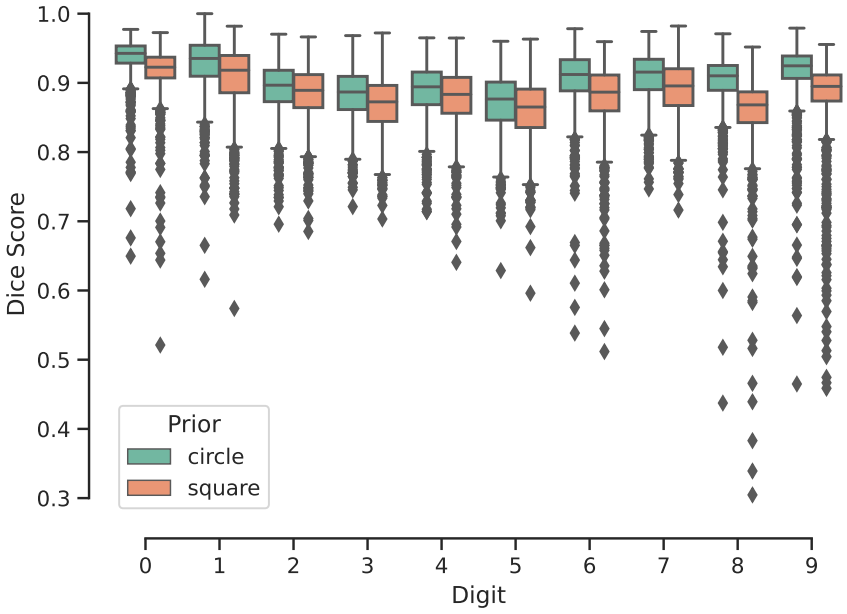


Figure 5.2: Segmentation performance across the digits when using the two sets of priors. The box plots show the mean and standard deviation for each digit, with outliers highlighted.

priors of substantially different shapes and orientations (but the same topology) can be used to segment structures to a high standard.

The segmentations obtained using the circle set of priors were found to significantly outperform those generated from the square priors ($p < 0.01$), as measured using an ANOVA test. Interestingly, from Figure 5.2, the largest differences in the mean between the two groups are observed for digits $0, 6, 8, 9$, which all contain holes and are relatively curvy shapes. This may be due to the sharp edges within the square set of priors, which can be difficult to smooth over whilst trying to prevent folding voxels. *As the model using the circle set of priors was found to significantly outperform the one using the square priors, it was used for the rest of the experiments.*

Qualitative examples of TEDS-Net segmentation performance when using the circle set of priors, across all digits, are shown in Figure 5.3. The segmentation performance between digits was significantly different ($p < 0.01$, as measured using an ANOVA multiple comparison tests). Figure 5.4 shows the average Dice score for each digit and statistical significance for each pair of digits, segmented using the

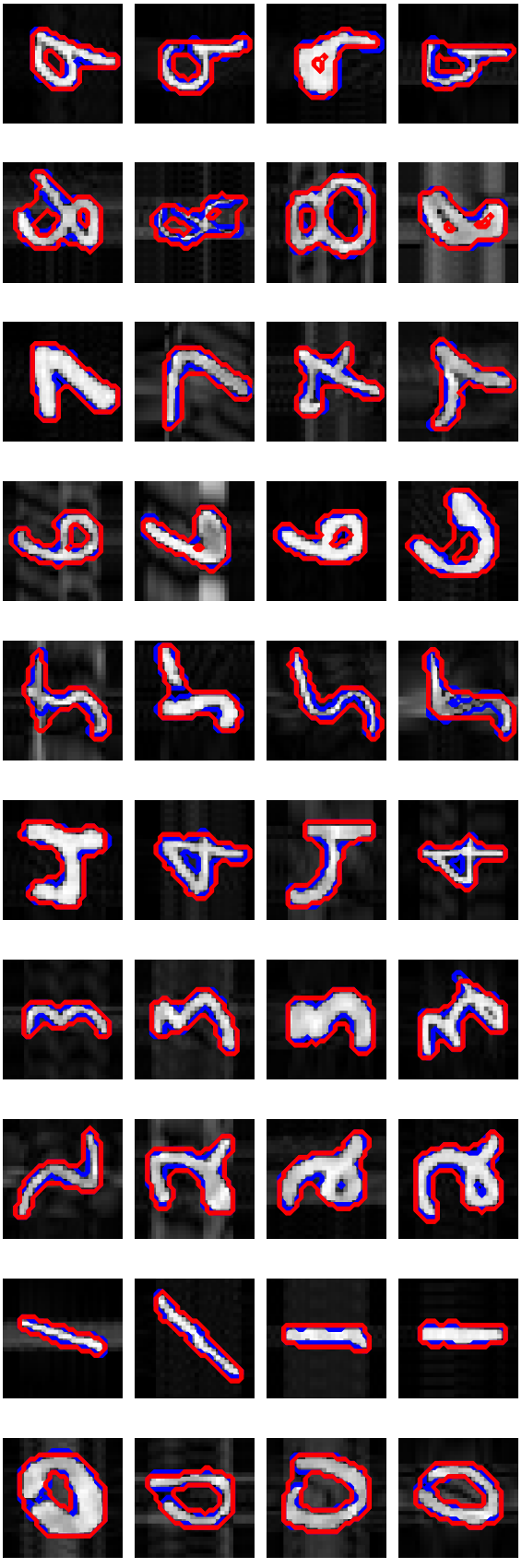


Figure 5.3: MNIST digit segmentation using TEDS-Net. The blue lines show the labels ($\mathbf{Y}$), whilst the red show TEDS-Net predictions ($\hat{\mathbf{Y}}$).

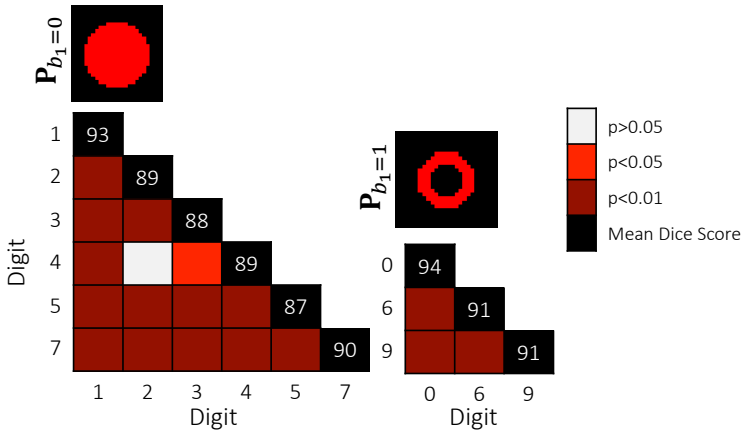


Figure 5.4: A statistical comparison of segmentation performance between digits segmented with the same prior. Significance (measured using an unpaired t-test and with Bonferroni correction [231]) and average Dice (shown on the diagonal) are presented. Despite the same prior being used, the performance of the majority of the digits was found to differ significantly. The digit "8" was the only one to be segmented with $\mathbf{P}_{b_2=2}$, it achieved an average dice of 91%.

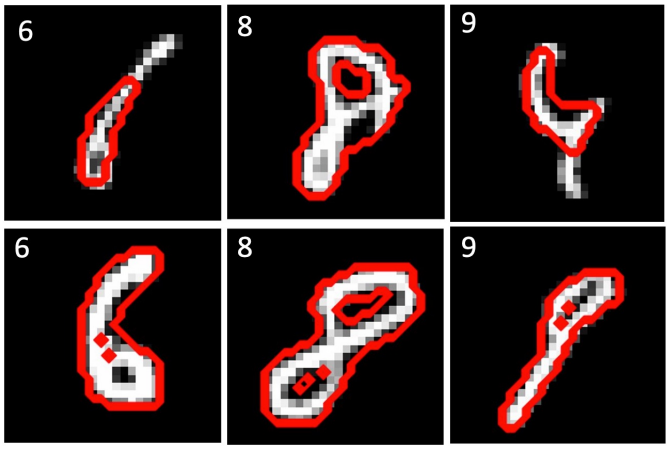


Figure 5.5: Examples of where TEDS-Net has not preserved the topology of the prior, these examples were randomly chosen.

same prior. Digits with simpler shapes (e.g. "1" and "0") were found to perform significantly better (measured using an unpaired t-test, corrected with Bonferroni correction [231]) than the more complex digits (e.g. "5" and "9").

In total, 99.53% of the 10,000 segmented structures retained the same topology as its prior. However, 71.43% of the generated fields ($\mathbf{\Phi}$) contained folding voxels, with the average fraction of folding voxels found to be 4.6% (equivalent to 36 voxels).

All the topological errors were found to be in digits $6, 8, 9$, with 33 cases of

incorrect topology for 8's, 11 for 6's and 3 for 9's. Figure 5.5 shows randomly selected examples of the topological errors from each digit. In the top row TEDS-Net has removed a hole from the prior, whereas in the bottom row, the prior hole has been separated into two holes. These topology violations could be caused by the folding voxels, however, other sources such as the downsampling of the images and thresholding of the final prediction could also break the topology. Interestingly, in the top row where a prior hole has collapsed, the underlying image is also missing a hole.

5.3.3 Discussion

As seen in Figure 5.3, the topology of the prior assigned to each digit was not always equivalent to the handwritten digit's labels due to the natural variation in handwriting, with 24% of the MNIST dataset found to differ from the topology I assigned. Despite this, TEDS-Net predictions were found to preserve the prior topology in the vast majority of the cases (99.53%). Further, Figure 5.3 shows that in some cases the Dice performance was negatively impacted to ensure topology preservation, despite the network being partially optimised to minimise Dice Loss, demonstrating how strictly topology preservation can be enforced within TEDS-Net. For instance, in the hand-drawn examples of *"0"*, the top of the loop was commonly left open, changing its topology. However, in TEDS-Net segmentations, the loop was completed, following the contours of the shape. This property of TEDS-Net could be desirable in medical segmentations when shadows or artefacts are obstructing parts of the structure of interest. A similar pattern can be seen for digits *"8"* and *"9"*, and in "5", where the labels often show two disconnected components.

Natural variation in handwriting is particularly seen in digits *"2"* and "*4*", where the digit can either be drawn as a single component with no hole or a component with a 2D-hole within it (e.g. *4* or 4). Therefore, depending on an individual's writing style, their topology can either be equivalent to $\mathbf{P}_{b_1=0}$ or $\mathbf{P}_{b_1=1}$. Initially, these digits were chosen to be segmented using $\mathbf{P}_{b_1=0}$, which often resulted in TEDS-Net surrounding the whole structure, and ignoring the hole, as shown in Figure 5.3. To investigate the alternative topological profile of these digits and to see if the fields

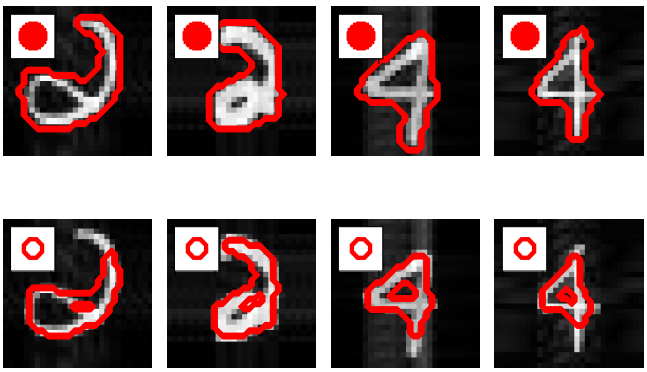


Figure 5.6: Results obtained by replacing the prior shape used for 2 and 4 with $P_{b_1=1}$ at inference time. The top left corner of each image shows the prior used to produce the segmentation (shown in red).

still preserve topology when using a different prior, at inference time I replaced the prior $\mathbf{P}_{b_1=0}$ with $\mathbf{P}_{b_1=1}$, and the resulting segmentations are shown Figure 5.6. When changing the prior, the segmentation performance decreases, which is expected as the network has not learnt the fields based on the new prior. However, the predictions contain the equivalent topology to the new given prior. This result further illustrates that the deformations are topology-preserving. However, as shown in Figure 5.3, there is a drop in segmentation performance. Therefore, this experiment shows that different topologies can be preserved using the same deformation fields but it is not advised to change priors at inference time.

This experiment demonstrates TEDS-Net's ability to segment structures of multiple topologies to a high standard within the same training cycle by using paired priors. Moreover, it also shows how strictly our network can enforce topology preservation and that this property often overrides the underlying topology seen in an image. This behaviour may be desirable in medical imaging, where shadows or artefacts obstruct regions of anatomy changing its observed topology. Further, it shows that structures can be segmented to a high standard when using very simple priors, with a vastly different shape compared to the desired structure.

This experiment has also highlighted some new limitations of TEDS-Net. Firstly, folding voxels were present in the majority of the generated fields, despite the additional smoothing modifications. This is likely to be due to the small size of the image paired with the large structure and prior size, which makes estimating flexible deformations, at the same time as preventing folding voxels, extremely challenging. This effect might also have been exacerbated as the digits ($\mathbf{Y}$) look vastly different from their paired rough priors ($\mathbf{P}$), requiring even more flexible fields. Therefore, it is likely that the network compromised between the field regularisation and the Dice loss in the optimiser. Further, this might suggest that there is a minimum resolution that is required so that the fields can achieve 100% topology preservation using the modifications built into TEDS-Net. Secondly, I found the performance was reduced by a small, but significant, amount when using the square priors, showing that the choice of prior shape can be important in TEDS-Net segmentation performance. Finally, in this experiment, I paired different priors to each input, which required prior knowledge. Therefore, this approach of using our method is only suited to segmentation where the topology is always known.

5.4 Multi-Structure Segmentation

So far, TEDS-Net has been shown to successfully segment structures of different topologies in 2D images. However, I have only looked at single-structure segmentation and have yet to investigate its ability for multi-structure segmentation. When performing multi-structure segmentation, the topology of the individual structures is still important, however, an additional measure of overall *scene* topology can also be considered, to assess the shape or position of each structure relative to one another. Scene topology can provide a key insight into anatomical plausibility, as some structures have a well-established scene from certain medical imaging views. For example, from the short-axis view of cardiac imaging, the myocardium is known to separate both the right and left ventricles, acting as a scaffolding between the two ventricles. Therefore, when segmenting this scene, the myocardium label should always be adjacent to both the right and left ventricles, and the ventricles should

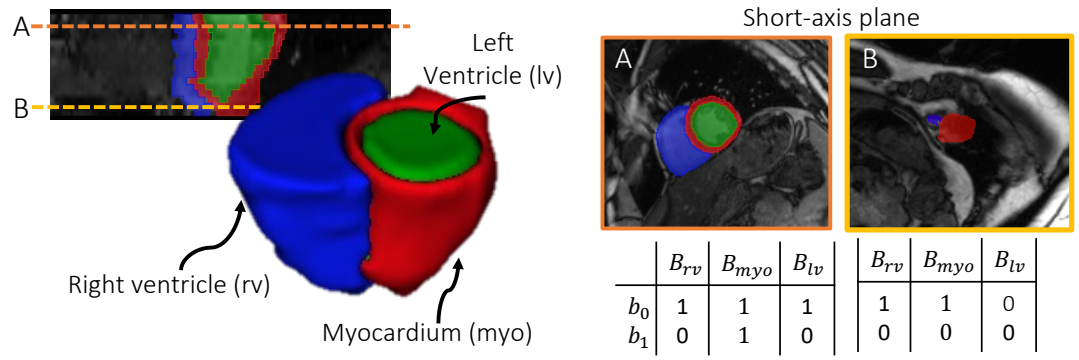


Figure 5.7: Segmentation of the heart shown in 3D and at two 2D cross sections in the short-axis. Depending on the location of the short-axis slice, the topology of each structure can vary. For TEDS-Net the topology must be known and therefore, only slices with the same topology as shown in Slice A were used. In fact, the myocardium extends around the right ventricle as well. However, when viewed in the short axis, the myocardium around the right ventricle is much thinner than around the left, and therefore is often unlabelled, like in the ACDC dataset. If the myocardium around the right ventricle was included, the topology would change.

never touch. Therefore, multi-structure segmentation offers an interesting task for TEDS-Net to see if it is able to preserve the topology of multiple individual structures as well as the overall scene topology.

Using prior knowledge of the topology of the scene and individual structures, a prior shape can be designed to follow both types of pre-defined topology. Then, by using the same learnt topology-preserving field, both the individual and scene topology should be preserved. In this section, I apply TEDS-Net to the task of multi-structure cardiac segmentation.

5.4.1 Methods

5.4.1.1 Dataset

To perform multi-structure segmentation, I used the same dataset and augmentations described in Section 4.5.1 but now segmenting the whole heart structure as shown in a 2D short-axis slice of the heart, which in this view is made up of the myocardium (MYO), right ventricle (RV) and left ventricle (LV). A cross-section of the heart is shown in Figure 5.7, displaying how different short-axis slices of the heart can contain the same structures but with different topologies. As TEDS-Net requires

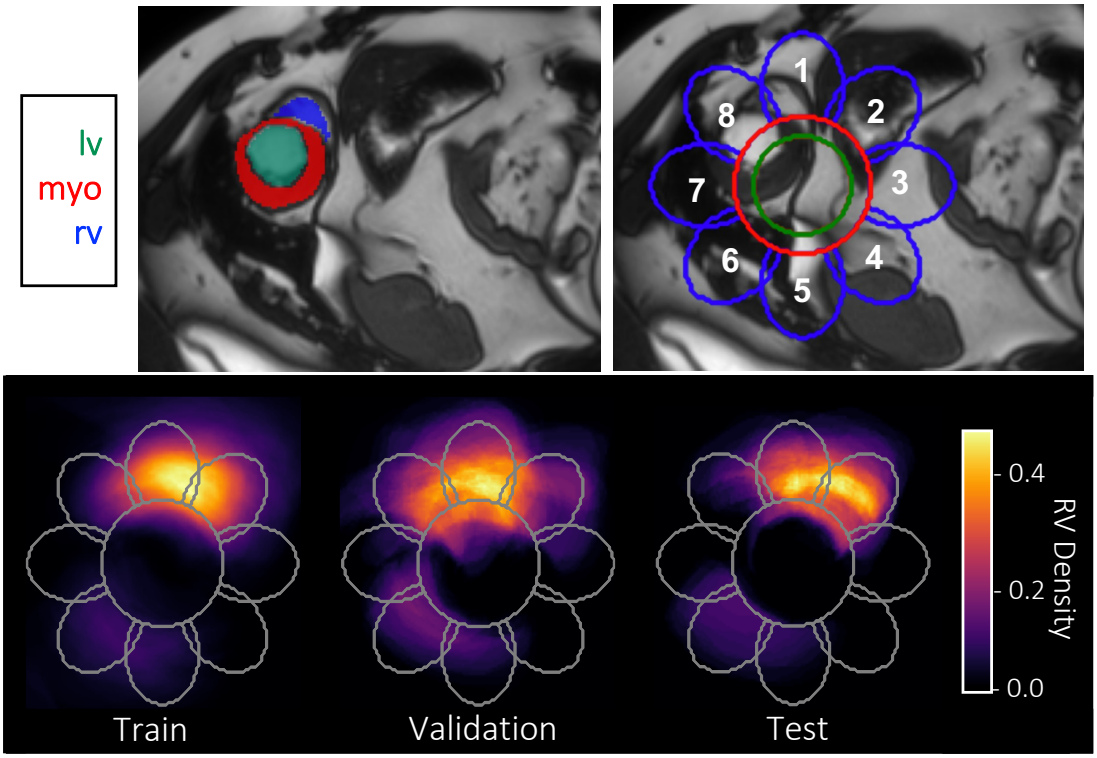


Figure 5.8: An overview of the different priors used for multi-structure cardiac segmentation. The position of the right ventricle (RV) was varied around the myocardium (MYO) as its position was found to be inconsistent across the train, validation, and test sets. The heat map shows the right ventricle (RV) density across each set with the outline of the different RV priors (P_{rv}), to show where the majority of the labels were located across the dataset.

prior knowledge of each structure's topology, I only selected slices from our ACDC dataset where the topology of each structure was as follows: $B_{\text{myo}} = \{b_0 = 1, b_1 = 1\}$, $B_{\text{rv}} = \{b_0 = 1, b_1 = 0\}$ and $B_{\text{lv}} = \{b_0 = 1, b_1 = 0\}$ assessed by measuring the label's topology, as shown in Slice A in Figure 5.7. It should be noted that the myocardium extends around the right ventricle as well. However, when viewed in the short axis (as shown in Slice A), the myocardium around the right ventricle is much thinner than around the left, and therefore is often unlabelled, like in the ACDC dataset. If the myocardium around the right ventricle was included, the topology would change to $B_{\text{myo}} = \{b_0 = 1, b_1 = 2\}$.

5.4.1.2 Prior Shape

When working with multiple anatomical segmentations, it is vital that not only do the individual structures have the correct topology but that their relationships to one another are also correct. For example, when viewed in the short axis as shown in Figure 5.7, the left ventricle is always encompassed by the myocardium. If a left ventricle segmentation had the correct topology but was no longer within the myocardium segmentation, then the segmentation would be incorrect, which would be highlighted by poor Dice and distance metrics. This example is extreme, but more subtle changes, such as gaps between the left ventricle and myocardium segmentations are more commonly seen and may not be as noticeable from the typically used metrics. By designing a prior shape that has the correct relationship between the structures, e.g. the left ventricle prior ($\mathbf{P}_{\text{lv}}$) fills the entire myocardium prior ($\mathbf{P}_{\text{myo}}$) and the right ventricle prior ($\mathbf{P}_{\text{rv}}$) is always touching the outside of the myocardium ring, and deforming it with a topology-preserving field, should preserve their relationship to one another. To assess this, the *scene* topology can be measured. The scene refers to all the structures within an image, which can be produced by binarising all the structures' labels, e.g. 0: *background*, 1: *foreground*. The scene topology then measures the topology of the binarised structures.

For the heart, the overall scene has a Betti score of $B_{\text{scene}} = \{b_0 = 1, b_1 = 0\}$, the equivalent to a solid circle. However, if n separate gaps were present between the left ventricle and myocardium, this would form holes within the scene, creating a Betti score of $B_{\text{scene}} = \{b_0 = 1, b_1 = n\}$. Therefore, measuring scene topology, paired with the other metrics (e.g. Dice and individual topology), can provide a good assessment of the relative position between the structures.

Across the dataset, the heart slices were unaligned and in different orientations, and therefore the position of the right ventricle relative to the myocardium varied. This inconsistency is shown in Figure 5.8 where a heatmap shows the density of right ventricle labels in the train, validation, and test set. I investigated whether the position of the $\mathbf{P}_{rv}$ relative to the $\mathbf{P}_{myo}$ affected the segmentation performance

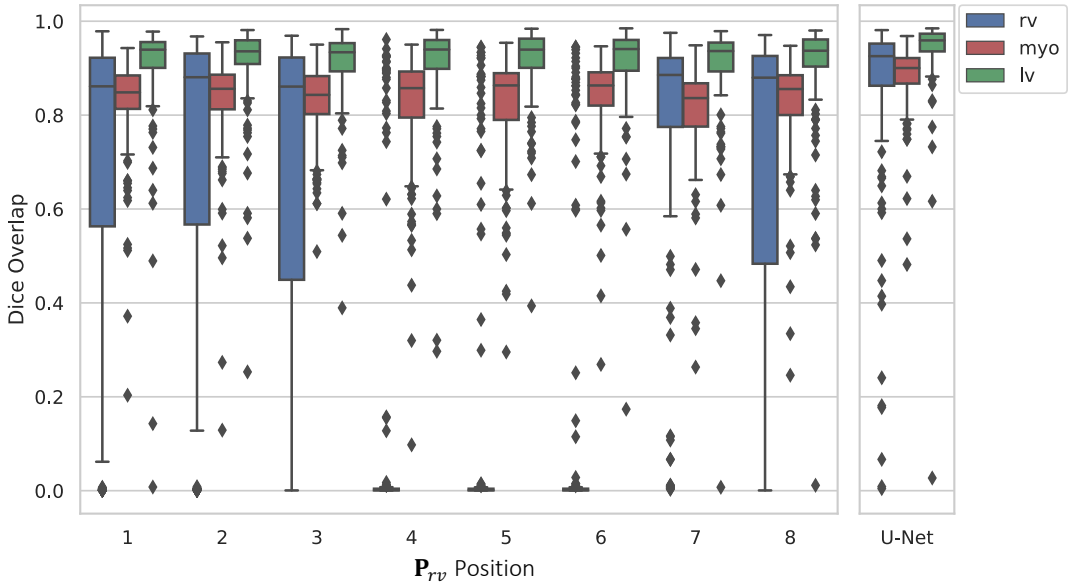


Figure 5.9: The segmentation performance when $\mathbf{P}_{rv}$ was at each position defined in Figure 5.8, evaluated using the validation set.

by designing a set of priors where $\mathbf{P}_{lv}$ and $\mathbf{P}_{myo}$ were kept the same but the position of $\mathbf{P}_{rv}$ around $\mathbf{P}_{myo}$, was varied.

5.4.1.3 Network Implementation

Moving from single to multi-class segmentation requires very little adaptation of TEDS-Net. The number of channels, one for each structure, in the prior label and output is increased to $c = 3$: $\mathbf{P} \in \mathbb{R}^{c \times h \times w}$, $\mathbf{Y} \in \mathbb{R}^{c \times h \times w}$, $\hat{\mathbf{Y}} \in \mathbb{R}^{c \times h \times w}$. However, importantly, the number of fields learnt stays the same, as a single field is used to deform all channels, to try to preserve the topological relationship between the channels. For multi-structure segmentation, the final image downsampling was changed from MaxPool to linear interpolation, as when working with multiple structures that share boundaries, MaxPool often leads to overlapping structures, violating topology. The architecture hyper-parameters and training implementation were kept the same as described in Section 4.5.2.

5.4.2 Results

Prior Shape

Figure 5.9 shows the Dice performance across the range of rv prior positions defined in Figure 5.8C, evaluated on the validation set. The majority of the rv labels within the dataset were in positions 1,2 and 8, as shown in Figure 5.8. However, when using the $\mathbf{P}_{rv}$ in these positions, there was a variation in right ventricle segmentation performance across the validation set, as shown with a large standard deviation. On visual inspection, generally, these networks successfully localised the right ventricle when the labels were found in positions $1-3$ and 8, but failed in all other cases. This suggests that the network did not learn to rotate the prior, as rotation was not required for the majority of the segmentations. Conversely, only a small fraction of rv labels were found in positions 4-6 but when using $\mathbf{P}_{rv}$ in these positions, the segmentation failed except for a small number of images where the rv was actually located in positions 4-6, which also suggests that these networks did not learn to rotate the prior. This might be due to the flipping augmentation, which provides the networks with more rv examples in positions 4-6 as the greatest density of rv labels are found in the opposite positions: $1, 2, 8$. Rotational augmentations ($\pm 5°$) were also applied, however, these rotations may have been too small to change the behaviour.

Finally, position 7 contained the lowest density of rv labels, however, the rv segmentation performance was better and more consistent when using $\mathbf{P}_{rv}$ in position 7, than the other priors. This is likely due to the fact that the network was forced to learn to rotate the prior, as there was minimal overlap between the right ventricle labels and $\mathbf{P}_{rv}$. Moreover, unlike in the examples above, image flipping should not negatively impact the network's ability to rotate, as position 7 and its opposite position 3 both have a low density of rv labels. *As position 7 resulted in the best performance, it was used for the rest of the experiments.*

Table 5.1: Segmentation performance and topology preservation rates between our TEDS-Net and the baseline U-Net, on each structure: the right ventricle (rv), the myocardium (myo) and left ventricle (lv) as well as the overall scene.

Method	Dice ↑				Correct Topology ↑			
	rv	myo	lv	scene	rv	myo	lv	scene
U-Net	0.88 ± 0.17	0.89 ± 0.10	0.95 ± 0.11	0.96 ± 0.10	99%	97%	99%	1%
TEDS-Net	0.78 ± 0.27	0.83 ± 0.10	0.93 ± 0.10	0.94 ± 0.05	100%	100%	100%	70%

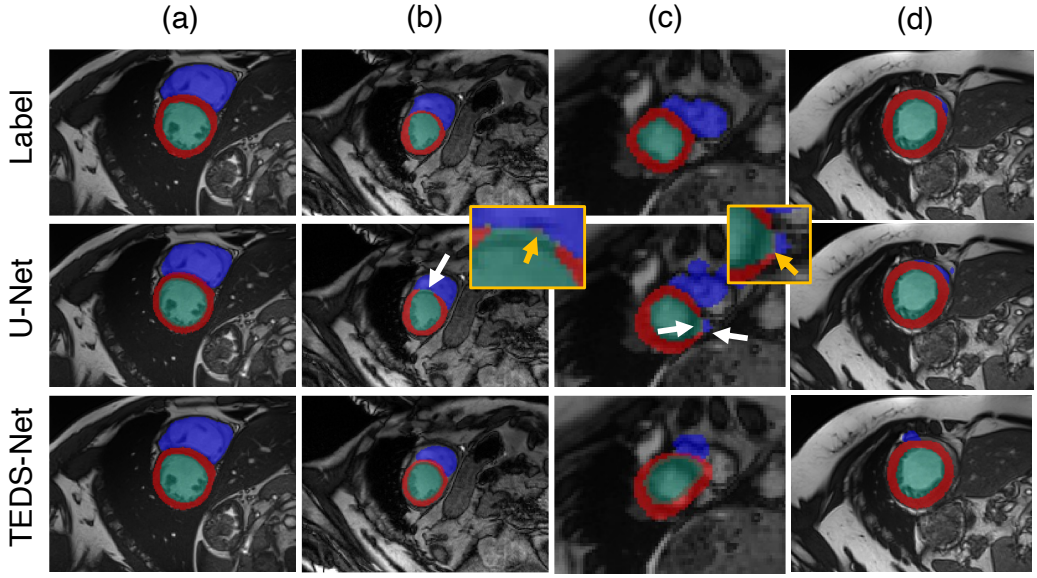


Figure 5.10: Four qualitative examples of cardiac multi-structure segmentation. Scene topology errors are shown in the zoomed panels with yellow arrows.

Comparison to Baseline

Table 5.1 shows the quantitative results from our multi-structure segmentation, on the held-out test set, between TEDS-Net and the U-Net, with qualatative results shown in Figure 5.10. TEDS-Net achieves 100% topology preservation for all three individual structures. This corresponds to the fields generated in TEDS-Net containing no folding voxels, calculated by measuring the Jacobian determinant at all points in the fields. The U-Net preserves topology for 99%, 97% and 99% of the RV, MYO and LV, respectively. However, there is a stark difference between the two networks when comparing the overall scene topology. The U-Net outputs are found to only have the correct scene topology in 1% of cases compared to 70% in TEDS-Net, where in the U-Net segmentations there are gaps that can commonly be found at the tissue interfaces, as shown in Figure 5.10c-d. The topological mistakes

found in TEDS-Net's *scene* were found to be overlapping channels, likely created in the downsampling and thresholding. Whereas, overlapping is prevented in the semantic U-Net due to the Softmax activation function.

The U-Net significantly (as measured with the paired t-test, corrected for multiple comparisons) outperforms TEDS-Net on Dice Score across all three structures and the overall scene, although it is competitive for the MYO and LV, and the overall scene. Between the two networks, the largest difference in Dice performance is found to be RV segmentation. TEDS-Net achieved an average Dice score for the RV of 0.78 ± 0.28, notably with a large standard deviation suggesting that performance is being skewed by outliers, as shown previously in Figure 5.9. Through visual inspection, the examples with the lowest RV Dice were found in scans where the RV was prominently in position 3, which would require a 180° rotation of the prior.

5.4.3 Discussion

This experiment demonstrates TEDS-Net's ability to perform multi-structure segmentation, using the ACDC dataset. I show how the same architecture and hyper-parameters can be used from our previous experiment, to still achieve promising performance, illustrating the robustness of TEDS-Net. In this experiment, TEDS-Net still achieved topology-preserving fields, despite the requirement for more complex deformations to transform the three channels, which corresponded to 100% topology segmentation across the individual structures. However, the topology of the overall scene (e.g. the combination of each label) was not always preserved. This finding highlighted how downsampling and applying a threshold to the predictions can impact the topology.

Scene topology can be used in multiple-structure segmentation, to provide an assessment of the relative position of each structure. As the vast majority of healthy anatomical structures in the body have a known arrangement, the prior knowledge of the anatomy incorporated into TEDS-Net has the potential to guide the arrangement of the structures as well as their individual topology.

In this experiment, I also showed that careful consideration of the prior shape is required when the positions of some structures relative to others are inconsistent. In this example, the position of the right ventricle relative to the myocardium was not consistent across the dataset, and I found that different priors had drastically different performances, which was likely due to the requirement of TEDS-Net to learn large rotations. However, in most medical imaging tasks scans there is a standard orientation in which scans are performed and therefore structure orientation should stay consistent, relieving the need for large rotations. In such a scenario, the prior should be designed in the same orientation as the standard view, with the flipping augmentation removed from training. Without a standard orientation, a large amount of augmentation is required during training to try and force the network to learn to rotate the prior. Further, initial rigid alignment of the images could be performed to circumvent this requirement, if the arrangement of the structures relative to the image orientation is consistent.

Thus far, TEDS-Net performance has remained relatively invariant to size (Section 4.6) and shape (Section 5.3.2) of the prior, illustrating that the prior does not have to be too similar to the final label to achieve promising segmentation results. However, the performance has varied drastically with alignment when performing multi-structure segmentation. Ideally, TEDS-Net's performance should be invariant to the prior with the only requirement being that the prior has the correct topology. As TEDS-Net's performance dropped when the network was required to learn a rotation, an extension to this project could be to initialise with a small network that learns to choose which prior shape would perform best for that image.

5.5 3D Liver and Few Slice Labelling

In this section, I extend TEDS-Net to perform segmentations in three dimensions, from CT scans. Similarly to our multi-channel extension, this requires little adaptation of the overall architecture described in Section 4.4.2.

I also explore TEDS-Net use for *minimally annotated data*. Collecting large amounts of labelled data can be extremely time-consuming, and tedious, and often

requires expert knowledge. This is further exacerbated in 3D, as the labelling must be performed in each plane, which can result in choppy segmentations (further discussed in Chapter 6.4.1.1). In recent works, methods have been proposed to reduce the need for a fully labelled dataset, by using weak-annotation strategies, such as; bounding-box labelling [232], scribble annotation [233] and one-shot learning [234]. Throughout this chapter, I class volumes with only a few slices labelled as *minimally annotated volumes*. Across the chapter, I investigate how to choose which slices to label and how many labelled slices are required. As TEDS-Net performs segmentation by registering a prior shape to the underlying anatomy in an input image or volume, I hypothesise that it will be able to perform whole structure segmentation whilst only being supervised by a few slices in each volume, e.g. using a minimally annotated dataset. As the fields within TEDS-Net are encouraged to be topology-preserving if some slice labels are missing, I hypothesise that TEDS-Net will be able to sensibly label the whole structure by using the prior knowledge of the topology of the structure, as shown in the MNIST experiments. Whereas, for traditional segmentation methods this might lead to gaps within the structure, as it is not encouraged to learn global shape features. In this section, I explore whether TEDS-Net can produce accurate segmentations using minimally labelled volume and investigate different realistic labelling scenarios.

5.5.1 Methods

5.5.1.1 Dataset

To perform 3D segmentation, I used the open-source Decathlon Liver dataset [235]. This dataset consists of 131 Portal venous phase CT 3D volumes and corresponding labels of the liver and tumours within the liver. The dataset varied drastically in dimensions and voxel spacing, with the axial plane separations between 0.7 to 5mm, across the dataset. The first task was therefore to preprocess the data. All the volumes were resampled to a voxel spacing of $3 \times 3 \times 3$mm using linear interpolation. The intensities within the CT volumes were then rescaled to between 0-1 using the same maximum (1000 HU) and minimum (-1000 HU), corresponding

to the Hounsfield units (Hu) in bone and air, respectively. In CT preprocessing, Hu rescaling is often performed around the Hu units of the structure of interest, to highlight it and dull the surrounding tissue. However, I chose to use the more generic approach, to stay consistent with the pre-processing used throughout the thesis, whilst increasing the difficulty of the task. All the volumes were then cropped or padded to match the target shape of $112 \times 112 \times 112$, which was chosen as it was greater than the size of the liver across the dataset and small enough to fit within the memory constraints of our hardware. The volumes were randomly cropped or padded around the liver labels, to ensure that all of the liver was included.

As the number of tumours and their location relative to the liver was inconsistent across the dataset, I combined the lesion and liver labels to perform a single class segmentation. With no tumours a liver has the topology equivalent to a solid sphere, however, if a tumour is located entirely engulfed by the liver, its topology would change. Moreover, depending on how many tumours (tn) were in the liver, and their location relative to one another, the topology would change further. Therefore, as the tumour can be located either on the surface of the liver (not changing its topology) or engulfed within the liver (causing a change), and tn is unknown, I chose to solely segment the liver (including any tumour labels as liver), as prior knowledge of topology is required for TEDS-Net.

From the 131 scans, 98 (75%) were used for training, 13 (10%) for validation and 20 (15%) for testing, chosen randomly.

5.5.1.2 Training Protocols

To investigate whether TEDS-Net can be used for 3D data and for minimally annotated volumes, I proposed three training protocols, shown in Figure 5.11. Firstly, I trained the 3D TEDS-Net with the fully labelled dataset to perform fully supervised liver segmentation. Next, I trained the 3D network on a minimally annotated dataset, where only some slices contained labels and the network was solely optimised using the losses calculated on those slices. This method is referred to as 2DLoss. The labelled slices were chosen based on several labelling scenarios

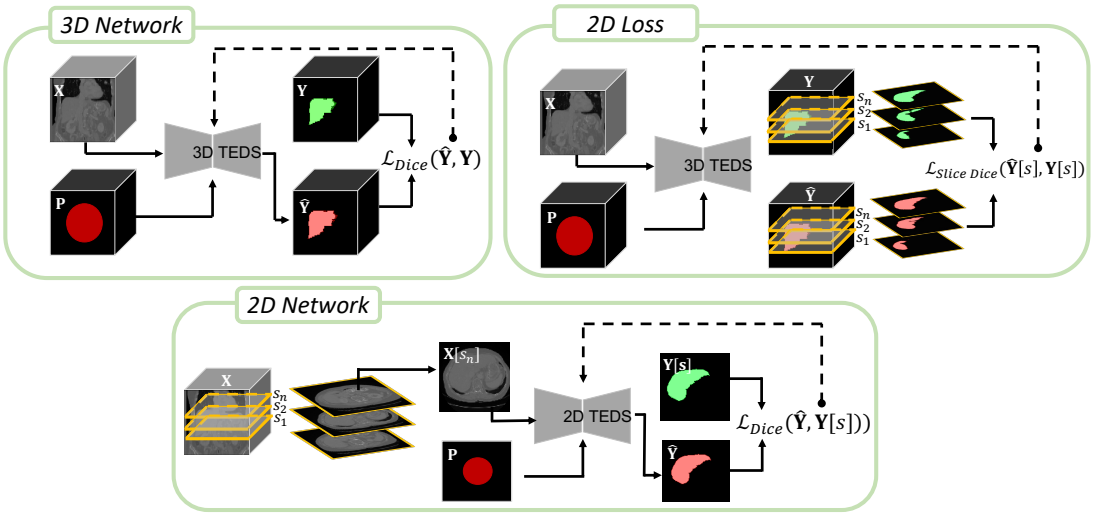


Figure 5.11: An overview of the three TEDS-Net training strategies. The fully supervised method is shown in the *3DNetwork*. Training using the minimally annotated dataset with the 3D TEDS-Net is shown in *2DLoss* and training using 2D inputs at set slices with a 2D TEDS-Net is shown in *2DNetwork*.

(described below). Finally, I trained a set of 2D networks using the 2D dataset made up of the same labelled slices described in our labelling scenario, which were used as the input, $\mathbf{X} \in \mathbb{R}^{112\times112}$. The final segmentation performance was evaluated by reconstructing the 3D liver, using the predictions ($\hat{\mathbf{Y}} \in \mathbb{R}^{112\times112}$) from each slice within a given volume. This method is referred to as 2DNetwork.

Across this experiment, I used labelled slices from the axial plane, as it is the most commonly used plane for annotations. Each scenario was designed to mimic possible labelling protocols, which would vastly reduce the time spent on labelling the whole volume.

Labelling Scenarios:

1. m **Liver Slices:** From each volume, m slices containing the liver were randomly selected.

2. b **Liver and** b **Background Slices:** From each volume b slices containing the liver and b containing background were randomly selected from each volume.

3. $\%p$ **of Liver Slices:** From each volume $p\%$ of the slices containing the liver were randomly selected. In the axial plane, the number of liver-containing

slices varied from 13 to 79. Therefore, selecting $n = 10$ random slices samples 77% of the liver in some cases and only 12% in others. To more evenly sample the liver across the dataset, p% of the slices containing the liver were randomly chosen from each individual 3D scan.

4. **Every j^{th} Liver Slice:** From each volume n slices containing the liver label were selected at even intervals. Each volume was assigned a random starting position, between 0 and 5 slices from the first slice containing liver, and every j^{th} slice was taken.

5. **Top, Middle, Bottom:** From each volume, the labelled slices from the top, middle and bottom of the liver were selected. As the top, bottom and middle of the liver can be hard to identify by eye, these slices were chosen within ± 3 slices of real position, whilst ensuring it still contained the liver. I investigated including all three labels (TMB), just the top and bottom (TB) and only using the middle slice (M).

5.5.1.3 Network Implementation

3D TEDS-Net: Moving from 2D TEDS-Net to 3D TEDS-Net is relatively simple, with the only changes required being to move from 2D to 3D convolutions, and increase the dimensions of the generated fields to 3D: $\mathbf{\Phi} \in \mathbb{R}^{3\times H\times W}$.

Due to the increased number of parameters required for 3D segmentation and memory limits, only the fine-tuning branch of TEDS-Net was used. This was also made possible as the livers were relatively well aligned initially, with the same spacing and similar sizes across the dataset. The TEDS-Net architecture was kept consistent with the original implementations (as described in Section 4.5.2), where the encoder-decoder was set up with $f = 12$ initial filters, a depth of $l = 4$ layers, and the fine-tuning branch being attached to $l_{\text{ft}} = 2$, whilst our topology preserving fields were generated through $h = 8$ composition layers, using Gaussian smoothing with $\sigma = 2$ voxels and upsampled to twice the resolution of the input image. For the fully labelled dataset, the same loss function was used as described

in Equation 4.5, with $\alpha = 2500$, chosen empirically by monitoring the ratio between the losses from the initial training loss curves.

For our minimally annotated experiment (denoted 2DLoss), the field regularisation loss function ($\mathcal{L}_{\text{Grad}}(\mathbf{\Phi}_{\text{FT}})$) was still used on the whole deformation field ($\mathbf{\Phi}_{\text{FT}} \in \mathbb{R}^{3\times228\times228}$), but the segmentation loss was only calculated on the set of labelled slices $S = \{s_1, s_2, \cdots, s_n\}$ and the mean Dice across the set was used for optimising:

$$\mathcal{L}_{\text{SliceDice}} = \frac{\sum_s^S \mathcal{L}_{\text{Dice}}(\mathbf{Y}[s], \hat{\mathbf{Y}}[s])}{n} \tag{5.1}$$

where, $\mathbf{Y}[s] \in \mathbb{R}^{112\times112}$ indicates a 2D slice at position s and n is the number of labelled slices used. The combined training loss was therefore set to, $\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{SliceDice}} + \alpha\mathcal{L}_{\text{Grad}}$, where $\alpha = 2500$.

As the liver has a topology that is equivalent to a solid 3D object, the prior shape was assigned to be a solid sphere with a radius of 30 voxels, chosen to be a similar size to the livers across the dataset.

2D TEDS-Net: To compare performance when only using the minimally annotated volumes, a 2D TEDS-Net was trained using the same set of labelled slices (S), denoted 2DNetwork. The 2D TEDS-Net is trained on each slice $\mathbf{X}[s] \in \mathbb{R}^{112\times112}$ using the label $\mathbf{Y}[s] \in \mathbb{R}^{112\times112}$, within the set S, to produce a prediction $\hat{\mathbf{Y}}[s] \in \mathbb{R}^{112\times112}$.

For this experiment, the prior shape $\mathbf{P} \in \mathbb{R}^{112\times112}$ was a solid circle with a radius of 30, similar to the 3D set-up. The network was set up with the same hyper-parameters described for the 3D architecture, still only using a single branch. The loss was calculated for each slice, using the same function described in Equation 4.5, and with a slightly reduced $\alpha = 2000$, chosen empirically from the loss curves.

At inference time, a prediction ($\hat{\mathbf{Y}}[s]$) is produced for every slice within a volume: $S_{\text{all}} = \{s_1, s_2, \cdots, s_N\}$, where $N = 112$ (the dimension in the axial plane), from $\mathbf{X}[s] \in \mathbb{R}^{112\times112}$, and the 3D-liver prediction is generated from reconstructing these 2D predictions and compared to the fully labelled liver volumes: $\mathbf{Y} \in \mathbb{R}^{112\times112\times112}$.

5.5.1.4 Baselines

As a comparison to the TEDS-Net implementations, I also trained a U-Net following the same setup; fully annotated, 2DLoss and 2DNetwork. The U-Net encoder-decoder was set up with the same hyper-parameters as TEDS-Net, to allow a fair comparison. The U-Net contained $l = 4$ layers, starting with $f = 12$ feature maps and was trained with Dice loss.

As a comparison to the minimally annotated dataset, I used Sli2Vol [234]. Sli2Vol is a SOTA method that segments a whole volume by propagating a single labelled slice across the volume. This method has been found to achieve extremely promising results, outperforming previous one-shot methods. This network first learns to reconstruct neighbouring slices in an unsupervised fashion, before performing segmentations. Sli2Vol was initially trained in a self-supervised fashion on a C4KC-KiTS dataset[1]. The network is trained on two neighbouring slices, e.g. $\mathbf{S}_i, \mathbf{S}_{i+1}$, and learns to reconstruct one slice from another, using an *affinity matrices* $\mathbf{A}_{i \to i+1}$. Key to this network is the introduction of an *edge profile generator*, which forces the model to pay attention to edges when performing the reconstruction. The trained model can then be applied to new datasets to perform segmentations. To do this, a single slice i from a volume $\mathbf{V}$ is labelled: $\mathbf{M}_i$. The model is then used to predict $\mathbf{A}_{i \to i+1}$ from two consecutive slices $\mathbf{S}_i, \mathbf{S}_{i+1}$ of $\mathbf{V}$, before $\mathbf{A}_{i \to i+1}$ is applied to $\mathbf{M}_i$ to calculate $\mathbf{M}_{i+1}$, this process is repeated for the next consecutive slices until the whole volume is labelled.

Therefore, Sli2Vol only requires a single labelled slice for each test set image and no labels during training. Whereas, when training the 3D U-Net and TEDS-Net, in the minimally labelled scenarios, n labelled slices are required per volume in training and zero labels are required in the testing phase. Further, as Sli2Vol propagates through the whole volume, the prediction time is notably slower than the other two methods, with the average time around 30s compared to 2s for 3D U-Net and TEDS-Net.

[1] dataset from: https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=61081171

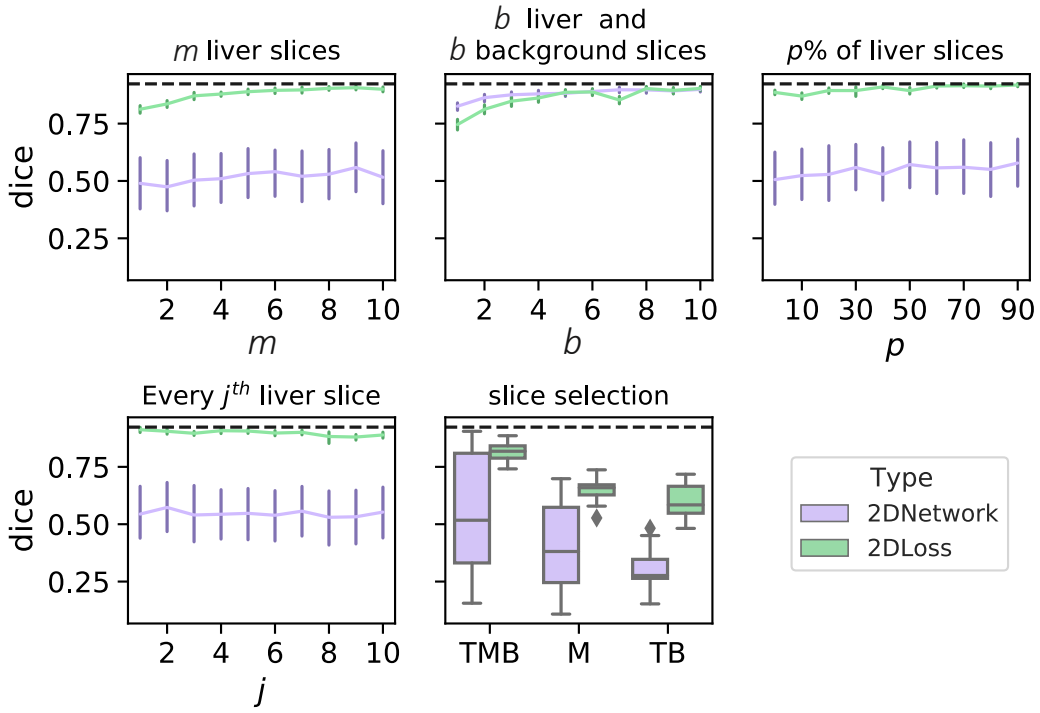


Figure 5.12: Comparing the segmentation performance when training TEDS-Net using different amounts of labelled data, with different labelling scenarios. The performance is compared between the 3D TEDS-Net trained using slice loss on selected slices (2DLoss) and between the 2D TEDS-Net which is trained on the labelled 2D slices and evaluated by reconstructing the 3D volume (2DNetwork). TEDS-Net performance on the fully annotated dataset is shown with a black dotted line.

5.5.1.5 Implementation and Evaluation

Each network was trained until convergence or 500 epochs, whichever came first, with a learning rate of 0.0001 using the Adam optimiser. The 2D networks were trained using a batch rate of 15, whereas the 3D networks with a batch of 5, due to memory constraints.

For each network, the segmentation evaluation metrics; Dice, Hausdorff Distance, Surface Distance and BCE (introduced in Chapter 2.2) are calculated on the 3D predictions, either on the output of the network or on the result obtained after a 3D reconstruction from each 2D slice within a volume.

5.5.2 Results

Figure 5.12 shows the liver segmentation performance when TEDS-Net is trained on minimally annotated volumes. Across the labelling scenarios and the number of labelled slices, the segmentation performance from the 3D TEDS-Net trained on minimally annotated volumes (2DLoss) shown in green in Figure 5.12, is found to be comparable to that of the fully annotated dataset (shown with the black dotted line). Across the different labelling types in the 2DLoss scenario, increasing the number of labelled slices, n, improved the segmentation performance, which is expected due to the increased amount of supervision. Moreover, the segmentation performance was found to be on par with the fully labelled baseline when using $p > 50\%$, $8 \leq m < 10$, $8 \leq b < 10$ or $j \leq 6$, for the 2DLoss experiments. This is an extremely promising result as it suggests that TEDS-Net can learn a structure's properties by using a small fraction of the labelled slices. Moreover, this could reduce the requirement for fully labelled volumes, which can be time-consuming due to the number of slices within a volume and the number of planes.

The 3D networks trained on minimally annotated volumes were found to perform better than their 2D counterparts across the methods (shown in purple in Figure 5.12 and denoted 2DNetwork) with the exception of *"n liver and n background slices"*, where the two lines begin to overlap. When including background slices into the 2D network, the network was able to learn what is *not liver*, whereas, for the other scenarios, the networks are only optimised based on slices containing liver. Moreover, the same prior (**P**) (2D solid circle) is paired with all 2D slices, including those which contain no liver. In the *"b liver and b background slices"* scenario, the network learns how to best compress the prior in slices containing no liver, whilst preserving topology, to reduce the Dice loss.

Table 5.2 compares the different TEDS-Net and baseline models using supervision from only a single slice from the liver volume, against the fully supervised methods. The performance between the two fully labelled datasets is extremely similar. The U-Net narrowly outperforms TEDS-Net on Dice, BCE, and average surface distance (ASSD). However, only the difference in BCE performance was found to

Table 5.2: A comparison of liver segmentation performance between fully supervised methods, and single slice annotations across different baselines.

Method	Dice ↑	HD [mm] ↓	ASSD [mm] ↓	BCE ↓	Topology ↑
Fully Supervised					
U-Net	92.9 ± 3.6	29.1 ± 36.8	1.44 ± 0.4	0.10 ± 0.05	60%
TEDS-Net	92.3 ± 2.2	13.0 ± 11.0	2.94 ± 1.5	0.11 ± 0.06	**100**%
2D Network (one slice annotated per volume)					
U-Net	49.8 ± 18.8	189.3 ± 12.3	66.5±7.9	1.71 ± 1.35	0%
TEDS-Net	49.0 ± 26.0	165.0 ± 13.8	68.2 ± 14.9	2.16 ± 2.13	**90**%
3D Network (one slice annotated per volume)					
U-Net	35.1 ± 8.1	141.1 ± 12.9	3.5 ±1.6	5.32 ± 1.37	0%
TEDS-Net	81.8 ± 3.8	28.1 ± 10.4	6.4 ± 1.3	0.57 ± 0.25	**100**%
Sli2Vol	80.4 ± 5.3	54.4 ± 20.9	2.4 ± 0.6	0.3 ± 0.11	55%

be statistically significant, as measured using FSL's PALM function for classical multivariate tests, corrected for multiple subjects, multiple network and multiple metrics [229]. Appendix B shows the statistics in full. Whereas, TEDS-Net outperforms the U-Net on Hausdorff Distance (HD), where the U-Net achieves an average of 29.1 mm compared to 13.0 mm from TEDS-Net. The large difference in HD between the two methods can be partially explained by the topology score. TEDS-Net achieved 100% correct topology compared to the U-Net's 60%.

On visual inspections, the topological errors were found to be disconnected components, as shown in Figure 5.13, which lead to a large HD. In the U-Net example shown in Figure 5.13, part of the spleen has incorrectly been labelled liver, a result found in 4/20 of the U-Net predictions. This may be due to the similarities in HU attenuation between the two organs [236]. However, in TEDS-Net the segmentation is enforced to be made from a single component, and therefore the network successfully ignores these intensity similarities.

In all three TEDS-Net implementations, the generated fields are found to contain no folding voxels, as measured with the Jacobian Determinants, and the two 3D networks result in 100% correct topology. Topology preservation is not guaranteed in our 2D model, as the deformation fields have been learnt independently for each slice, so although the same prior is used, there is no guarantee that the warped shapes will align with the neighbouring slices. Despite this, the network achieved

90% topology preservation. Neither of the one-slice U-Net implementations resulted in correct topology across the dataset, which was largely due to incorrect predictions in slices containing no liver. However, in an additional experiment, when training the 2D U-Net using $b = 1$ liver and $b = 1$ background slices, the segmentation performance drastically improves to a Dice score of 73.1 ± 0.07. Although this produced more realistic segmentations, the topology performance was still poor, with artefacts found around the edges of the volume and gaps between correctly labelled regions of the liver. Moreover, it is important to note that the U-Net was not designed for minimally annotated datasets and receives no prior knowledge of the liver shape, and therefore is not expected to excel in the minimally annotated tasks.

Using a single labelled slice during training, TEDS-Net was found to outperform Sli2Vol on Dice, HD and topology, whereas, Sli2Vol performed better on BCE and ASSD. However, there are some differences between the two network implementations. Firstly, TEDS-Net required a single labelled slice per training volume, whereas, Sli2Vol requires a labelled slice per testing volume. Therefore, although both methods require minimal annotation, once TEDS-Net is trained, it can be applied to large datasets without the requirement for more manual annotations.

Large Hausdorff distances were found in both the minimally annotated U-Net models and 2D TEDS-Net. As these networks are solely trained on slices that contain the liver, the networks have not learnt what to do in slices containing no liver. Therefore, since the networks expect the liver in each slice, it is predicted even in background slices, which results in large distance errors. Each input to the 2D TEDS-Net is paired with the $\mathbf{P}$, even in cases with no liver. As the fields within TEDS-Net are enforced to be topology-preserving, the prior cannot be fully compressed. Therefore, the liver is labelled in each slice of the 3D reconstructed prediction, as shown in Figure 5.13, which results in a poor HD and Dice score.

From the 2D cross-section of the predictions shown in Figure 5.13, the fully annotated U-Net and TEDS-Net predictions correspond closely to the labels. However, when looking at the 3D renderings of the whole liver, all the networks achieve a much smoother finish than the labels, with TEDS-Net being the most

Figure 5.13: Qualitative comparison of the Liver segmentation using the two fully supervised approaches and TEDS-Net methods that use a single labelled slice.

smooth. The rough edges of the label might be due to the preprocessing steps, such as downsampling, which can result in sharp corners and a blocky appearance of the liver. From our previous experiments, TEDS-Net has been shown to often produce overly smooth predictions, due to the additional modifications included to more tightly enforce topology preservation. Despite the smoothness, the 3D rendering of TEDS-Net corresponds well to the main features and edges of the liver seen on the label. From the 2D Loss TEDS-Net, the prediction shown in the cross-section (bottom row), over-segment parts of the liver, and the 3D rendering shows an even smoother liver with a large fin-like artefact. Similar artefacts are found across the other predictions from this model. Therefore, although the 2D Loss TEDS-Net achieves a promising segmentation performance based on Dice, BCE and topology, on visual inspection very noticeable errors are still present, which are reflected in the relatively high HD and ASSD. This result shows again why a range of segmentation metrics are required to accurately assess segmentation performance.

5.5.3 Discussion

Across this experiment, I have shown how TEDS-Net can be trained using minimally annotated datasets to produce high-quality 3D segmentation. In the 2DLoss approach, the network is only optimised using the n labelled slices. Therefore, as TEDS-Net perform segmentations by warping the 3D prior shape with learnt deformation fields, for a single slice to be accurately segmented, TEDS-Net must predict a good approximation of the whole field. Thus, the fields can be learnt from the guidance of just a few labelled slices, which has the potential to reduce the burden of manual segmentations.

One of the most striking findings from this experiment was the relatively small drop in performance when using only $n = 1$ labelled slice to optimise the 3D TEDS-Net (2DLoss) compared to a fully annotated baseline. In this experiment, the labelled slice was chosen at random across the liver. Thus, across the dataset, the network was likely supervised by slices distributed across the different livers, allowing the network to learn the appearance of the liver from different slices.

However, when using a single slice from the middle of the liver (M) for each volume, the performance drops. This implies that TEDS-Net requires a distributed set of slices across the liver in the training set to accurately segment the whole liver when operating in a minimally supervised setting.

Further, the proposed labelling protocols were designed to be easily implemented when annotating new datasets. I have shown that labelling $n = 5$ random liver slices, achieves competitive performance to fully annotated data, with the potential of reducing labelling time by 90% (as the average number of axial slices = 52).

Crucially, in these experiments, I performed minimal adjustments to the TEDS-Net architecture proposed in Chapter 4, to demonstrate the ability of TEDS-Net without performing a vast amount of fine-tuning of parameters. The prior used in this experiment was the only shape investigated for this task, to show that TEDS-Net can achieve promising performance without optimising the prior shape, a potential limitation found in previous experiments.

Although TEDS-Net achieves extremely promising Dice, HD, BCE and topology results, often the segmentations are overly smoothed, an effect previously noted in Chapter 4. This result is likely due to the modifications implemented within TEDS-Net to try and enforce topology-preserving fields and will be discussed further in the next section.

In this experiment, I was only able to segment the liver, despite labels being available for tumours and the liver. This was because the topology of the tumours and the relative scene topology between the liver and the tumor was unknown. This further highlights that TEDS-Net requires prior knowledge of the structures of interest and their scene arrangement, and that it only works for known topologies.

5.6 Conclusion

In this Chapter, I have shown how TEDS-Net can be used to perform segmentations in a variety of tasks, whilst keeping the architecture and hyper-parameters largely the same as those proposed in Chapter 4, to show that in-depth hyper-parameter tuning is not needed when applying TEDS-Net to new datasets and tasks. By

doing this I showed how TEDS-Net is easily extended to work for multiple structure segmentation and in 3D. Further, I have shown how TEDS-Net can deform a range of topologies, from different image types, and combine different topologies within the same training cycle. Our proposed minimally-labelled protocols have shown that TEDS-Net can learn effectively using a partially labelled dataset, with the potential of reducing manual labelling by 90%, whilst achieving competitive performance. I have also outlined some of the limitations of this method, including the need for prior knowledge of anatomy, over-smoothing of structures, and difficulty in performing rotations in unaligned datasets.

The increased smoothness of TEDS-Net's predictions can be especially problematic for structures that are small and thin. For example, I tried to use TEDS-Net for both hippocampus and pancreas segmentation from 3D MRI and CT scans, respectively. Although TEDS-Net was able to correctly localise the structures, it was lacking detail and the shapes were overly bulbous. This resulted in the best models achieving a Dice of 58.8 ± 0.11 and 67.9 ± 0.09 for the hippocampus and pancreas, respectively, compared to the U-Net that achieved 71.6 ± 0.08 and 78.1 ± 0.11. Future work on TEDS-Net should primarily focus on how to reduce the unwanted smoothing artefacts whilst still enforcing topology preservation. One possible solution could be to move away from the discrete representations of the prior and fields, which require additional smoothing to prevent topology violations. In recent work, priors have been represented as meshes, with vertices lying on non-regular coordinates [237, 238], they are then deformed using learnt diffeomorphic fields by finding the deformations at non-regular coordinates using an ODE solver. Representing the shapes in this way relieves the requirement for resampling post-transformation, which may reduce the need for smoothing.

6

Characterising the Healthy Cortical Plate Development

6.1 Overview

Anatomical segmentations are often required as an intermediate step in medical image analysis pipelines. In previous chapters, I introduced our topology-preserving segmentation network (TEDS-Net) and demonstrated its ability to preserve topology in medical segmentations on a wide range of tasks. Here, I integrate TEDS-Net into a larger image analysis case study: analysing the developing cortical plate (CP) from ultrasound image volumes of the fetal head. Structures within ultrasound volumes can be challenging to segment due to large acoustic shadows obstructing regions of interest. Therefore, I utilise TEDS-Net's ability to guide segmentations in regions of poor contrast to segment the CP. As the segmentations are encouraged to be topologically correct, this then enables in-depth morphological analysis of the developing CP surface to be performed. Performing such analysis across a large, healthy cohort can provide healthy baselines of CP properties, which can be used for comparison to individuals. Further, comparing our measures to previously reported fetal growth patterns, can further validate TEDS-Net segmentation performance [41, 169, 175, 220].

Early versions of the work presented here have been published in:

Madeleine K Wyburd, Mark Jenkinson, and Ana IL Namburete. "Cortical plate segmentation using CNNs in 3D fetal ultrasound." Annual Conference on Medical Image Understanding and Analysis. Springer, Cham, 2020.

Madeleine K Wyburd, Mark Jenkinson, and Ana IL Namburete. "Deep learning-based assessment of second-trimester cortical plate development in three-dimensional ultrasound" in The International Society of Ultrasound in Obstetrics and Gynecology (ISUOG) 2021

Versions of this work have also been presented at the following:

Madeleine K Wyburd, Mark Jenkinson, and Ana IL Namburete. "Artificial Intelligence can Measure Fetal Brain Development from 3D Ultrasound" at STEM for BRITAIN 2022 -*Bronze Award for Engineering*

Madeleine K Wyburd, Mark Jenkinson, and Ana IL Namburete. "Using Deep Learning to Segment the Developing Cortical Plate from 3D Fetal Ultrasound" at Radiological Society of North America (RSNA) 2020

6.2 Introduction

In Chapter 3 I reviewed the motivation behind studying the developing CP and previous characterisation methods. Here I provide a brief overview of the key background required for this chapter.

As a human fetus develops, the smooth (lissencephalic) cerebral cortex begins to fold, a process known as *gyrification*, which continues through childhood and adolescence [111]. Gyrification allows for the cortical surface area to scale almost linearly with volume, which is associated with increased cognitive abilities [110, 111]. In healthy fetal development the occurrence of major peaks and troughs in the folds, otherwise known as *gyri* and *sulci*, has been shown to be temporally consistent across a large population [126, 188]. Abnormal cortical folding has

frequently been associated with cognitive impairment in adolescents and adults [141, 239] and more recently, deviations from the expected timings and depths of sulcal development have been identified in at-risk pregnancies [114, 115]. Having the ability to characterise and quantify cortical growth during development is therefore of huge importance, as it could help monitor pregnancies to ensure that the cortical development is at the expected rate and identify any biomarkers for risk and disease.

A plethora of methods to monitor fetal cortical development in clinical care have been proposed, including measuring depths or angles of individual fissures [156, 173] and grading the appearance of folds [168], as discussed in Section 3.2.3. However, such methods have been found to be extremely time-consuming and user-dependent [174]. Moreover, as these methods have been designed for specific 2D ultrasound planes, they require consistent localisation of the correct ultrasound plane and risk losing valuable 3D shape information. Thus, future methods to analyse fetal brain development should be performed in 3D to more accurately describe the complex morphological properties of the CP.

With the recent advances in fetal MRI (e.g. correcting for fetal motion and acquisition improvements [221] [240]), the 3D properties of the fetal brain have begun to be measured [112, 115, 176, 187, 188], as discussed in Section 3.3.2. In these studies, the CP is commonly extracted from MRI images before the morphological properties of the surfaces are measured. These pipelines are often aided by pre-existing analysis methods [213] and MRI toolboxes, such as FreeSurfer and BrainVisa [185, 191], which were previously developed for adult MRI and have since been adapted for newborns and fetuses [115, 184, 193]. Such toolboxes can offer automatic CP segmentation and analysis of the surfaces. Despite the success of such packages in fetal MRI studies, they are largely incompatible with other imaging modalities, including ultrasound (US) due to domain differences. Compared to fetal MRI, ultrasound has the advantages of being relatively low-cost, available worldwide, and having shortened acquisition times (in the scale of seconds) and it is the modality of choice in prenatal care. Therefore, it would be highly beneficial to develop a pipeline to analyse the developing CP from ultrasound volumes.

From the cortical plate properties measured from MRI, healthy atlases of fetal development have been generated [134, 176, 221]. Healthy atlases are required as a benchmark to compare individuals to, in order to ensure they are following the schedule trajectory. However, to our knowledge, there is no equivalent ultrasound-derived atlas. As discussed in Section 3.1.2.1, fetal brain measures and longitudinal emergence of structures are found to differ between ultrasound and MRI, and therefore healthy MRI atlases are an unsuitable benchmark to use for comparisons of an individual's properties as derived from ultrasound scans. Thus, there is a need for an atlas describing healthy cortical plate properties generated from ultrasound volumes.

This chapter can be broken down into four main sections. Firstly, I introduce the dataset used throughout this work (Section 6.3). Secondly, I explore automatic CP segmentation with TEDS-Net compared to baselines (Section 6.4). Next, I propose methods to analyse the developing cortical surface of individuals (Section 6.5). Finally, I present the healthy CP properties across the cohort compared to a range of previous MRI and 2D ultrasound studies (Section 6.6).

6.2.1 Contributions

In this chapter, our two main contributions are:

1. The development of an automated pipeline to analyse the cortical plate from ultrasound volumes.

2. The generation of healthy population atlases describing a range of cortical plate properties between 18 to 26 gestational weeks (GW).

For the first contribution, I propose a method of segmenting the developing CP from ultrasound volumes, which can be challenging due to large acoustic shadows and speckle artefacts. Therefore, I utilise the ability of TEDS-Net to anatomically guide segmentations in regions of contrast, to *segment through the shadows*. Next, I propose a number of automated methods to measure key properties

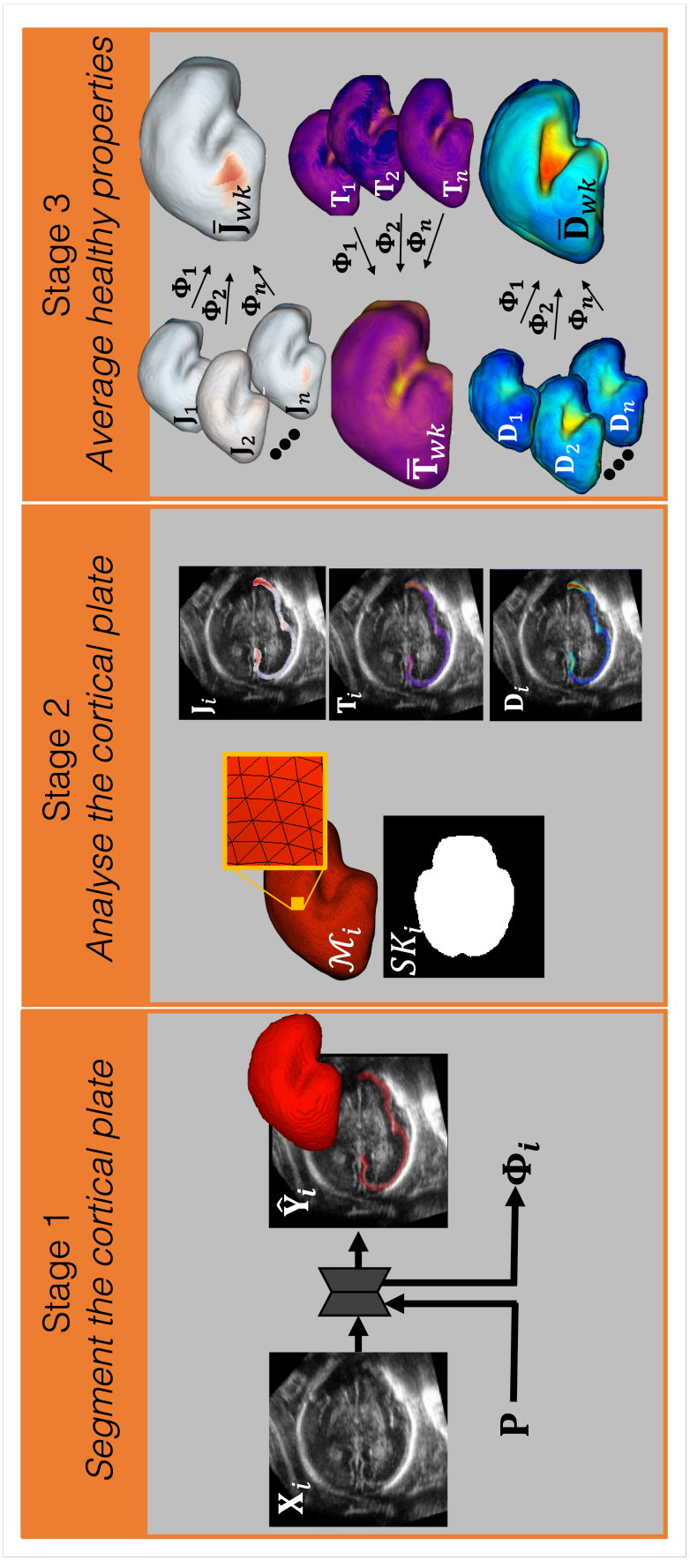


Figure 6.1: An overview of the pipeline used to measure the developing cortical plate. Step 1 is described in Section 6.4, Step 2 in Section 6.5 and Step 3 in Section 6.6.

of the developing CP, which are verified by comparing them to previously published MRI results. For the second contribution, I combine all the derived measures to generate expected healthy growth trajectories over gestation. An overview of these contributions is shown in Figure 6.1.

6.3 Image Dataset

Throughout the rest of this chapter, three-dimensional (3D) ultrasound volumes of the fetal brain obtained from the INTERGROWTH-21st Fetal Growth Longitudinal Study (FGLS) [241] is used. In this Chapter, I focus on the less studied cortical development in the second trimester, more specifically between 18 to 26 GW, which contains the beginnings for fissure development and growth [126, 129, 133]. Moreover, this period is when the anomaly scan is usually performed in routine prenatal care, as discussed in Section 3.2, and therefore, an important age to characterise in order to be eventually integrated into the routine assessment.

6.3.1 INTERGROWTH-21st

The INTERGROWTH-21st FGLS aimed to describe the optimal growth and neurodevelopment across gestation in a healthy group of fetuses, to establish new international standards of fetal and newborn development [241]. To achieve this ambitious goal, longitudinal ultrasound scans were performed on over 4,000 pregnancies, worldwide. The study was performed over eight urban areas, geographically delimited to ensure the study was representative of broad population characteristics. In total, 4,321 healthy pregnant women were enrolled, who satisfied a set of criteria that define them as having a low risk of impairment to fetal growth or diagnosis of fetal abnormality. For each pregnancy, a longitudinal set of 3D brain ultrasound scans were performed using a Philips HD-9 ultrasound machine using the same scanning-protocol, every 5 ± 1 week from as early as 14 GW and up until term ($\sim$40 GW) [241].

The fetal head scanning protocol required the fetuses to be positioned on one side so that the beam was tangential to the mid-plane. Further, as discussed in Section

3.2.2, due to reverberation artefacts from the skull only the hemisphere furthest from the probe (the distal hemisphere) contains clearly visible tissue boundaries, so throughout this work, only the distal hemisphere is analysed. Therefore, as the side on which the fetuses were lying was inconsistent (with approximately two-thirds of fetuses found to be on their left), the dataset is broken down into the "left" and "right" hemispheres.

Health Assessment In this chapter, I present the expected healthy growth trajectories across the second trimester. Further, to verify the analysis, the properties of the CP from our proposed pipeline are compared to previous measures of healthy development. Therefore, it is crucial that the fetuses included are in optimum health. As the INTERGROWTH-21st consortium aims to study healthy development, strict entry criteria were set for the mothers as well as follow-up cognitive tests performed on infants at 2 years of age. The women included had to have no clinically relevant obstetric, gynaecological or medical history. Moreover, there was further exclusion based on nutrition, education and socioeconomic status, with more details described in [242]. The follow-up assessment was performed on a subset of the initial INTERGROWTH-21st dataset (n= 1,181 children), where cognitive, language, motor and behaviour skills were assessed [243, 244].

Age Estimation Throughout this contribution, longitudinal CP development is assessed across gestation. Therefore, having a reliable measure of gestational age (GA) is required. Gestational age was based on the last known menstrual period (LMP), provided that the routinely used crown-rump length (CRL) age estimation was in agreement within 7 days.

6.3.2 Ultrasound Atlas

In order to perform supervised CP segmentation, CP ground-truth labels are required. Manual annotation is an extremely time-consuming task, with a single CP label taking around four hours, due to the complexity and size of the structure, filling the majority of the slices in each plane. This task is made even more challenging in

ultrasound volumes, where the artefacts vastly vary between individuals, obstructing different regions of CP, which can result in unreliable segmentations. Therefore, to perform supervised segmentation on our subset of $M = 537$ volumes, labelling alone would take $2,148$ hours, almost a year (assuming 40 hours/week), which is unfeasible within the scope of a single DPhil. Therefore, the first challenge in this chapter is to develop a pipeline to efficiently and accurately label the CP across the dataset.

One method to reduce the time taken performing manual segmentation is to use atlas-based segmentation, as discussed in Section 2.1.2. In this approach, individual volumes are registered to a labelled template, before the labels are propagated into the individual space using an inverse transform. Therefore, a labelled atlas with corresponding fields is required.

Atlas Generation In previous work, a weekly longitudinal fetal brain atlas was developed using a selected optimum subset of INTERGROWTH-21$^{\text{st}}$ dataset which was resampled to $160 \times 160 \times 160$ with the voxel sampled at $0.6 \times 0.6 \times 0.6$ mm to and manually aligned to a common reference space [42]. The optimum subset was partially selected based on the following criteria:

1. Cranium occupies $\geq 50\%$ of the image
2. Interhemispheric fissure and falx cerebri are clearly visible in the entire supratentorial region
3. Sylvian fissure is visible in the distal hemisphere
4. Distal cranial hemisphere is unobstructed by shadows

A qualitative comparison of examples of included and excluded data is shown in Figure 6.2. In the excluded dataset, large shadows are often obstructing regions on the distal side and key structures, such as the Sylvian fissure (SF), are often hidden. These criteria aligned with our goal of studying CP and, therefore, this subset was deemed suitable for our project. In total, 889 fetuses (1,059 images) were included in atlas construction, out of the total 4,321 fetuses (56,659 images) in the INTERGROWTH-21$^{\text{st}}$ dataset.

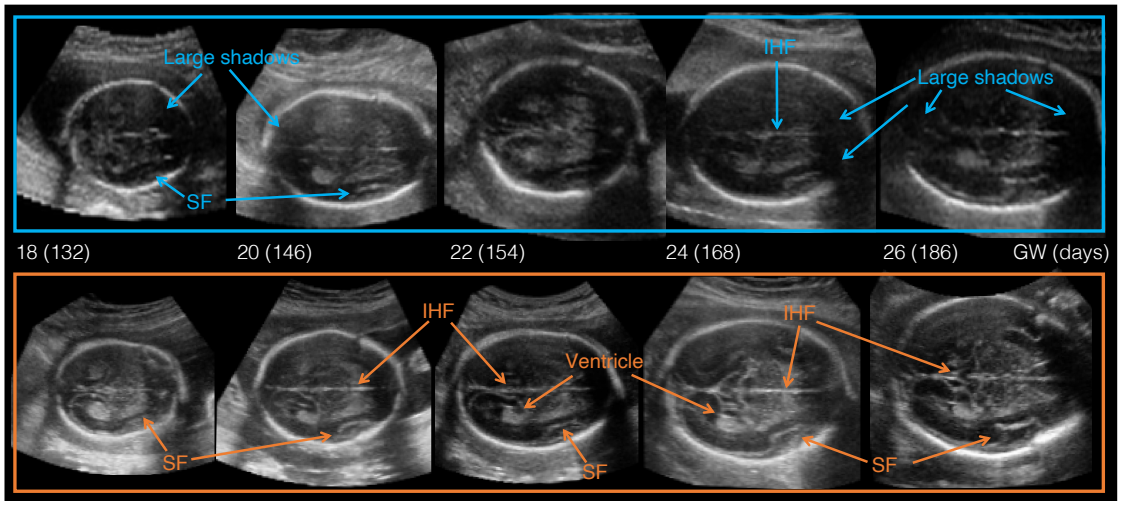


Figure 6.2: Qualitative overview of scans included (orange box) and those excluded (blue box) based on quality. The scans are shown in the transventricular plane. For the excluded dataset, labels have been added to show missing structures or artefacts to explain their exclusion.

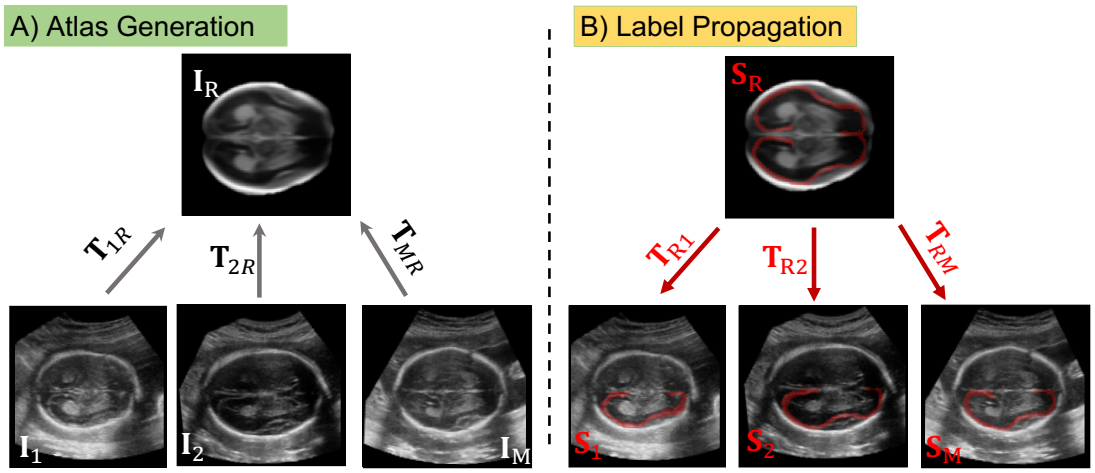


Figure 6.3: An overview of the generation of the INTERGROWTH-21st weekly atlas, generated in [42] and how the learnt deformations can be used for atlas-based labelling.

The weekly atlases were developed through diffeomorphic groupwise registration [93, 245], in which set of non-rigid transformations $T_{m\to R} = \{\mathbf{T}_{1R}, \mathbf{T}_{2R} \dots \mathbf{T}_{MR}\}$ were calculated to align a set of M images $I_m = \{\mathbf{I}_1, \mathbf{I}_2 \dots \mathbf{I}_M\}$, into a weekly atlas reference space R to generate the atlas ($\mathbf{I}_R$): $\mathbf{I}_R = \mathbf{T}_{mR}(\mathbf{I}_m)$, as shown in Panel A in Figure 6.3.

As these transforms are invertible, it is possible to move between the atlas and individual's space ($\mathbf{I}_m \leftarrow \mathbf{I}_R$), allowing atlas-based segmentation to be performed.

Therefore, by manually labelling the CP on the weekly ultrasound atlases, the labels can be propagated back to all the individuals used to create the atlas (shown in Panel B in Figure 6.3), dramatically reducing the time spent on manual annotation. Moreover, as ultrasound artefacts are inconsistent across individuals and as the atlases are an average of the individuals involved in building the atlas, the atlases contain much fewer shadowing artefacts than individual scans, and therefore the CP is more visible and easier to annotate.

Across this chapter, data is only included from the manually aligned optimum subset and between 18 to 26 GW, the age range of interest and a period of active folding, for training and evaluating segmentation performance, as these are the only volumes where ground truth labels can be generated using atlas-based segmentation. This decreased the size of the dataset to 537 volumes from 499 fetuses.

6.4 Segmenting the Cortical Plate

As commonly performed in MRI investigations, the first step in the CP analysis pipelines is usually to segment the cortical plate. In this work, supervised deep learning is used to segment the CP. This section begins with a discussion on methods, in which I describe how the cortical plate ground-truth labels are generated (Section 6.4.1.1), followed by reviewing the CNN architectures used (Section 6.4.1.3). Finally, the segmentation results are presented and compared between the different methods (Section 6.4.2.2).

6.4.1 Methods

6.4.1.1 Label Generation

In what follows, I describe the pipeline used to generate the CP labels for each individual through atlas-based label propagation. An overview of this pipeline is shown in Figure 6.4.

The first step was to manually annotate the CP in each weekly atlas. Manually labelling the CP was performed on 2D slices of the volumes and as the CP is visible in the majority of the slices within the 160^3 volume, it is important to

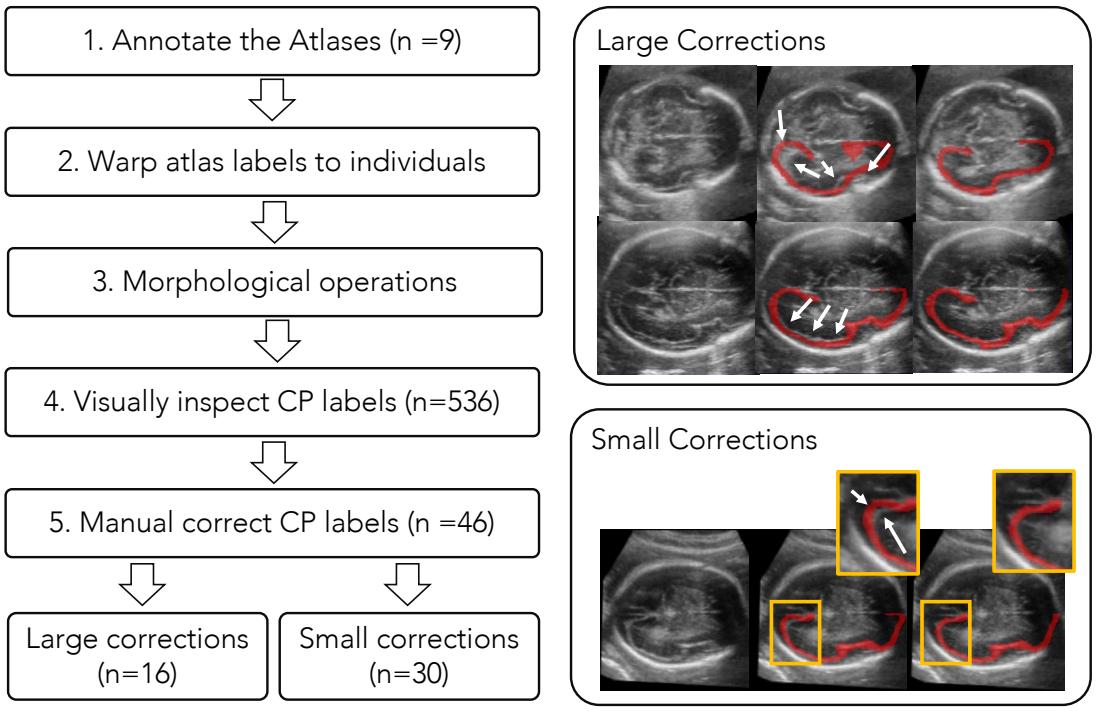


Figure 6.4: An overview of the pipeline used to generate the ground-truth cortical plate labels.

ensure that the label looks correct in each 2D plane (e.g. axial, coronal, sagittal). Further, it was important that each weekly atlas was consistently labelled. To tackle this, I designed and followed a labelling protocol to generate CP labels for each individual, as shown in Figure 6.5.

Each CP segmentation on an atlas map took around 4 hours and was performed on 3DSlicer [246]. Axial views of the final labelled atlases are shown in Figure 6.6 for each gestational week.

From the CP annotations in the atlas reference space, $\mathbf{S}_R$, individual CP labels ($\mathbf{S}_m$) were found by deforming the atlas label for the corresponding gestational week to each individual's space: $\mathbf{S}_m = \mathbf{T}_{Rm}(\mathbf{S}_R)$, as shown in Figure 6.3. A threshold of $t = 0.2$ was applied to each inverse transformation of the atlas label, followed by a closing operation using box kernel size of $c_k = [4 \times 4 \times 4]$, which removed any holes generated in the deformation. These values were chosen empirically, by visually inspecting the labels generated with different morphological operations and threshold values. I investigated applying opening, dilation, closing, and a combination of the

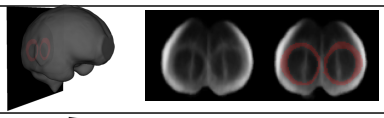
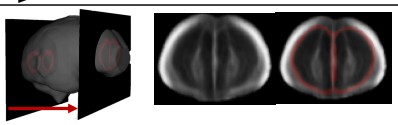
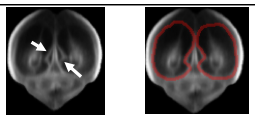
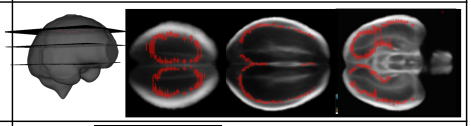
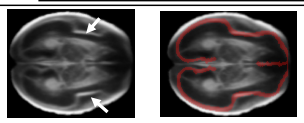
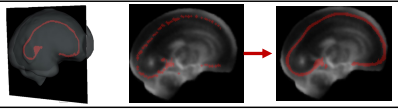
Begin in the coronal plane, starting at the posterior. Find the first plane where the CP is visible.	
Label the CP, moving from posterior to anterior, until we reach the last slice where CP is visible.	
Ensure the **Calcarine fissure** is well labelled	
Move into the Axial Plane, where the CP is likely to be jiggered. Start at the superior and begin to correct the CP label.	
Start at the superior and begin to correct the CP label. Ensure the **Sylvian fissure** is correctly labelled.	
Repeat again in the sagittal plane.	

Figure 6.5: Cortical plate (CP) labelling protocol for each of the weekly ultrasound atlases. The white arrows highlight the fissures referred to in the text.

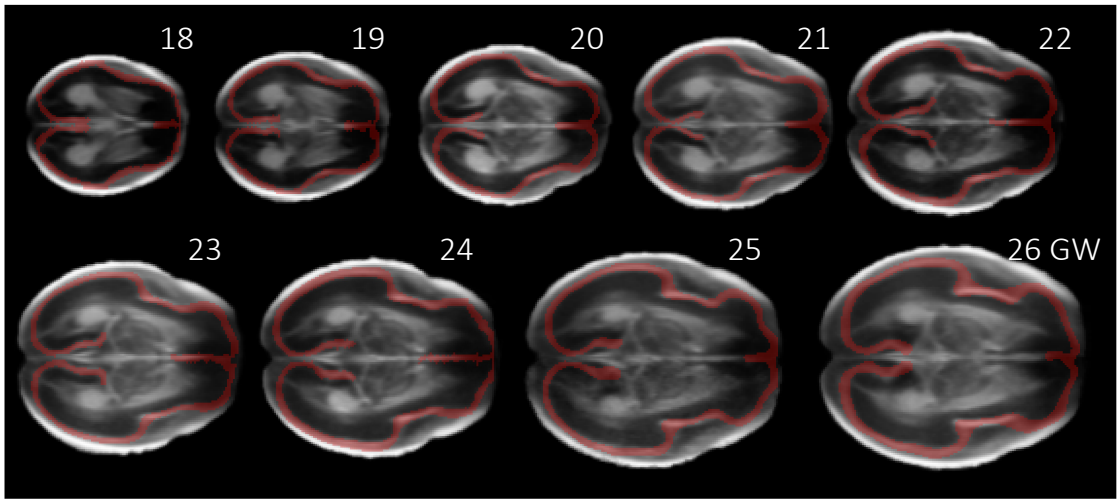


Figure 6.6: The manually annotated cortical plate labels ($\mathbf{S}_R$) are shown at each gestational week in the transventricular plane.

three for three kernel sizes: $[2 \times 2 \times 2], [3 \times 3 \times 3], [4 \times 4 \times 4]$. For each morphological combination, a range of threshold values ($t = \{0.2, 0.4, 0.6\}$) was investigated.

Although atlas-based segmentations facilitate time-efficient labelling, it is difficult to measure how reliable the deformed segmentations are. To ensure the warped individual CP labels were accurate, I visually inspected each of the predictions. From visual inspection, 46/536 (9%) of the scans needed adjustments, with 16 requiring large changes and 30 needing small adjustments, with examples of each shown in Figure 6.4. These images were corrected on 3DSlicer.

6.4.1.2 Data Preparation

The fetal head ultrasound scans ($\mathbf{X}_m \in \{160 \times 160 \times 160\}$) were used as the network input and the CP labels ($\mathbf{Y}_m \in \{160 \times 160 \times 160\}$)[1] as the ground-truth, where $m \in [0, 537]$. The network predictions had the same dimensions as $\mathbf{Y}_m$ and will be denoted as: $\hat{\mathbf{Y}}_m \in \{160 \times 160 \times 160\}$.

Data-split To train the CP segmentation networks, the dataset was split into three subsets, namely: training, validation and testing. The training dataset was used to fit the model, testing was used to evaluate the final model's performance, and validation was used to choose the best model and perform the intermediate evaluation. Therefore, it was important that the test set was unseen during the training phase. Moreover, as the dataset contains longitudinal data (e.g. multiple subsequent scans from each fetus), it is also necessary to ensure that each fetus is only in one of the three subsets so that the models remain blind to subject-specific anatomical characteristics within the test set.

As the appearance and size of the CP are known to change with GA [134, 175], it is important that each subset contains a similar age distribution. Further, as the task is separated into left and right cerebral hemispheres, an even distribution of hemisphere sides is also required. Figure 6.7 shows the data-split used throughout

[1]Note that, I am changing $\mathbf{S}_m$ to $\mathbf{Y}_m$ here, to keep consistent with previous notations in TEDS-Net.

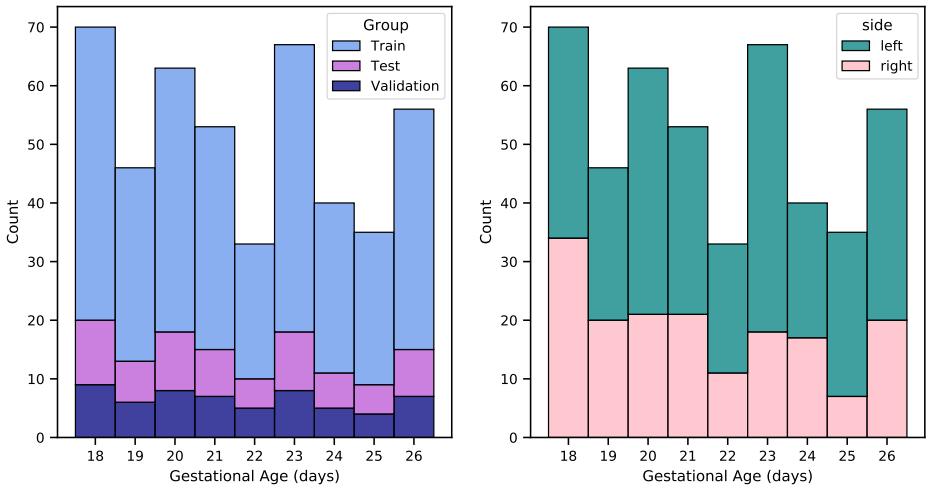


Figure 6.7: INTERGROWTH-21st training data split across gestation.

this experiment. Numerically, the dataset is split into 334 (72%) training, 59 (12%) validation and 70(15%) testing.

Data Augmentation Data augmentation was applied to the training dataset to prevent the networks from overfitting to the training dataset and to ensure the networks were robust to small changes in alignment. Isotropically scaling ($\pm$20%), translation ($\pm$20 voxels) and rotations ($\pm 5^{\circ}$) were randomly applied to each training volume at each epoch.

6.4.1.3 Deep Learning Approaches to Cortical Plate Segmentation

Atlas-based segmentation enables fast label generation for a set of individuals $I_M = \{\mathbf{I}_1, \mathbf{I}_2, \cdots, \mathbf{I}_M\}$. However, it requires a labelled atlas and the registration transformation between an individual and the atlas. As the alignment transformations are only available for scans used to create the ultrasound atlas (e.g. 1059 scans), atlas-based segmentation can not be used on the majority of the INTERGROWTH-21st volumes (e.g. over 55,000 scans), without first aligning them to the templates. As discussed in Section 2.1.3, deep learning methods offer an alternative technique to segment new volumes.

Segmenting the developing CP from ultrasound volumes can be challenging due to the large shadowing artefacts obstructing regions of the tissue, which can lead to gaps within the segmentations. Moreover, in order to perform follow-up surface analysis, the CP is required to be topologically correct. Therefore, as proposed in Chapter 4, I utilise TEDS-Net, as it strongly encourages topology preservation and has the ability to anatomically guide the segmentation in regions of poor contrast. In previous attempts to segment the CP whilst preserving topology, I implemented multi-task learning as discussed in Appendix D.1.

To the best of our knowledge, no deep learning-based methods have attempted to segment the CP from challenging ultrasound volumes. Therefore, to find a suitable baseline architecture, I turned to previous methods designed to segment the fetal CP from MRI volumes (as previously discussed in Section 3.3.1) [178–180, 203, 247, 248]. As in Dumast et al. [203], Dou et al. [178], and Khalili et al. [179], a U-Net style (encoder-decoder) architecture was used to perform CP segmentation, the same was implemented as the baseline [23].

TEDS-Net

For the task of CP segmentation, the TEDS-Net architecture is kept relatively consistent with that described in Chapters 4 and Chapters 5. Similarly to the 3D liver segmentation (see Section 5.5), the fetal head volumes are relatively well-aligned, so the *bulk* deformation branch was removed. Moreover, as each fetal brain scan has dimension $\mathbf{X}_m \in \mathbb{R}^{160\times160\times160}$, memory constraints[2] become a factor when implementing TEDS-Net. Removing the bulk branch reduces the number of parameters required for training, decreasing the memory demand.

In the original implementation of TEDS-Net, the bulk branch is involved in both prior alignment and size adjustment. As the input volumes are aligned but not scaled, removing the bulk branch might negatively impact TEDS-Net's ability to adjust for size. Moreover, between 18GW and 26GW, the volume of the CP is found to rapidly increase [175] and thus the average size of the CP varies with

[2]Training is performed with an Nvidia GeForce GTX 1080 GPU, with a memory of 10GB

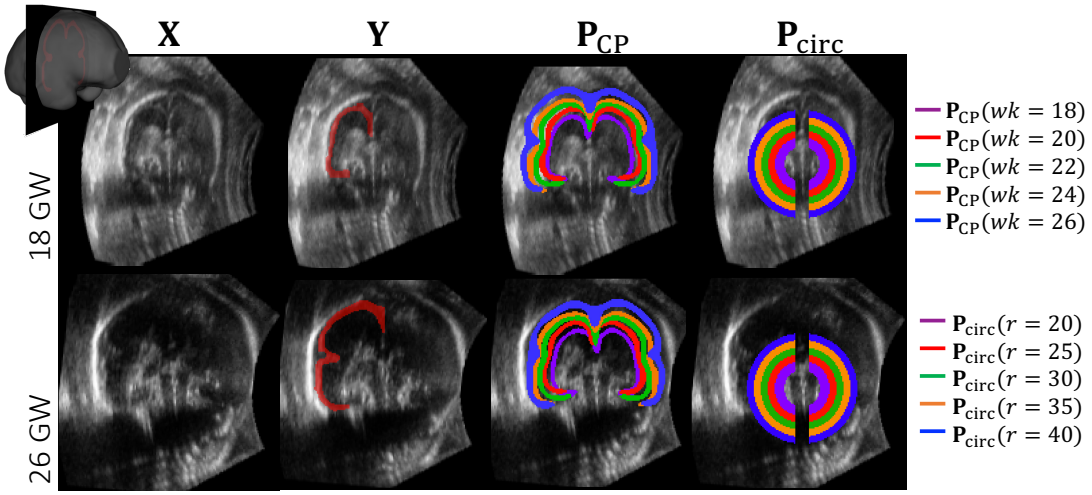


Figure 6.8: The set of priors investigated for TEDS-Net segmentation. The priors are shown on volumes between the lower (18 GW) and upper (26 GW) age range used to highlight the size differences of the CP.

age. Therefore, I hypothesise that the size of prior shape $\mathbf{P}$, is likely to play a key role in segmentation performance.

Prior Shape To investigate the impact that $\mathbf{P}$ had on the segmentation performance, I designed a range of priors, at different sizes, mimicking the different CP sizes across gestation. As an initial experiment, I used a set of hollow spheres ($\mathbf{P}_{\text{circ}}(\text{r})$) with radii: $r \in 20, 25, 30, 35, 40$ voxels (or $r \in 12, 15, 18, 21, 24$ mm) compare to the image dimensions of $\mathbf{X} \in \mathbb{R}^{(160\times160\times160\text{voxels})}$ (or $\mathbf{X} \in \mathbb{R}^{(96\times96\times96\text{mm})}$). Next, I generated a series of priors that more closely resembled the CP ($\mathbf{P}_{\text{CP}}(\text{wk})$), by applying Gaussian-smoothing to the atlas CP labels $\mathbf{S}_{R,\text{wk}}$ for $\text{wk} \in 18, 20, 22, 24, 26$GW. An overview of the priors in relation to the fetal brain and CP labels is shown in Figure 6.8. As the left and right hemispheres were treated as separate tasks, the priors were separated along the midplane.

U-Net

For one baseline a U-Net style architecture was used. To better regularise a CNN, *spatial dropout* is commonly added to the filters, as previously described in Section 2.1.3 [59, 60]. I hypothesised that dropout would help the network perform in regions of shadows, as it enforces the network to adapt in obstructed

regions, rather than relying on the image information at each individual pixel, thus improving topology performance. Therefore, I compared TEDS-Net to the U-Net with and without spatial dropout. In the case of dropout, I applied spatial dropout at the end of each convolutional block, to probabilistically mask out 20% of the activations, chosen empirically.

Network Implementation All models were trained for 400 epochs, using the Adam optimiser, a learning rate of 0.0001 and implemented in Python 3.7 using PyTorch (version 1.7.1). To facilitate a fair comparison between the U-Net and TEDS-Net, both architectures were constructed with $l = 4$ layers, each with two convolutional blocks, initialised with $f = 6$ feature maps, which doubled at each layer.

6.4.2 Segmentation Results

6.4.2.1 TEDS-Net Prior

The performance of each prior is reported on our validation set and shown in Figure 6.9. To investigate the presence of statistical differences between the prior performance, ANOVA multiple-comparison tests were performed between each set of priors: ($\mathbf{P}_{\text{CP}}(\text{wk})$ and $\mathbf{P}_{\text{circ}}(r)$). The difference in performance between the $\mathbf{P}_{\text{CP}}$ priors were not found to be significant ($p > 0.05$). Conversely, $\mathbf{P}_{\text{circ}}$ priors were found to differ significantly ($p < 0.01$), where $\mathbf{P}_{\text{circ}}(r = 20)$ was found to perform significantly worse than the other circle priors, surviving multiple-comparison corrections on paired t-tests [231][3].

I hypothesised that the CP segmentation performance would be better when the size of $\mathbf{P}$ was similar to the individual's CP size, due to the missing bulk deformation branch. In other words, when using $\mathbf{P}(wk = 18)$, I hypothesised that the segmentation performance would be higher for individuals at 18 GW than the older ages. To investigate this, segmentation performance (Dice) is plotted with respect to gestational age, to reveal any relationship between the two. For each prior, the segmentation performance was found to decrease with age, evidenced by

[3]A note on statistical tests: In this chapter, I will be performing multi-comparison corrections for each hypothesis rather than the whole chapter or thesis

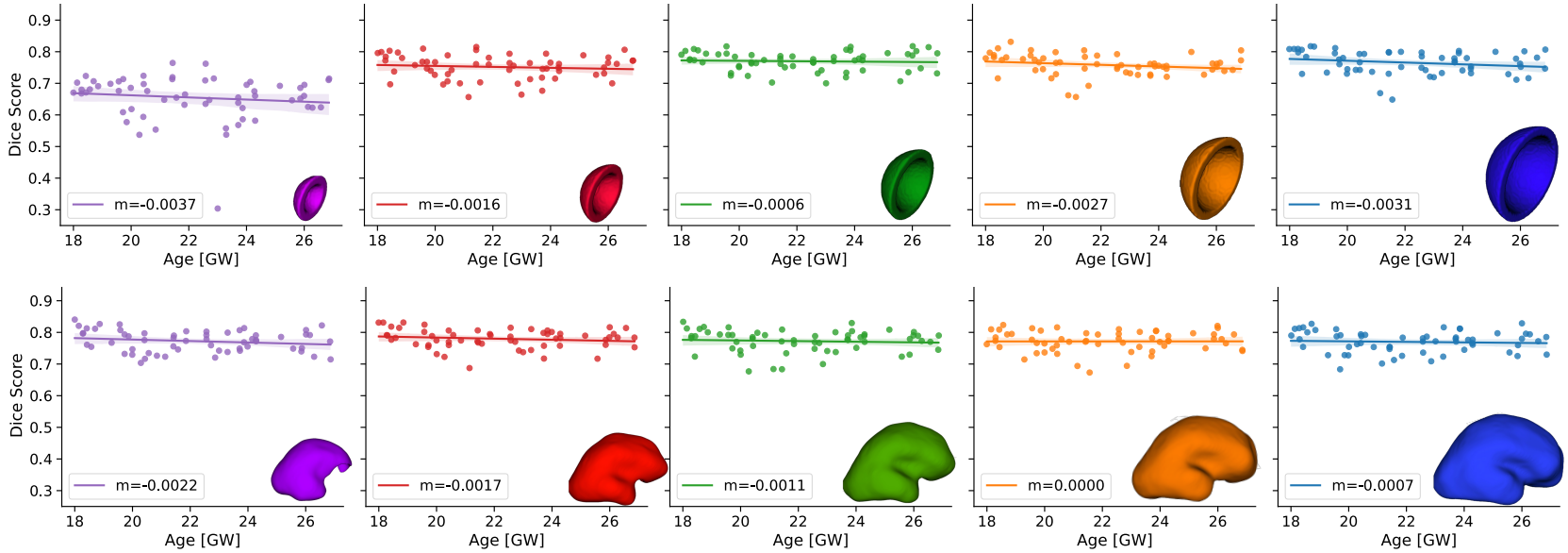


Figure 6.9: TEDS-Net segmentation across the different priors. For each experiment, the gradient m is measured to show the trend of segmentation performance over age. The colours of the priors correspond to the labels in Figure 6.8.

a negative gradient ($m < 0.0$) in Figure 6.9. The negative correlation between age and Dice score was found to be significant (r=-0.13, p<0.05) for the set of circle priors ($\mathbf{P}_{\text{circ}}$) but not for the set of CP priors ($\mathbf{P}_{\text{CP}}$) (r=-0.09, p=0.1), measured using repeated-measures of analysis of covariance (ANCOVA) [249].

To investigate whether the segmentation performance was significantly better at one gestational week compared to another (rather than across the whole gestational range), I performed an ANOVA multiple-comparison t-test between each gestational week, for each prior. No significant differences were found between the gestational weeks when segmenting using any of the $\mathbf{P}_{\text{circ}}$(r) priors. The same was true for $\mathbf{P}_{\text{CP}}$(wk) priors, except for $\mathbf{P}_{\text{CP}}$(wk=18) and $\mathbf{P}_{\text{CP}}$(wk=26), where the performance was found to significantly vary with age ($p < 0.05$). To investigate this statistically significant result further, unpaired t-tests with Bonferroni correction [231] were calculated between the Dice scores at each gestational week pairing. In both cases, only one pairing was found to be significantly different: between 18 and 20GW for $\mathbf{P}_{\text{CP}}$(wk=18) and between 18 and 21GW for $\mathbf{P}_{\text{CP}}$(wk=26). As the size of the cortical plate depends on age, measuring the significance between each gestational week helps determine whether TEDS-Net is performing better for labels that match the prior's size. From our statistical investigations, I show that for the majority of the cases, the performance variation with age is not significant.

As the segmentation performance variation for different priors was not substantial, albeit statistically significant, as shown by the shallow slopes in Figure 6.9, the best performing prior $\boldsymbol{P}_{\text{CP}}(\text{wk} = 26)$ *was used for the rest of the experiments.*

6.4.2.2 Baseline Comparison

In this section, I compare the performance of TEDS-Net to two U-Net baselines. The average segmentation performance across the different networks for cortical plate segmentation is shown in Table 6.1, with qualitative examples shown in Figure 6.10.

The U-Net and U-Net + dropout significantly ($p < 0.01$) outperformed TEDS-Net on Dice and BCE. However, TEDS-Net significantly outperforms HD. No statistical difference was found in ASSD across the three groups. These statistical

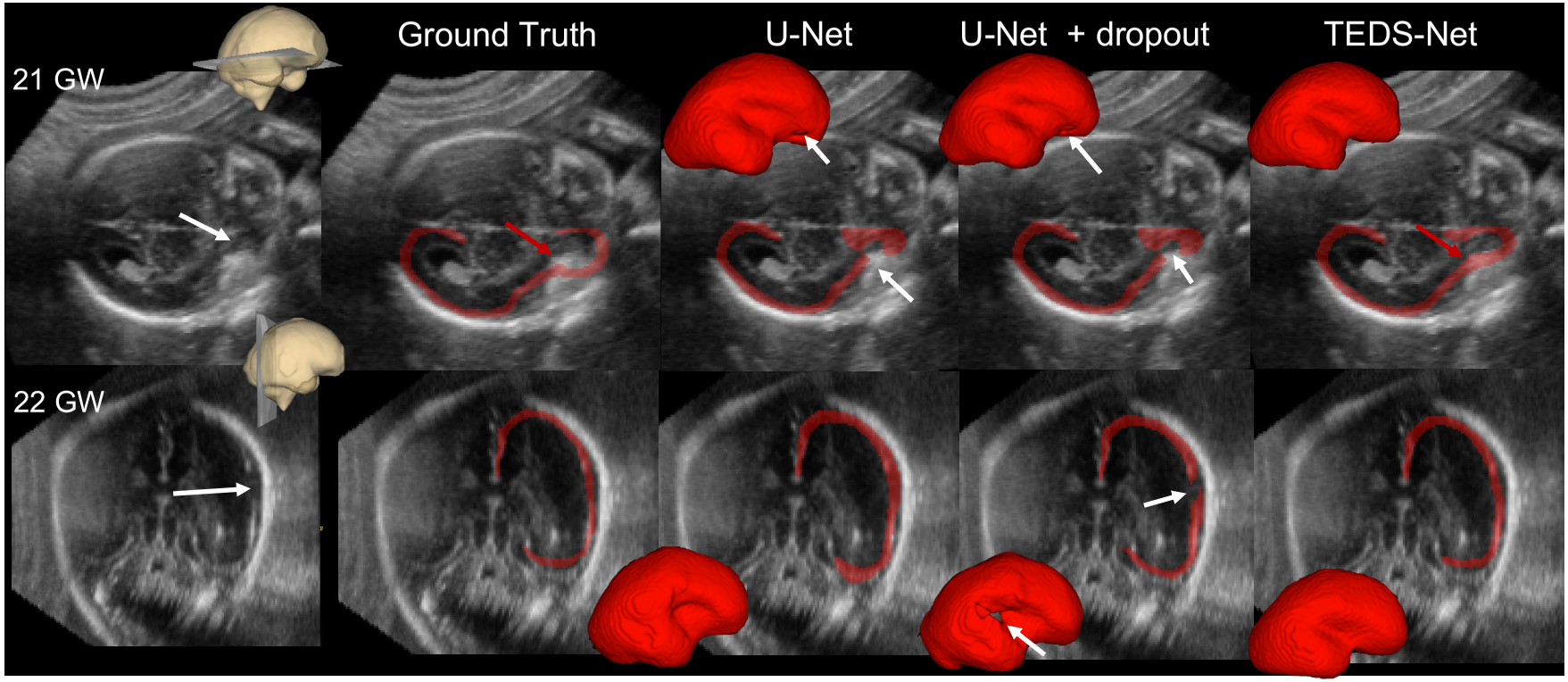


Figure 6.10: Example segmentations across the different networks, showing topological errors found in cortical plate segmentations. The white arrows point out the location of topological errors within the segmentations.

Table 6.1: Cortical Plate segmentation performance measured using Dice (↑), average surface distance (ASSD) (↓), 95th Hausdorff distance (HD) (↓), Binary cross entropy (BCE) (↓) and Topology performance (↑). For each metric, the best-performing network is highlighted in bold.

	Dice	ASSD [mm]	HD [mm]	BCE	Topology
U-Net	0.79 ± 0.05	0.37 ± 0.11	1.81 ± 2.22	$\mathbf{0.09 \pm 0.05}$	84%
U-Net + dropout	$\mathbf{0.80 \pm 0.05}$	$\mathbf{0.36 \pm 0.10}$	1.48 ± 0.09	$\mathbf{0.09 \pm 0.05}$	91%
TEDS-Net	0.77 ± 0.04	0.42 ± 0.08	$\mathbf{1.46 \pm 0.10}$	0.10 ± 0.05	**100**%

tests were performed using FSL's PALM function for classical multivariate tests, corrected for multiple subjects, multiple networks and multiple metrics, with the full statistical results shown in Appendix B [229]. TEDS-Net was the only network to achieve 100% topology preservation, corresponding to no folding voxels found within the fields. As hypothesised, including dropout into the U-Net architecture helped improve the topology performance from 84% to 91%, showing that dropout can help the network adapt to regions of poor contrast.

In Figure 6.10, qualitative examples of CP segmentations are shown, highlighting regions of topological errors. The top row shows a cortical plate segmentation in the axial plane. For both of the U-Net architectures, a "hole" is formed in the same location. On the underlying scan, in this region the CP is obstructed (shown with a white arrow). Similarly, the bottom row shows a hole in the U-Net + dropout segmentation in a region of the shadow. In both these cases, TEDS-Net is able to guide the segmentation through the shadows based on the fixed topology of the anatomical priors. These examples help illustrate the limitations of semantic segmentation, showing that such methods rely strongly on pixel information and if such values are corrupted (e.g. obstructed by shadows), the networks often miss-assign their class label rather than referring to learnt higher-level structure information.

As TEDS-Net generated topologically correct segmentations, I continue to use TEDS-Net predictions from here on, to facilitate a thorough analysis of the developing surface.

One of the benefits of deep learning compared to atlas-based labelling is its potential to segment new volumes almost instantaneously. The trained TEDS-Net

model was applied to two new datasets, firstly to a larger cohort of unlabelled INTERGROWTH-21st dataset (as shown in Appendix D.2), and secondly, to a dataset collected using a different ultrasound probe and independently evaluated by a neurosonographer (as shown in Appendix D.3).

6.5 Morphological Measurements

As discussed in Chapter 3, the developing CP can provide a vital insight into fetal development, and key geometric properties, such as volume and depth, have been found to significantly differ between clinically different fetal cohorts [114, 115]. In the adult brain, the curvature is often assessed, however, for fetuses, non-curvature-based metrics (such as depth) are found to consistently differentiate groups [216]. In this section, I describe the automated methods used to measure the developmental properties of the segmented cortical plate.

Packages and toolboxes have been designed to analyse the cortical surface from adult MRI scan [185, 191], which have more recently been adapted for newborn and fetal MRI scans. Often these packages are part of a complete pipeline, starting from CP segmentation, so the analysis techniques have been designed for a certain CP format. For instance, in such pipelines, the CP is often defined as the space between two closed surfaces: the inner and outer cortical boundary, with the gap created by the brainstem capped (see Section 3.3.2 and Figure 3.9). Methods, such as measuring thickness using potentials [213] depend upon this format (as discussed in Section 3.3.2). However, as only the distal hemisphere is visible in ultrasound, it is challenging to produce two closed surfaces and separating the inner- and outer-boundaries is non-trivial. Therefore, the majority of analysis methods utilised in the MRI pipeline are incompatible with our segmented CP volumes from the distal hemisphere and thus new methods are required.

In this section, I describe the methods used to analyse the developing CP. The measures are separated into two categories; *global* and *local*. Global properties return a single value describing the CP, whereas, local properties contain measures at each position along the CP.

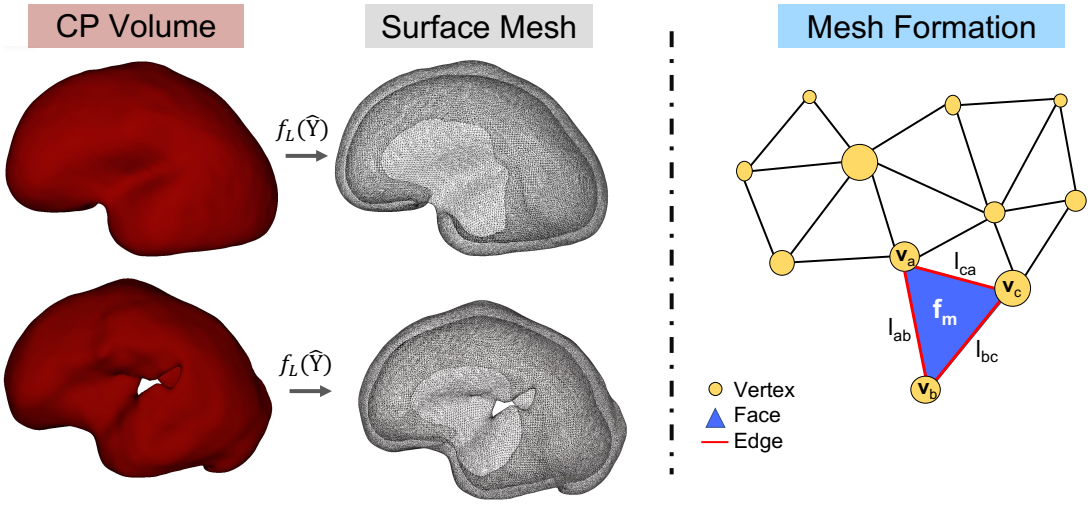


Figure 6.11: Qualitative examples of surface meshes generated from predicted CP volumes. The surface mesh encompasses the volume of the CP, around both the inner and outer cortical plate boundary. The top row shows a mesh generated from the TEDS-Net prediction, whereas, the bottom row shows a U-Net prediction with a topological error. The right-hand side illustrates the building blocks of a triangle mesh.

Global:

- Volume (V)
- Surface Area (SA)
- Sylvian Depth (SD)

Local:

- Surface Depth ($\mathbf{D}$)
- CP Thickness ($\mathbf{T}$)
- Relative Growth ($\mathbf{J}$)

6.5.1 Global Properties

Volume and Surface Area

Volume and surface area are amongst the most commonly used CP measures in fetal development [112, 114, 175, 209, 212, 216]. To calculate the volume of each CP volume ($\hat{\mathbf{Y}}_i$), I summed all the voxels ($\mathbf{v}$) labelled as CP: $\sum_{\mathbf{v}\in C}\mathbf{v}$, where $C = \left\{\mathbf{v} : \hat{\mathbf{Y}} = 1\right\}$, and multiplied V by the voxel volume (e.g. 1 voxel = 0.6mm^3).

To ease surface area computation, the CP surface was first converted into a triangular mesh $\mathcal{M}_i = f_L(\hat{\mathbf{Y}}_i)$ using the Lewiner marching cube approach (f_L) [250]. Crucially, this method preserves the topology of the underlying structure, which becomes critical when the CP becomes more convoluted, to prevent neighbouring

folds from merging together. Figure 6.11, shows an example of a CP surface mesh ($\mathcal{M}_i$) generated from a CP prediction with and without topological errors, generated from the U-Net (bottom) and TEDS-Net (top), respectively. The topological errors seen in the CP volume are transferred to the surface mesh and therefore can interfere with the properties measured from the mesh. Therefore, the topological guarantees provided by TEDS-Net are essential for this analysis.

Triangular surface meshes are defined in 3D using a set of N vertices $V = \{\mathbf{v}_1, \mathbf{v}_2 \dots \mathbf{v}_N\}$, where each vertex describes the 3D location of a point on the mesh: $\mathbf{v}_m = x, y, z$, and a set of M faces $F = \{\mathbf{f}_1, \mathbf{f}_2 \dots \mathbf{f}_M\}$, where each face is made up of three vertices, e.g. $\mathbf{f}_m = \mathbf{v}_a, \mathbf{v}_b, \mathbf{v}_c$. To compute the CP surface area, the surface area of each triangular face within the mesh was calculated and then summed together. The surface area of each triangle ($\mathbf{f}_m$) was calculated by measuring the length (l) of each edge (shown in Figure 6.11): $l_{ab} = \sqrt{\sum_{d\in x,y,z} \left(\mathbf{v}_a(d) - \mathbf{v}_b(d)\right)^2}$, $l_{bc} = \sqrt{\sum_{d\in x,y,z} \left(\mathbf{v}_b(d) - \mathbf{v}_c(d)\right)^2}$, $l_{ca} = \sqrt{\sum_{d\in x,y,z} \left(\mathbf{v}_c(d) - \mathbf{v}_a(d)\right)^2}$. From the sides the area was calculated using Heron's formula: $A(\mathbf{f}_m) = \sqrt{s(s-l_{ab})(s-l_{bc})(s-l_{ca})}$, where $s = (l_{ab} + l_{bc} + fl_{ca})/2$. Therefore, the total area of a surface mesh can be written as: $A(\mathcal{M}_i) = \sum_m^M A(\mathbf{f}_m)$.

Sylvian Fissure Depth

As discussed in Section 3.1.3, the Sylvian fissure (SF) is one of the most prominent folds on the CP surface in the second trimester and undergoes a rapid transformation during the gestational range studied here (18 to 26 GW) [137, 168, 169, 173]. Measuring the SF depth has a well-established protocol in 2D ultrasound images [169], as discussed in Section 3.2.3. In short, from the transventricular axial plane, SF depth is defined as the straight line between the border of the inner cortex to the inner parietal bone, perpendicular to the midline. This method is limited as it reduces a 3D structure into a 2D linear measure and is found to be extremely user-dependent [168], therefore, an automated 3D method could improve the accuracy and reliability of such a measure. Using the prior shape ($\mathbf{P}$) and

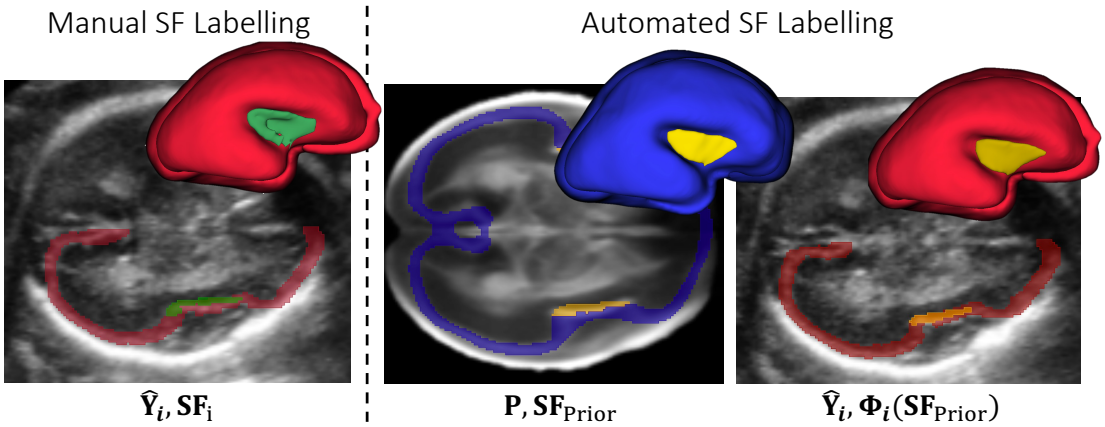


Figure 6.12: An overview of the Sylvian Fissure (SF) labels (**SF**). The 3D renderings show the inner cortical plate boundary. The predicted cortical plate volume ($\hat{\mathbf{Y}}_i$) is shown in red and the prior (**P**), in blue. On the LHS, $\hat{\mathbf{Y}}_i$ with the SF manually labelled ($\mathbf{SF}_i$) is shown in green and on the RHS, the labelled prior $\mathbf{SF}_{\text{Prior}}$ and the corresponding warped label using the learnt deformation fields $\boldsymbol{\Phi}_i$ are shown in yellow.

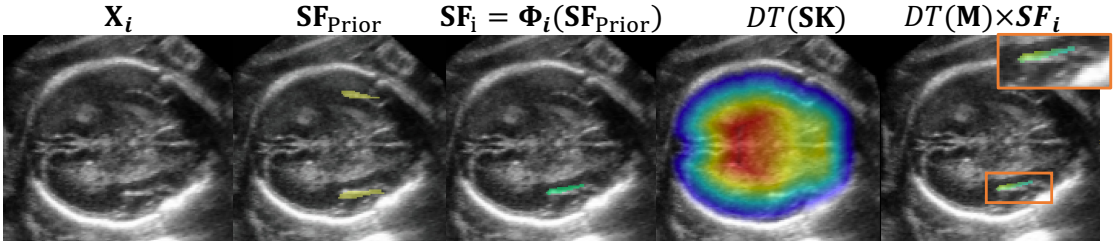


Figure 6.13: Automated pipeline for measuring SF depth. The SF seen on the prior (**P**) is labelled: $\mathbf{SF}_{\text{Prior}}$, before being warped into an individual's space using the learnt deformation fields from TEDS-Net ($\boldsymbol{\Phi}_m$). The distance transform (DT) from the skull masks (**SK**) is then calculated and multiplied by the deformed SF position, to give a set of SF depths.

the learnt deformation fields ($\boldsymbol{\Phi}_m$) from TEDS-Net, I propose an automated 3D method for measuring SF depth.

At the gestational age range studied here (18 to 26 GW), the SF starts out as a long, (relatively) flat fissure, whose walls become steeper with progressing gestation. We are interested in classifying the depth from the skull to the inner boundary of the CP across the flat section (referred to as the *base*) of the fissure (as shown in Figure 6.12). Therefore, to measure depth, this section of the SF must be obtained.

Manually labelling the base of the SF would be a tedious and time-consuming task, so instead, I proposed an automated method to detect this region. The prior shape ($\mathbf{P}(\text{wk} = 26)$) used in TEDS-Net was based on the atlas CP label at

26 GW, and therefore, a region which resembles the "SF" is visible on the prior surface, as shown in blue in Figure 6.12. Moreover, as TEDS-Net predicts the field ($\mathbf{\Phi}_m$) between each individual's CP prediction ($\hat{\mathbf{Y}}_m$) and the prior ($\mathbf{P}$), labels can be moved between the two spaces. Therefore, I labelled the inner SF boundary ($\mathbf{SF}_{\text{Prior}}$) on our prior shape ($\mathbf{P}$), as shown in Figure 6.13. Then, for each individual, i, the prior SF labels were deformed into the individual space, using the learnt TEDS-Net deformations: $\mathbf{\Phi}_i(\mathbf{SF}_{\text{Prior}})$, returning the SF base for that individual. To measure the distance of the SF boundary from the skull, I used skull masks ($\mathbf{SK}_i$) that were generated in previous work by registering a pre-existing MRI Fetal atlas with skull masks to our ultrasound atlas [43, 134], and propagating the masks back to each individual [43]. To calculate the closest point of the skull to each voxel inside the skull using the Euclidean distance transform:

$$DT(\mathbf{SK}_i) = \min_{\mathbf{p}\in\mathbf{SK}_i} \sqrt{\sum_{j=1}^{3}(q_j - p_j)^2} \tag{6.1}$$

where $\mathbf{q} = q_1, q_2, q_3$ represent locations of background voxels and $\mathbf{p} = p_1, p_2, p_3$ are locations of other voxels inside the skull. To convert the distance from voxels to mm, $DT(\mathbf{SK}_i)$ was multiplied by the voxel spacing.

The Euclidean distance values at the warped SF positions were then found: $SD = DT(\mathbf{SK}_i) \times \mathbf{\Phi}_i(\mathbf{SF}_{\text{Prior}})$. An overview of this pipeline is shown in Figure 6.13. To characterise the SF base depth, from SD I calculated: the mean and median depth, to represent the average depth, and the 5^{th} and 90^{th} percentile, to report the shallowest and deepest parts, respectively, whilst adjusting for outliers.

Validation This method is based upon the assumption that the learnt deformation fields will warp the SF region in the prior ($\mathbf{SF}_{\text{Prior}}$) to the SF of an individual ($\mathbf{SF}_i$): $\mathbf{SF}_i == \mathbf{\Phi}_i(\mathbf{SF}_{\text{Prior}})$. Despite the prior shape resembling a cortical plate surface, this correspondence is not guaranteed through deformations. Therefore, to validate our method and to assess whether the warped position of the SF can be used to measure SF depth, I compared manual SF labels to our automated method. To do this, the inner SF boundary was labelled in 5 volumes across gestation and

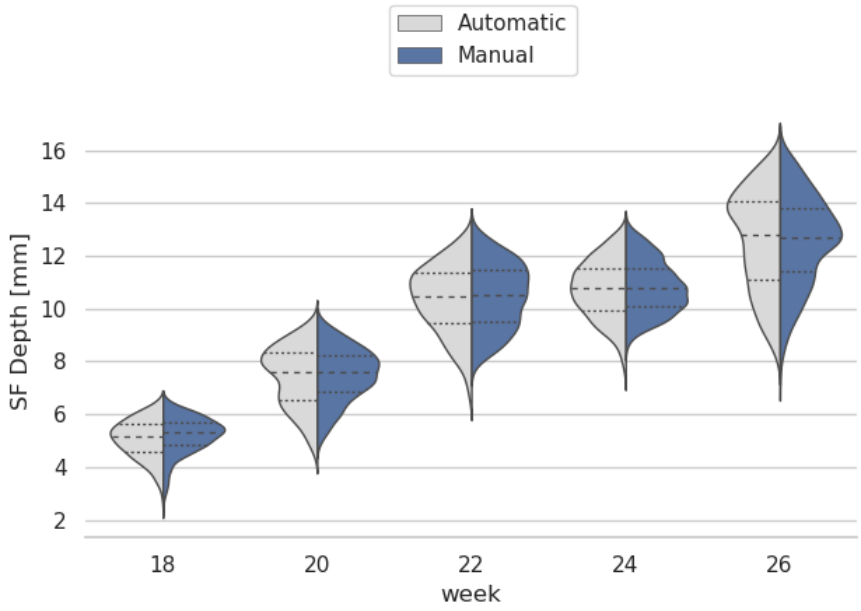


Figure 6.14: Comparing the SF depths measured for manually obtained SF labels and from our automated method.

the corresponding depths at those locations were calculated. Figure 6.12 shows an example of the manual SF label on a CP, compared to the SF labelling on the prior and the warped SF label for the same individual.

The SF depths measured from the manual SF labels and the deformed labels are shown in Figure 6.14, across gestation. The manual labels were performed for fetuses at 18, 20, 22, 24 and 26 GW. As the SF changes dramatically over this age range, it was important to ensure that the automated method was accurate across gestation and not just at the ages that most resemble the prior (e.g. 26 GW). Between the two methods, the distributions of SF depths show a high correspondence, with the mean SF depth being found to neatly align. This is a highly encouraging result, as it suggests that the deformation fields warp the SF region of the prior onto the SF on an individual's CP, thus validating our automated method.

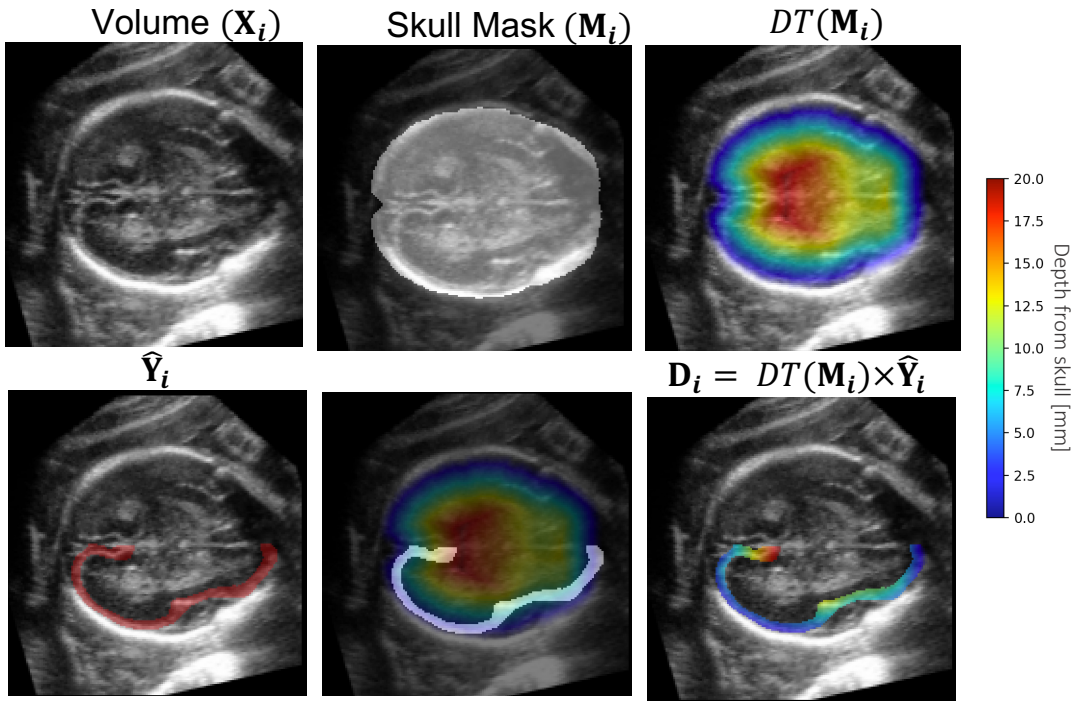


Figure 6.15: Pipeline to measure the depth at each location on the predicted cortical plate. First, the closest Euclidean distance from the skull to each voxel within the skull is found ($DT(\mathbf{M}_i)$). Next, the CP volume ($\hat{\mathbf{Y}}$) is multiplied by the distance map, to find the distance from the skull to the CP.

6.5.2 Local Properties

6.5.2.1 Surface Depth

To measure the local depth of the predicted cortical surface (e.g. the depth from the inner skull to the inner cortical boundary), I used a similar method as described for measuring SF depth. $DT(\mathbf{SK}_i)$, described in Equation 6.1, was multiplied by the CP segmentations $DT(\mathbf{SK}_i) \times \hat{\mathbf{Y}}_i$, generating a volume containing the CP depth: $\mathbf{D}_i$. An overview is shown in Figure 6.15.

6.5.2.2 Thickness

Thickness is traditionally measured by finding the shortest distance between a point on the outer CP boundary and a point on the inner boundary, as discussed in Section 3.3.2. However, as I do not have a visible separation between the two boundaries, I defined thickness as the distance between two perpendicular points (s and $s^{\perp}$) on the cortical plate boundary. To calculate this, I computed the orthogonal

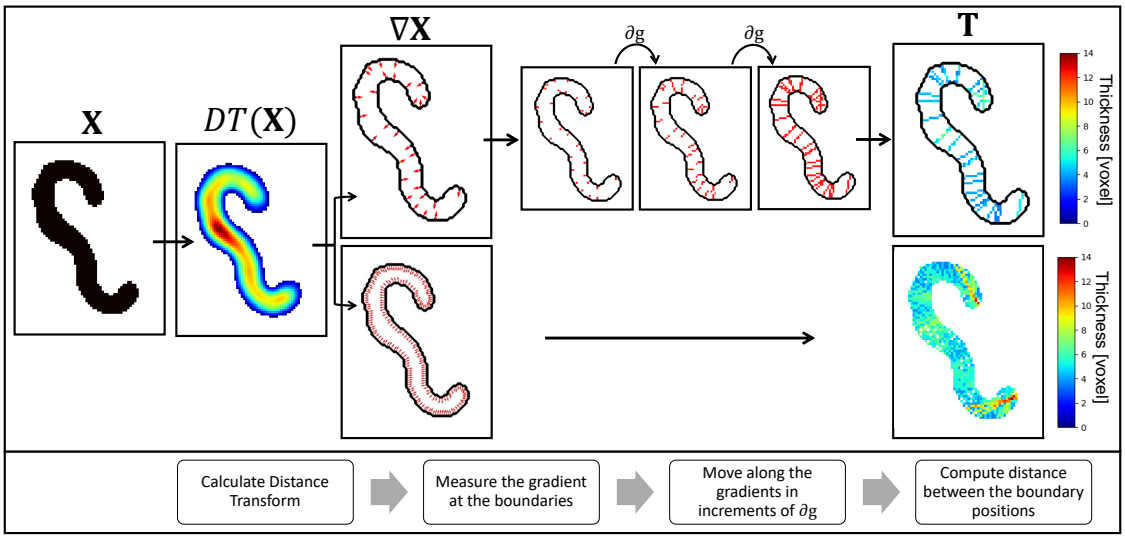


Figure 6.16: An overview of the thickness measurement pipeline on a simple 2D example ($\mathbf{X}$). First, the closest distance ($DT(\mathbf{X})$) from the edge of $\mathbf{X}$ to each pixel inside $\mathbf{X}$ is found. Next, the gradient ($\nabla\mathbf{X}$) at the boundary of $\mathbf{X}$ is found. The thickness at each boundary point is then found by moving along the gradient in steps of ∂g until a new boundary point is reached.

projection at each location on the surface and found the closest boundary point along the gradient vector. However, this method does not guarantee that the mappings are one-to-one, such that no point has multiple points mapped to it, and it also does not guarantee invertibility. These effects are more pronounced in heavily convoluted volumes (such as the adult brain) and will be largely reduced in surfaces with a smaller radius of curvature, (such as the fetal brain). Therefore, I implement this method but with appropriate smoothing to compensate for such effects and with awareness of its limitations.

Figure 6.16 shows an overview of the thickness measurement method in a simplified 2D example. Firstly, the contour of the cortical plate is found by morphological eroding of the binary CP segmentation, and subtracting the resulting volume from $\hat{\mathbf{Y}}_i$. Using a distance transform of the cortical plate labels ($DT(\hat{\mathbf{Y}}_i)$), the unit gradient at each position on the contour was found, generating a vector map $\nabla\hat{\mathbf{Y}}$. For each position along the contour, $\mathbf{s}_c$, the closest perpendicular contour point $\mathbf{s}_c^{\perp}$ was found by travelling along the gradient, in steps of $\partial g = 0.5$ voxel. Once two perpendicular coordinates were found, the Euclidean distance (d_c) between them was measured. The two coordinates $\mathbf{s}_c, \mathbf{s}_c^{\perp}$ and each CP voxel the algorithm

travelled through were then assigned d_c. This process was then repeated for all points along the contour. If any voxel was assigned multiple values of d, the average was calculated after all values were known. The thickness measures were then smoothed using a Gaussian filter with $\sigma = 1$, generating a final thickness volume $\mathbf{T}_i$, which had the same dimensions as $\hat{\mathbf{Y}}$: $160 \times 160 \times 160$.

6.5.2.3 Jacobian Determinant

From the learnt deformation fields in TEDS-Net ($\mathbf{\Phi}$), I can also calculate the Jacobian Determinants of the field, which show us how much the prior has been stretched or shrunk in order to segment the CP. This in turn provides information about the relative growth of the cortical plate over age. From each field generated by TEDS-Net, the Jacobian determinant ($\mathbf{J}_i$) was measured as described in Equation 2.9.

6.6 Healthy Population Averages

The methods discussed in Section 6.5 describe how the cortical plate properties can be measured from each segmented CP volume: $\hat{\mathbf{Y}}_i$. For the global properties, such as volume, a single global value is returned per scan, which can be directly compared between individuals. On the contrary, for the local properties, such as surface depth ($\mathbf{D}_i$), a value is given at each position in the cortical plate, where there is no precisely known anatomical correspondence between different individuals. Therefore, to compare multiple individuals and calculate meaningful averages, the local measures must be transformed into a common reference space. This is commonly done by using an average fetal brain shape or atlas [41, 175, 207]. In this section, I first describe the methods used to average the local properties in a reference space, before presenting the characteristics across our healthy test subset.

6.6.1 Methods

In this investigation, I explore the use of two reference spaces, which can be used as surface templates to both averages and visualise the healthy population properties.

In order to average individual properties in a common reference space, an alignment transformation is required between the space and each individual.

The reference spaces used are as follows:

1. **TEDS-Net Prior:**

   In TEDS-Net, a transformation field ($\mathbf{\Phi}_i$) that deforms the prior ($\mathbf{P}$) into an individual's CP label ($\hat{\mathbf{Y}}_i$) is learnt, where $\hat{\mathbf{Y}}_i = \mathbf{\Phi}_i(\mathbf{P})$. As $\mathbf{\Phi}$ is invertible, the TEDS-Net prior shape can be used as a reference space and template for the calculation of the average properties across the individuals. To do this, each property (e.g. depth, $\mathbf{D}_i$) is simply transformed into the prior space using the inverted field ($\mathbf{\Phi}^{-1}$): $\mathbf{D}_{P,i} = \mathbf{\Phi}_i^{-1}(\mathbf{D}_i)$. This represents a substantial benefit of the TEDS-Net approach, as an additional registration step is not required. As the same prior was used for each GW, the same base geometry is used for each age, and therefore, longitudinal changes at each vertex point on the base surface can be compared.

2. **Atlas Template:**

   For the second reference space I used the ultrasound atlas, described in Section 6.4.1.1. As the volumes used throughout this work were originally labelled using atlas-based label projection, the transformations (T_{Ri}) between the CP of the weekly atlases ($\mathbf{S}_R$) and the corresponding individual CPs are known, as shown in Figure 6.3. Therefore, to move between the spaces I use: $\mathbf{D}_R = \mathbf{T}_{Ri}(\mathbf{D}_i)$. Using weekly CP atlas labels as a template means there is a different template for each week that more closely resembles the true morphology of the individual. However, each vertex on these surfaces can no longer be directly compared, unlike the previous case with the TEDS-Net prior.

Both reference spaces have different advantages, and by computing the two I can compare their corresponding average measures, to ensure they are both accurately representing the population. For both methods, once all of the individual CP property matrices $\{\mathbf{D}_i, \mathbf{J}_i, \mathbf{T}_i\}$ were transformed into the reference space, the

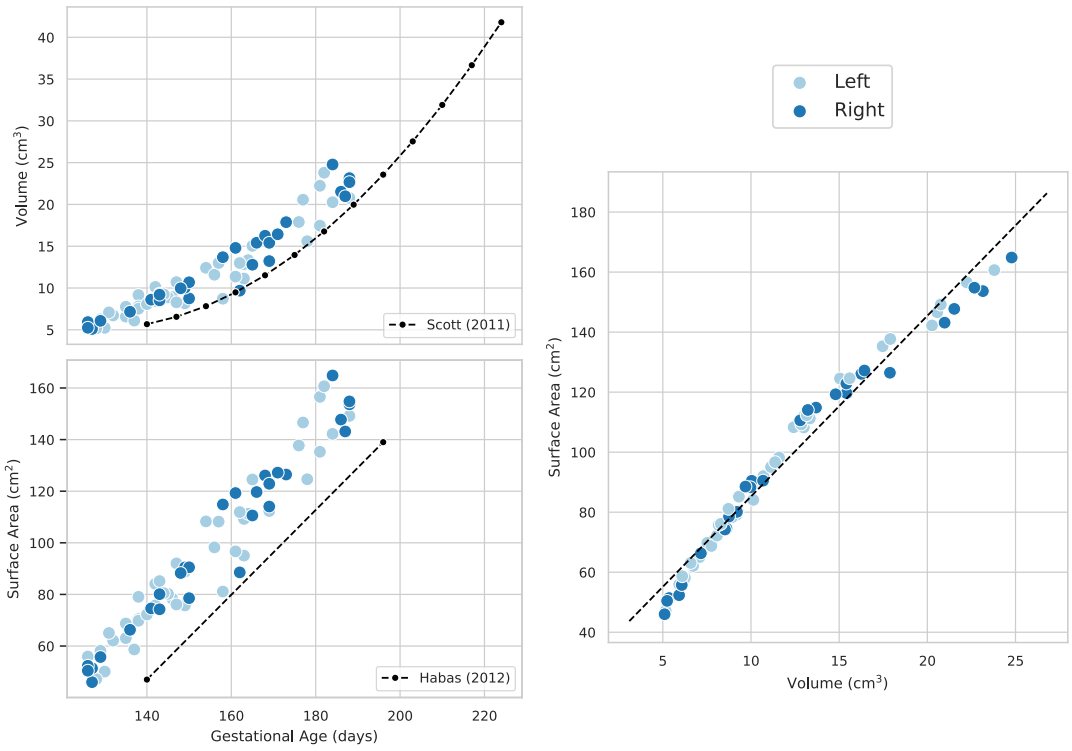


Figure 6.17: Measures of cortical volume and surface area of the CP from a single hemisphere, measured automatically from the TEDS-Net predictions. The measurements are reported over gestation and separated into hemisphere sides (e.g. left and right). Note that, in Scott et al. [175] the volume of both hemispheres combined was reported, so the measurement presented here has been halved, whereas Habas et al. [134] reported the surface area of the inner boundary of both hemispheres.

average for each position was calculated at each gestational week, resulting in an average volume at each GW: $\left\{\bar{\mathbf{D}}_{GW}, \bar{\mathbf{J}}_{GW}, \bar{\mathbf{T}}_{GW}\right\}$. These averages were then converted to surface meshes $\mathcal{M}$, by computing the surface mesh of the template surface and finding the average measure at each vertex position.

6.6.2 Results and Discussion

Here I present the morphological properties derived from the TEDS-Net cortical plate segmentations, compared to previously reported cortical plate properties.

6.6.2.1 Global Measures

Surface Area and Volume Figure 6.17 shows the measured surface area and volume from the segmentation CP volumes (as described in 6.5.1). The volume

measurements are compared to an MRI study [175] that measured the volume across both hemispheres. To compare the volume measurements between the two studies, the MRI measurements were halved to represent an average of the two hemispheres. Both sets of results show that CP volume increases with age, following a similar quadratic pattern, that maintained a strong rate of growth. The volumes derived from MRI were found to be smaller but close to our measurements. Previous literature has shown that differences between the two modalities are expected, likely due to differences in image acquisition, with measurements derived from ultrasound often found to be larger [251], which may explain the differences found here.

The cortical plate surface area was found to linearly increase with age, a trend previously found in Habas et al. [134] for the same gestational window. In Habas et al. [134], the surface area of the *inner* CP boundary across *both* hemispheres was measured from MRI volumes. This methodology differs from ours as I measured the *whole* surface area (inner and outer) of a *single* hemisphere. Therefore, between these two methods, I do not expect the surface area to be exactly equivalent, as the inner CP surface area is less than the outer, thus the surface area reported in Habas et al. [134] should be less than ours but following the same trend. As shown in Figure 6.17, the MRI-derived measures are lower than the TEDS-Net CP, as expected.

Cortical folding allows the cortical surface area to scale almost linearly with volume, increasing the surface area whilst being constrained by the skull volume. Increased surface area is often associated with increased cognitive abilities, as it allows for increased connectivity [252]. The relationship between the two across our dataset is shown in Figure 6.17. Linear regression analysis revealed a strong correlation between the two measures ($R^2 = 0.97, p = 1.3 \times 10^{-55}$).

Sylvian Fissure Depth The deepening of the SF across gestation is shown in Figure 6.18, where for each individual the $5^{\text{th}}, 50^{\text{th}}$ and 95^{th} percentile values of SF depth are reported. Our SF depths are compared to growth charts of 2D SF depths, which were generated from the wider INTERGROWTH-21^{st} dataset [169]. Between the two methods, there is a high overlap, and both methods show the SF deepening

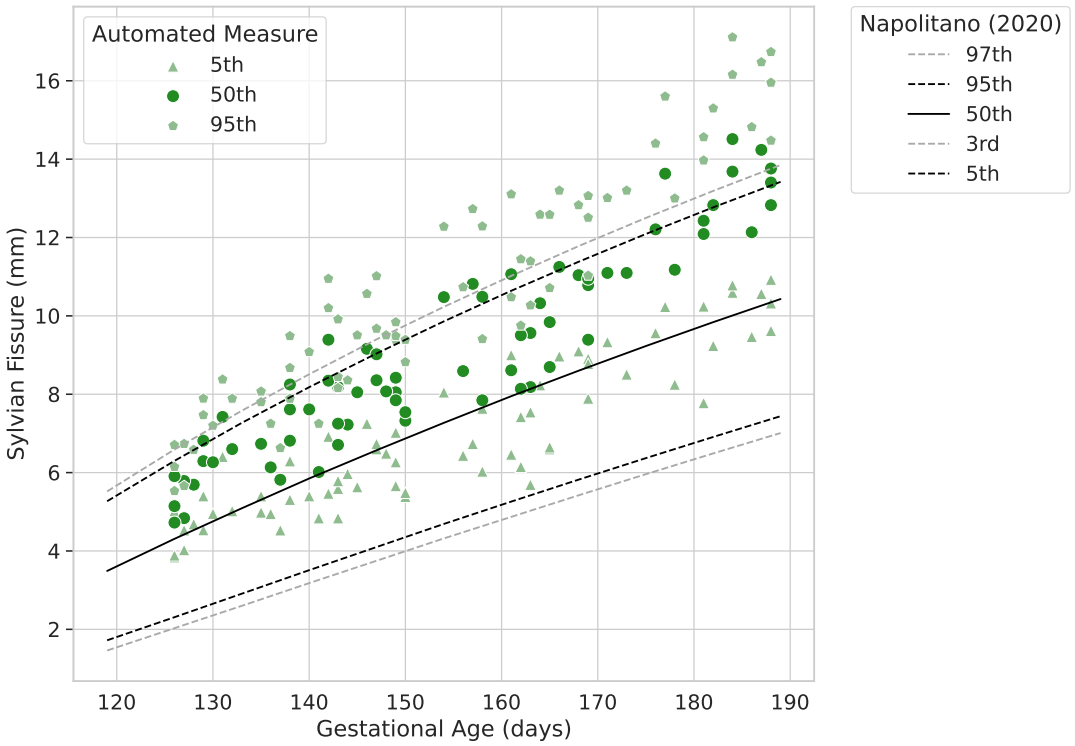


Figure 6.18: The Sylvian Fissure (SF) depth measured using our automated method. From the set of SF depths for each individual, the 5^{th}, 50^{th} and 95^{th} percentile was reported. In Napolitano et al. [169], the SF depth was measured in 2D from the larger INTERGROWTH-21^{st} dataset, where the average (50^{th}), upper (95^{th}, 99^{th}) and lower bounds (5^{th}, 3^{rd}) of healthy SF depth were found, as shown here.

at a similar rate. Our automated 3D measures are found, on average, to be larger than the 2D measures. Differences between these two methods were expected, as the measured distances are fundamentally different between 2D and 3D. Moreover, the location of the SF was also inconsistent between the two methods as in Napolitano et al. [169] the SF depth is consistently measured at the same position (as discussed in Section 3.2.3), whereas in our method I measured depth at every position along the inner boundary of SF, and computed the averages and percentiles.

Hemisphere Asymmetry As there are known asymmetries in the cortical plate between the two hemispheres, with the left hemisphere often found to be larger [188, 253–257], our results are separated by hemisphere. On visual inspection, the left and right seem consistent for both surface area and volume. However, to probe

Table 6.2: Statistical test results (p-values) between the two hemisphere sides before multiple-comparison correction. None of these tests survived Bonferroni corrections.

	p values (L-R)		
	Volume	Surface Area	SF Depth
Gestation	0.35	0.92	0.17
18 GW	0.20	0.14	0.25
19 GW	0.49	0.51	0.05
20 GW	0.52	0.27	0.02
21 GW	0.48	0.49	0.23
22 GW	0.10	0.09	0.24
23 GW	0.80	0.88	0.63
24 GW	0.03	0.01	0.85
26 GW	0.48	0.76	0.36

whether they differ statistically, a set of statistical tests were performed. Firstly, to assess if the overall trend between the left and right hemispheres was significant, a general linear model (ANCOVA) was employed, which evaluates whether a dependent variable (e.g. volume) varies significantly between two independent factors (e.g. side) whilst controlling for the effects of a 'nuisance' or covariant variable (e.g. age), which affects the trend. Both surface area and volume were found not to differ significantly between the two hemispheres, with p-values of $p = 0.9$ and $p = 0.3$, respectively.

In previous literature, cortical asymmetries have been found to vary with gestational age [188, 255], with Rajagopalan et al. [188] finding regions of asymmetry shifting posteriorly with age and Kasprian et al. [255] showing regional asymmetries present from 20GW. Therefore, to investigate whether such differences are present at different periods of gestation, an unpaired t-test was performed at each gestational week between the two hemispheres' sides and corrected using Bonferroni correction [231]. No significance was found for either volume or surface area at any gestational week, implying for these measurements no strong asymmetries were found across gestation, as shown in Table 6.2.

Asymmetrical Sylvian fissures were amongst the first discoveries of asymmetry in the fetal cortex [116, 258, 259]. This asymmetry is thought to be liked to handedness and right-left functional differences [116, 259], and is thought to be present in Neanderthal man and is seen in some apes [259]. From post-mortem

studies, Sylvian fissure asymmetry was found to start around 25 GW [126], whereas, in more recent in-utero imaging studies, morphological asymmetries are found as early as 20 GW [130]. However, most of these methods report differences in shape, but not depth [171]. To investigate whether the depth of the Sylvian fissure differs significantly between left and right hemispheres, the ANCOVA and weekly unpaired t-tests (as performed for volume and surface area), were performed between the SF depths and no significant results were found ($p > 0.05$). An overview of the uncorrected test results between the hemisphere sides is shown in Table 6.2.

6.6.2.2 Local Measures

Sulci Depth The average sulci depth projected on both the Atlas and Prior templates is shown in Figure 6.19 and Figure 6.20, respectively. These figures look at the depth on the inner cortical surface, whereas Figure 6.21 and Figure 6.22 shows the depth on the outer cortical surface.

Between the two base surfaces, there is a high correspondence between the reported depths. For example, at 18 GW inner CP depth was found to vary little across the surface (∼2.5mm), with a slight increase around the Sylvian fissure (∼5mm). This pattern was observed on both the atlas and prior geometry. Additionally, between 20-24W, an increase of depth is localised at the posterior of the Sylvian fissure seen on both the atlas and prior geometry. These findings are promising and show that both methods return similar averages.

The Sylvian fissure is shown to be the deepest surface on both the inner and outer cortical boundary, which is expected for this gestational age range [188]. At 26GW, the deepest part of the Sylvian fissure is shown to be between $15 - 20$mm, corresponding well to the Sylvian fissure depth reported in Figure 6.18 [169]. The depth of the inner cortical surface surrounding the Sylvian fissure remains similar between $18 - 24$ GW, with the depth increasing for 25 and 26 GW. Conversely, the depth of the outer surface remains consistent across the entire gestational range, implying that cortical thickness is increasing.

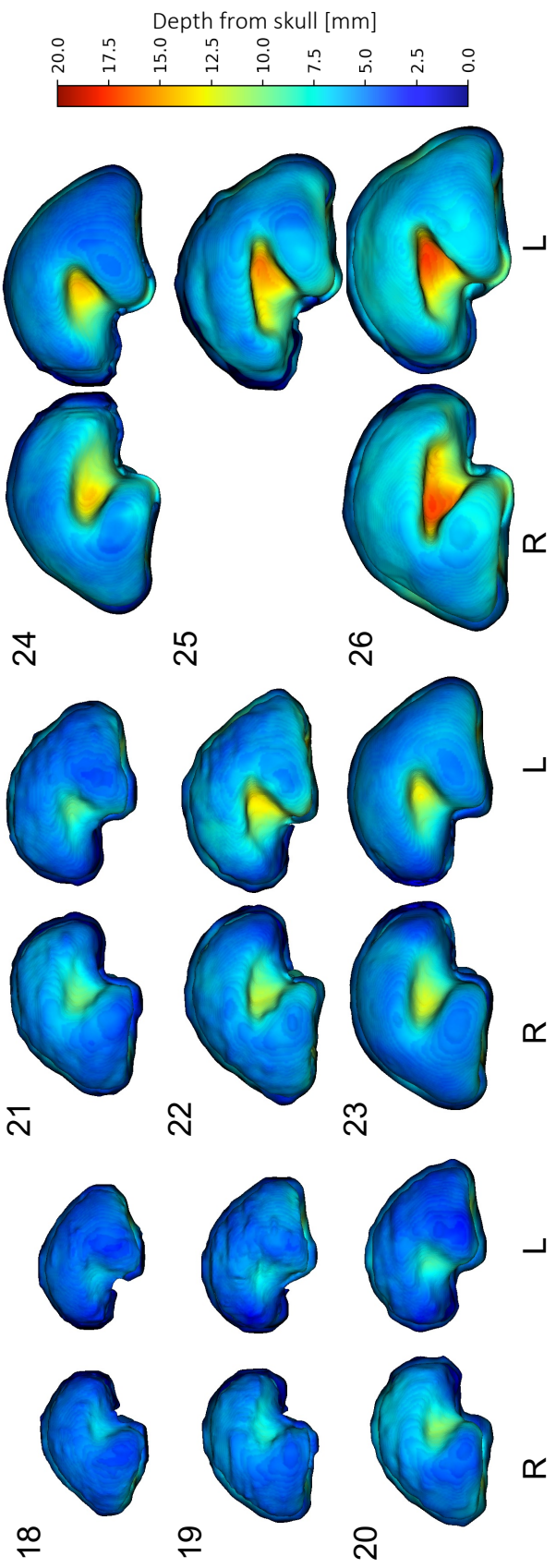


Figure 6.19: The average depth of the inner cortical plate boundary projected onto the atlas CP template for each gestational week. Note that, there was only one volume of the right hemisphere at 25 GW, so do not present an average for this age and side.

Figure 6.20: The average depth of the inner cortical plate boundary projected onto the TEDS-Net prior.

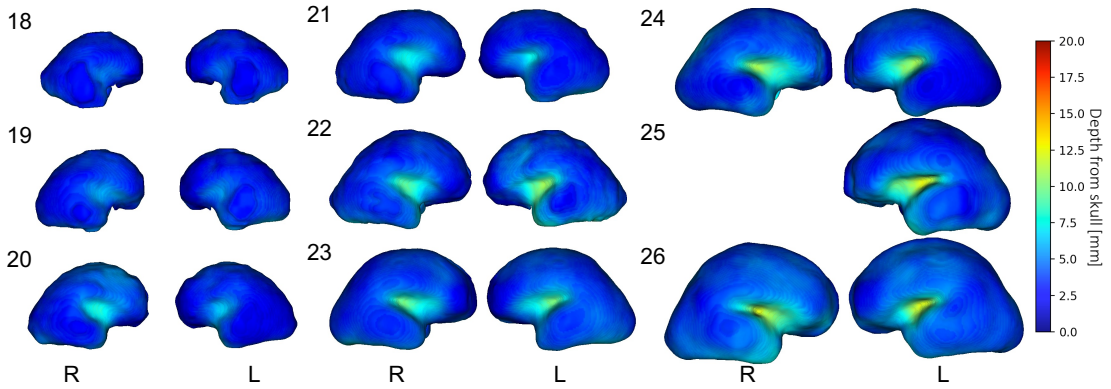


Figure 6.21: The average depth of the outer cortical plate boundary projected onto the atlas CP template for each gestational week.

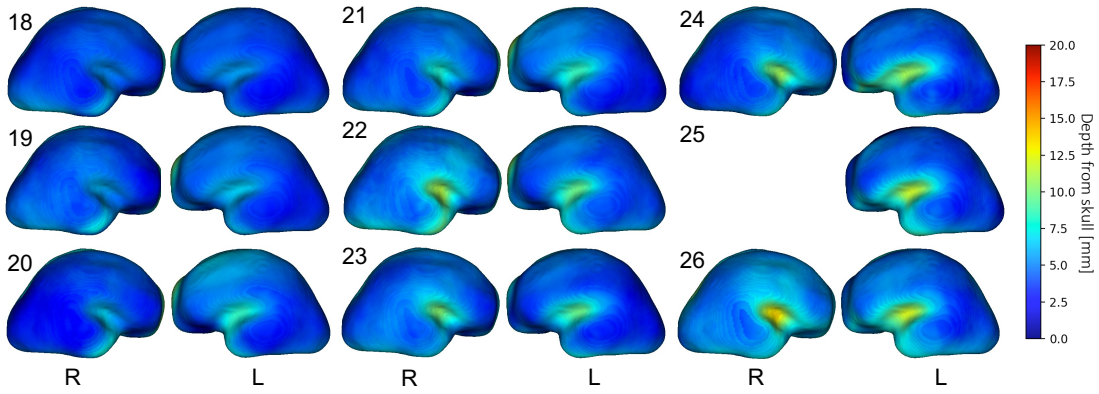


Figure 6.22: The average depth of the outer cortical plate boundary projected onto the TEDS-Net prior.

As discussed in Chapter 3 (Section 3.1 and Section 3.3.2), the central sulcus begins to emerge and deepen from around 25 GW and seen in MRI and ultrasound imaging [133, 176, 212]. However, in both Figure 6.20 and Figure 6.19, there is no sign of this fissure. On initial inspection of the predictions ($\hat{\mathbf{Y}}_i$), as shown in Figure 6.23, it was clear this fissure was not identified by TEDS-Net. I concluded that this was likely due to topology-preserving modifications within TEDS-Net, which limit the deformation's flexibility, making it challenging to deform the prior to fit the tight and small fold. However, when inspecting the corresponding label ($\mathbf{Y}_i$), it had also incorrectly labelled this sulcus. As $\mathbf{Y}_i$ was produced by atlas-based propagation, on inspecting $\mathbf{S}_{\mathrm{R}}$ and $\mathbf{I}_{\mathrm{R}}$ at 26 GW, it was clear that this sulcus is not visible on the atlas, and therefore, the central is often unlabelled in the propagated

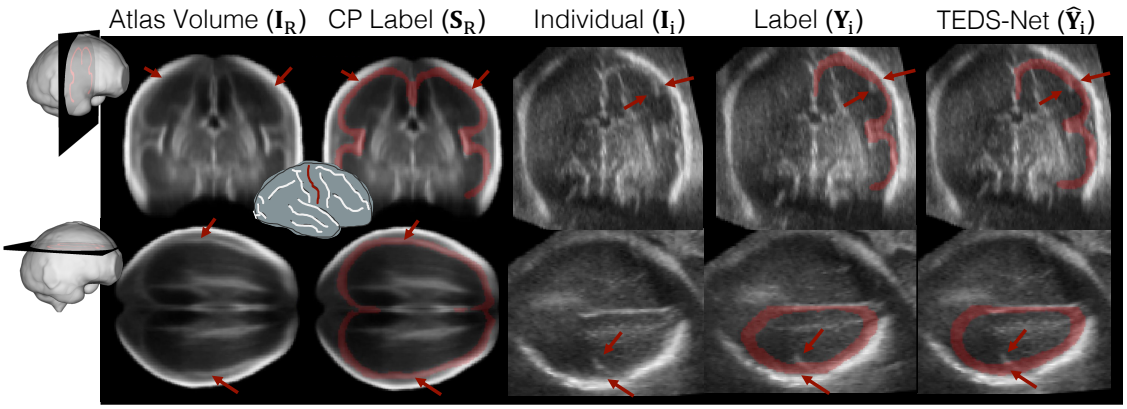


Figure 6.23: An overview of the central sulcus at 26 GW. The first column shows the ultrasound atlas at 26 GW and its corresponding CP label, as described in Section 6.3.2. From the volumes, the central sulcus is not clearly visible, with the arrows showing where the sulci should be. The next column shows a fetal brain at 26 GW, where the sulcus is visible, albeit not fully formed. However, as the central sulcus is not visible in the atlas labels ($\mathbf{S}_R$), it is also missing in the propagated labels ($\mathbf{Y}_i$). The sulcus is also missing in the TEDS-Net prediction ($\hat{\mathbf{Y}}_i$).

labels. Therefore, it is likely that the central was not visible in the sulcal depth renderings, due to this mistake in labelling. However, this does not rule out the hypothesis that TEDS-Net may struggle to segment tightly folded structures.

Thickness The average cortical thickness across our healthy population is shown in Figure 6.24 on the Prior template. The cortical thickness remains relatively uniform across gestation, with the majority of the plate between $1-2$mm deep, which is in keeping with the MRI literature for a similar gestational age range [212, 220].

As shown in Figure 6.24, with age the SF becomes thicker and is shown to be the thickest part of the cortical plate. This is likely due to SF operculization, where the SF begins to fold over itself, increasing its thickness. As shown in Pistorius et al. [168]'s Sylvian fissure grading chart (see Figure 3.8), in healthy pregnancies the Sylvian fissure begins to form an acute angle on the posterior side from around 24 GW. Figure 6.25 and Figure 6.26, show the thickness projected onto the atlas templates at two slices in the axial and coronal plane, respectively. From these figures, it can be seen that the posterior of the Sylvian fissure begins to fold over itself, shown with an increase of thickness from around 22 GW (shown with yellow arrows),

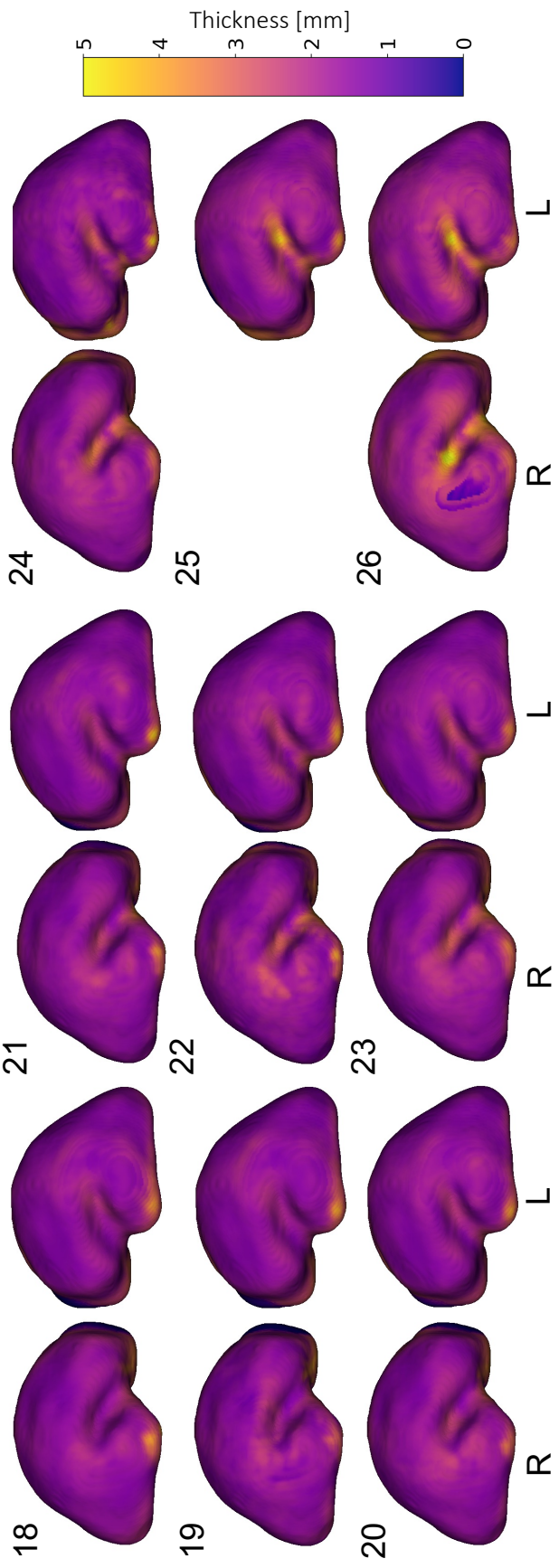


Figure 6.24: The average thickness values at each location across the CP at each gestational week projected onto the prior shape surface.

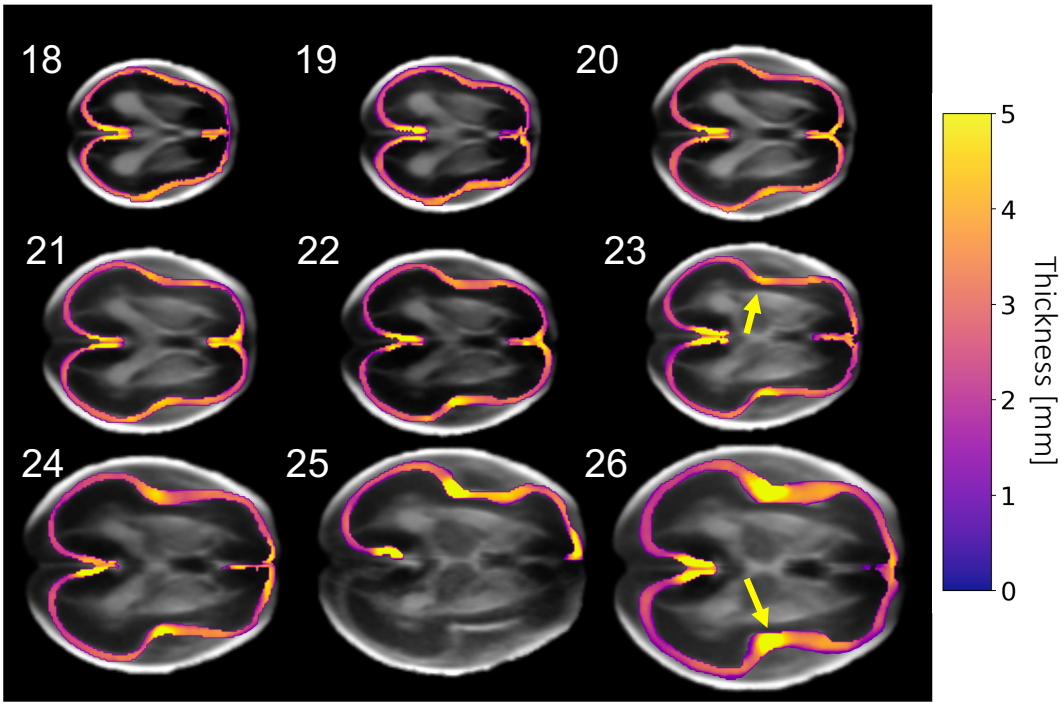


Figure 6.25: The thickness values averaged in the atlas space and shown in the transventricular axial plane. The yellow arrows show regions of Sylvian fissure opercularization.

which is two weeks earlier than Pistorius et al. [168] described. This difference might be a consequence of the difference between the 2D and 3D analyses. The Sylvian fissure is thickest at 26 GW, an age that Pistorius et al. [168] would expect the fissure to have gone through $> 50\%$ operculization, resulting in the SF doubling over itself.

Relative Growth The average Jacobian determinant mapping at each gestational week is shown on the prior surfaces in Figure 6.27 and the volume in Figure 6.28. The Jacobian determinant describes how much the prior (which was based on the 26 GW atlas label) was stretched or squeezed to produce the segmentations at each gestational week. Both the volume and surface representations are shown, as surface renderings can be misleading since they may just describe changes in thickness. The volume measurements are overlaid onto the 26 GW atlas for landmark references, however, it should be noted that the prior is not equivalent to the 26 GW CP atlas label but instead is a smoothed version, and therefore the overlap with the underlying CP is not exact.

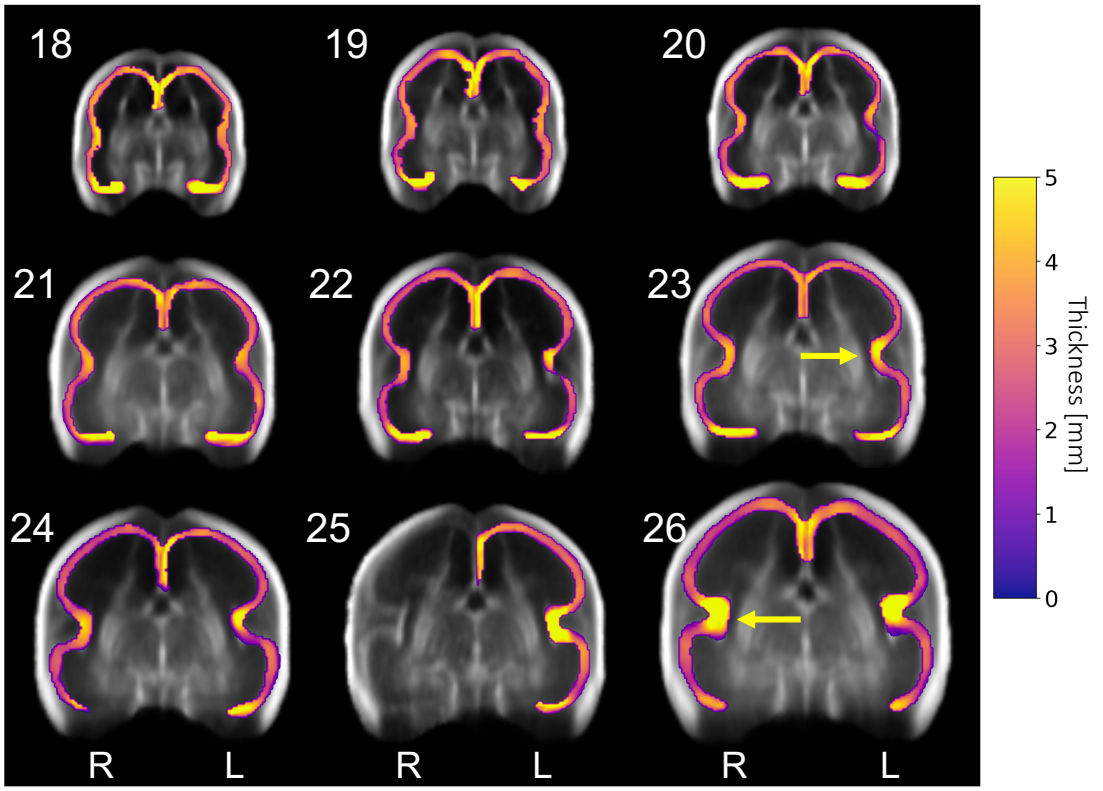


Figure 6.26: The thickness values averaged in the atlas space and shown in the coronal plane. The yellow arrows show regions of Sylvian fissure opercularization.

The most striking change in the cortical morphology across this period of the second trimester is the increase in size. As the prior used for this experiment most closely resembled the average CP at 26 GW, it is expected that the fields for younger ages will be compressing the prior. This trend is seen in the figures, where the prior is shrunk for all ages except 26 GW, which is stretched in certain regions. The shrinkage is more pronounced for lower gestational ages, showing that younger ages are associated with transformations that change the size more, which is expected. A notable increase in prior size is shown in the posterior Sylvian fissure region at 26 GW (shown in Figure 6.27 and Figure 6.28). This increase may be associated with the increased thickness of the Sylvian fissure, shown in Figure 6.25, where the fissure begins to fold over itself, requiring an increase in prior thickness. There is also an increase in growth in the dorsal region at 26 GW, as shown in Figure 6.28. At this gestational period, the central, post-central and pre-central fissures should

Figure 6.27: Visualisation of the average Jacobian determinants on the prior, showing the regions stretched and squashed to produce the cortical plate segmentations for each gestational week.

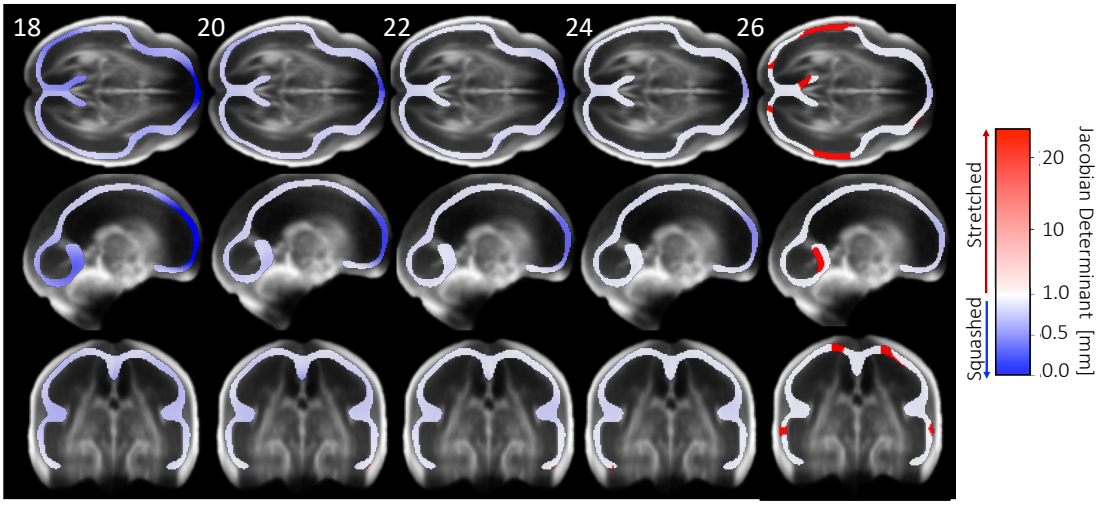


Figure 6.28: Visualisation of the average Jacobian determinants on the prior, showing the regions stretched and squashed to produce the cortical plate segmentations for each gestational week. Each row shows a different brain plane, axial (top), sagittal (middle) and coronal (bottom).

begin to emerge around this region, which could explain this growth [126, 129, 133].

6.7 Limitations

Across this chapter I have proposed an automated pipeline to segment and analyse the developing cortex, returning a set of expected morphological properties for healthy development. However, the methods are not without limitations.

This chapter has focused on characterising the cortical development in the second trimester, a period where the secondary fissures begin to emerge [126, 133]. Across this period, the cortical surface is still relatively smooth, with the majority of the folding happening in the third trimester, with the emergence of the tertiary fissures (as described in Section 3.1). However, the pipeline proposed in this section is likely to be unsuitable for more convoluted surfaces. This is partially due to the topology-preserving fields within TEDS-Net, which may not be flexible enough to delineate tight and small folds across the surface, as seen with the central fissure at 26 GW. Moreover, the current method used to measure thickness is not robust for highly convoluted volumes. Therefore, it is likely that 26 GW is the maximum age for which our proposed analysis is suitable.

When characterising the cortical development from MRI volumes, typically both hemispheres are delineated, allowing the separation of the inner and outer cortical plate boundary, as discussed in Section 3.3.2. The analysis (e.g. surface area or sulcal depth) is then reported across both hemispheres and usually from the inner cortical boundary. Due to the acoustic shadows in ultrasound volumes, the analysis in this chapter was only performed from a single hemisphere (the distal hemisphere) from each scan. Therefore, I was unable to separate the two boundaries and was only able to visualise the local changes on the inner boundary through mesh rendering algorithms. Moreover, this means that it was hard to directly compare our analysis to previous MRI measures. In future work, developing a method to separate the two cortical plate boundaries would be a priority, in order to facilitate comparisons and better visualisations. Further, if these boundaries were separate, sophisticated MRI analysis methods, such as measuring the thickness with electric potentials [213], could be utilised.

In this chapter, the CP was only characterised for the held-out test set, which was made up of 15% (70 scans) of the optimal INTERGROWTH-21$^{\text{st}}$ subset. In this subset, there was only a single scan at 25 GW with the right hemisphere visible and therefore, average local measures were not shown for this age and hemisphere side. Further, the subset of data used for training and testing in this chapter only represents about 2% of the total INTERGROWTH-21$^{\text{st}}$ dataset. Therefore, to better represent the population, the characterisation of the CP should be performed on a larger subset. Appendix D.2 shows preliminary work on applying the pipeline to a larger population of $N = 1,643$ of INTERGROWTH-21$^{\text{st}}$ scans.

The analysis pipeline designed in this chapter has been proposed for routine ultrasound scans collected in the second trimester. However, although the volumes used throughout were routinely obtained they required pre-processing before being applied to the proposed pipeline. Namely, the volumes were aligned to a common coordinate space and resampled to a spacing of 0.6 mm^3. Recent work has begun to automate the process of ultrasound image alignment using deep learning [260].

Therefore, using such methods our pipeline could be integrated into a larger toolbox to facilitate analysis directly from ultrasound outputs.

6.8 Conclusion

The contributions of this chapter can be separated into two parts. First, I have integrated TEDS-Net into a complete medical imaging pipeline, showing the importance of topology for downstream medical imaging tasks. Second, from the analysis of our topologically correct CP segmentations, I have presented the first 3D properties of this developing surface from ultrasound images.

Cortical plate segmentations from ultrasound volumes can be an extremely challenging task, due to large acoustic shadows obstructing regions of the CP. Therefore, as TEDS-Net was able to anatomically guide the segmentation in regions of poor contrast, it was able to excel in this task, generating topologically-correct surfaces in all cases. In this chapter, I have shown why topology is important for downstream analysis, e.g. to measure both global and local surface properties. Further, I have shown how the prior and learnt deformation fields within TEDS-Net can be utilised to generate average properties over the population, providing a common space template.

To further validate the segmentations produced by TEDS-Net, I analysed the surfaces and compared the properties to previous work. To do this, I designed new methods to analyse a single hemisphere's CP properties. Both local and global CP properties were found to have high correspondence to previous 2D ultrasound and 3D MRI studies. Moreover, to the best of our knowledge, this is the first 3D analysis of the cortical plate performed from ultrasound.

7 Conclusion

7.1 Summary

Anatomical segmentations are often required in medical imaging pipelines, to extract relevant information about a structure of interest. Therefore, to ensure precise anatomical measurements the segmentations must accurately represent the true anatomy. To accurately evaluate a segmentation's performance a wide variety of metrics is required. Throughout this thesis, I have frequently found that the best-performing network based on one metric (e.g. Dice) was not the best-performing on another (e.g. HD), thus showing that measuring a single metric often does not provide a complete picture. Further, I have discussed how measures of the overall shape, e.g. topology, are often neglected in segmentation evaluation, and how traditional metrics do not reflect large topological errors. Moreover, I have shown how the optimisation and extraction of features from CNNs, the most popular medical image segmentation method, can contribute to incorrect segmentation topology.

In medical imaging pipelines, segmentations are often required as an intermediate step, used to automate measures of the structure (e.g. volume or perimeter). However, if a segmentation has the incorrect topology, extracting meaningful measures can become extremely problematic. In this thesis, I have proposed a deep learning

segmentation method (TEDS-Net) that strongly enforces segmentations to have the correct topology, whilst remaining competitive to SOTA baselines on typically used metrics (e.g. Dice and HD). TEDS-Net utilises topology-preserving fields that are generated using traditional methods (e.g. the Scaling and Squaring approach) but with additional modifications to try and overcome the limitations of discrete sampling and storage. Through extensive hyper-parameter exploration, I have shown that under or over-smoothing can negatively impact both the topology and pixel-wise performance of the segmentation, and found a stable range of parameter values that resulted in no folding voxels and promising segmentation performance.

In this thesis, I have shown how TEDS-Net can be used to segment a range of structures from a variety of scenarios: from multiple paired priors to multi-structure segmentation, and from different imaging modalities in different dimensions. Further, I illustrated how it can be used to segment a 3D volume, whilst only training on a few labelled slices. Finally, TEDS-Net was integrated into a large medical imaging pipeline, to characterise the cortical plate development from routine ultrasound scans.

7.2 Contributions

The main contributions in this thesis are summarised below:

1. **Development of a Topology Preserving Segmentation Method**
   My first contribution is the development of TEDS-Net, an automated deep-learning segmentation network. TEDS-Net incorporates prior knowledge of the known topology of an anatomical structure of interest, through a prior shape that is designed by the user. TEDS-Net learns two topology-preserving deformation fields, one generated at a low resolution used to localise the prior, before a second higher-resolution field is used to fine-tune the prediction. As the topological guarantees of continuous fields begin to break down when working in the discrete domain, TEDS-Net employs additional modifications to try and combat some of the discrete violations. These modifications

mainly involved removing large local gradients within the fields and decreasing interpolation artefacts by reducing the grid spacing. Gaussian smoothing was applied between each h composition layer within the Scaling and Squaring implementation, with the amount of smoothing controlled by σ, further, the fields were encouraged to be smooth through a field regularisation term that penalised large gradients, whose contribution to the total loss was controlled by α, β. Within TEDS-Net, h, σ, α, β, were all found to play a large effect in preventing folding voxels within the generated fields and a range of suitable values for each parameter was found.

2. **Achieving Topologically-Accurate Segmentations from a Range of Popular Segmentation Tasks**

   As topology is rarely assessed in segmentation analysis, the extent of topological mistakes in popular segmentation tasks is unknown. I applied TEDS-Net to a range of popular open-source segmentation datasets and compared them against SOTA baselines. First, using the open-source ACDC cardiac dataset [222], TEDS-Net was applied to myocardium segmentation, where its segmentation performance was competitive with four baselines, with no network significantly outperforming the others across four segmentation metrics. However, TEDS-Net outperformed all on topology, achieving 100% correct topology, compared to the next best performing baseline, TopoLoss, which achieved 95%. TEDS-Net was then extended for multi-structure segmentation using the same cardiac dataset, but now also including the right and left ventricle. In this, I showed how topology can also be used in combination with other segmentation metrics to evaluate the relative position of different structures to one another. Next, TEDS-Net was applied to digit segmentation [27] by pairing each digit to a different prior, depending on the digit's assumed topology. Due to the natural variation in handwriting, the digits often deviated from the given topology, therefore, this task was a good opportunity to see how strictly TEDS-Net enforced correct topology when it would impede the Dice performance. TEDS-Net preserved the prior topology

in 99.55% of the cases, despite only 76% of ground truths having the same topology as the prior. Finally, TEDS-Net was applied to 3D liver segmentation from the decathlon dataset [235]. In this, I showed how TEDS-Net can be used to segment volumes using minimally annotated datasets, outperforming previous SOTA one-slice methods.

3. **Segmentation of the Cortical Plate from Ultrasound Volumes**
   Using TEDS-Net, I was able to segment the cortical plate from challenging ultrasound volumes. Delineating structures from ultrasound volumes can be challenging due to large acoustic shadows, which obstruct the underlying anatomy. Using the prior knowledge of the cortical plate shape, TEDS-Net was able to anatomically guide the segmentation in regions of shadows. To the best of my knowledge, to date, cortical plate segmentation in 3D has only been performed from MRI scans, and at older gestational ages. In this thesis, the cortical plate is extracted between 18 and 26 GW, a period of initial growth that aligns with the anomaly scan performed in routine prenatal care.

4. **Characterisation of Healthy Cortical Development from Ultrasound**
   From the topologically correct cortical plate segmentation, morphological properties of the developing structure were extracted. From each individual, global measures of volume, surface area and SF depth were performed, followed by local measures of thickness, relative growth and surface depth. These local measures were averaged across the healthy population by propagating the measures into a reference coordinate system. As TEDS-Net learns the deformation fields between an individual and the prior shape, the prior shape can be used as a common reference template, to project local properties on. From the test subset, I characterised the healthy cortical plate properties between 18 and 26 GW. To the best of my knowledge, 3D characterisation has not been performed from ultrasound scans, the modality of choice in prenatal care. Our reported measures fitted well with previous MRI studies.

7.3 Limitations and Future Work

Across this thesis, I have identified some limitations of the methods, however, such limits often provide opportunities for future work.

TEDS-Net Implementation

1. **Prior Knowledge of Topology**

   The TEDS-Net implementation requires prior knowledge of the anatomy of interest. More specifically, a prior shape must be designed with the same topology as the target structure. Therefore, TEDS-Net is unsuitable for tasks where the topology of a structure can vary between subjects or is unknown, e.g. tumour segmentation. Further, as the topology of a 3D structure can vary between 2D slices, for 2D segmentation careful preprocessing is currently required, to ensure only specific slices where the structure exhibits the desired topology are chosen. This can be extremely time-consuming, and often only possible if labels are already available. Moreover, as seen from the multi-structure cardiac segmentation, the orientation of a prior can make a large difference in performance. Furthermore, as shown in the MNIST example, often the input's true topology did not align with the assigned topology, which negatively impacted the other segmentation metrics. Therefore, it is clear improvements are required in the prior assignments. To overcome these limitations, one possible solution would be to include an additional network into TEDS-Net, tasked with learning the most appropriate prior for each given input. To achieve this, a set of priors of different characteristics and/or topologies could be designed to describe the range of topologies or orientations exhibited in the dataset. For example, in the multi-structure segmentation case, the additional network could choose which of the eight prior shapes would perform best for each input, removing the requirement to learn challenging rotations.

2. **TEDS-Net Prediction's *Smooth* Characteristics**

   To overcome the limitations of discrete sampling and storage, topology-preserving modifications are implemented in TEDS-Net, which increases the smoothness of the predicted fields. Although these modifications improve the topological performance of the segmentations, they also can result in overly smoothed segmentations, which risks losing the finer segmentation details. This unwanted effect has been shown throughout this thesis. For example, increased thickness was found in the myocardium segmentation and linked to the value of σ. Moreover, the liver renderings were overly smoothed compared to the blocky ground truth labels. As smooth segmentations are a direct consequence of the smoothing modifications, to directly overcome this effect, we need to reduce the requirement for smoothing within TEDS-Net. One possible way to achieve this would be to no longer store the predictions and priors at discrete intervals, but instead to describe the structures as meshes. In this, the vertex coordinates can be stored at continuous intervals, which may reduce the resampling artefacts and therefore the need for additional modifications.

3. **Convoluted Cortical Plate**

   In this thesis, the cortical plate was segmented using TEDS-Net between 18 and 26 GW. At 26 GW, the network was unable to accurately delineate the beginnings of the central sulcus, a small and tightly folded fissure. Although the ground truth labels did not accurately outline this structure, it is also likely that the topological modifications within TEDS-Net limit the flexibility of the field. More specifically, large differences in the field's gradients are required to deform the prior into such a fold, which the additional modifications within TEDS-Net try to prevent. Therefore it is likely that this is the upper age limit that TEDS-Net can accurately delineate before the smaller tertiary folds begin to emerge. In future work, attention should be paid to ensure that the ground truths accurately represent the smaller fissures so that we can

investigate if TEDS-Net can delineate such structures and therefore be used to segment the cortical plate at older gestational ages.

4. **Translation into the Clinic**

   The motivation to design a topology-preservation segmentation model was to ensure that the predicted segmentations were suitable for subsequent steps in medical imaging pipelines, as discussed in Chapter and illustrated in Figure 1.1. Throughout this thesis, I have shown that TEDS-Net can produce topologically-accurate segmentation that can facilitate automated measures of a structure's properties, such as perimeter (Chapter 4) or surface area (Chapter 6). However, despite these promising results, there are still large barriers preventing its translation into medical pipelines [261]. Although deep-learning-based methods are beginning to receive FDA approval for applications in medical imaging, rigorous evaluation is required for adoption, to ensure that the methodology is suitable across a wide demographic and insusceptible to domain shifts caused by different scanners. The translation of TEDS-Net would be task-specific, e.g. myocardium segmentation, and before it reached the clinic it would have to be shown to work effectively on a vast amount of data. Moreover, I would advise TEDS-Net to always be used to support the clinical workflow, with clinicians approving the segmentation before computing the measurements.

Cortical Plate Characterisation

1. **MRI Comparison**

   The ultrasound-derived characterisations of the cortical plate were compared to previous MRI studies and 2D ultrasound results, to help further validate the results. However, often comparison was not like for like, as the MRI results were often looking at slightly different structures or surfaces. For example, in MRI the surface area is measured from the inner cortical boundary across both hemispheres. However, as only one hemisphere is clearly visible in ultrasound volumes, all the results presented here were derived from a single

hemisphere per volume. Further, as it is challenging to extract the inner and outer boundaries from a single hemisphere, our surface area combines both boundaries. In future work, I would like to develop a method that could cap the brain stem in ultrasound, as done in MRI analysis, to help separate the two boundaries. This would facilitate a more direct comparison to previous work and would allow the use of more sophisticated analysis tools, such as electrical potentials to measure the cortical thickness [213].

2. **Cohort Size**

   The expected healthy cortical properties presented in this thesis were derived from a cohort of $n = 70$ from the INTERGROWTH-21$^{\text{st}}$ dataset, with the segmentation network trained on a further $n = 393$ volumes. Although this subset contained fetuses from all over the world, it is still a relatively small group. Further, due to a small sample size, I was unable to characterise the local properties at 25 GW when the right hemisphere was visible. To overcome these limitations, in future work I hope to characterise the cortical plate across a larger subset of the INTERGROWTH-21$^{\text{st}}$ dataset, to get a more accurate representation of the population. In preliminary work, I applied my analysis pipeline to $n = 1,643$ scans as shown in Section D.2, however, I hope to extend this further to a larger subset ($n = 9,995$) of newly aligned INTERGROWTH-21$^{\text{st}}$ brain scans.

3. **Future Applications**

   In this thesis, I have begun to characterise fetal brain development from routine ultrasound scans. One of the long-term goals of this project was to develop a benchmark of healthy cortical characteristics, to compare at-risk pregnancies too. As discussed in Section 3.1.4, differences in cortical growth have commonly been identified in preterm newborns and fetuses with congenital heart disease. However, for the former, this has only been seen after birth, and in the latter only assessed from 25 GW. Therefore an exciting avenue for future work would be to investigate whether any differences arise earlier

in pregnancies using our automated pipeline. In the work presented here, I studied population growth over gestation, however, in future work it would be interesting to study longitudinal development for individuals separately, to generate more personalised growth projections. As the INTERGROWTH-21st dataset contains longitudinal ultrasound brain scans, with each fetus scanned every 5 weeks, this data can be utilised to perform such analysis.

Appendices

A
Baseline Training

A.1 VoxelMorph

VoxelMorph implementation was designed for brain volume registration, so in order to apply it to the task of myocardium segmentation, certain parameters required tuning. In previous work, the weighting, ω, between the segmentation and field regularisation loss functions was found to play a key role in balancing registration and field smoothness performance [96, 262]. To allow a fair comparison, we first investigated the effect of this value, as shown in Figure A.1, in order to find the optimal performance for this segmentation task. A trade-off between segmentation performance and field smoothness is shown across values of the regularisation term, with optimal topology performance found at $\omega = 3$, which we, therefore, continued to use throughout this work. The field weighing terms used for VoxelMorph's implementation are 100 times smaller than that used for TEDS-Net, shown in Figure 4.11. This is because the two methods use different grid coordinate systems ($\mathbf{x}$), which results in different orders of magnitudes for the deformations and their derivatives. It should also be noted that VoxelMorph is primarily an image registration method so has not been optimised for segmentation.

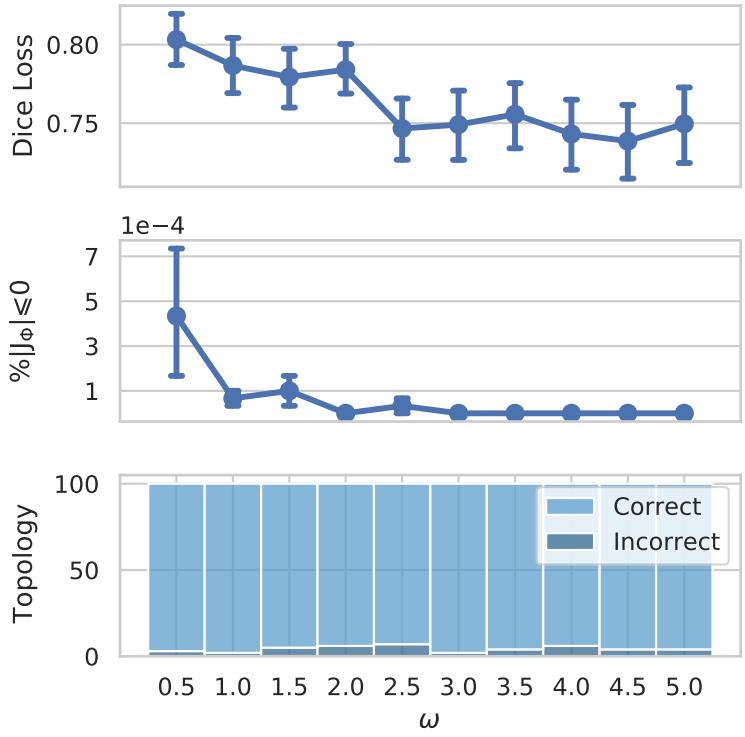


Figure A.1: The effect of the field regularisation weighting in VoxelMorph's implementation. The segmentation overlap (Dice), folding nature of the generating field and topology performance are measured. Results performed on the validation set.

A.2 TopoLoss

As in VoxelMorph's implementation one crucial parameter, ϵ has been found to play a key role in performance, and therefore the value of ϵ was varied to find the optimum performance, on the validation dataset, shown in Figure A.2. Due to the long runtimes of TopoLoss, shown in Table 4.3, for this experiment we only used 15% of the training set for fine-tuning. The best topology performance was found using $\epsilon = 0.2$, which was then used for the final experiment where we fine-tuned using 100% of the training dataset.

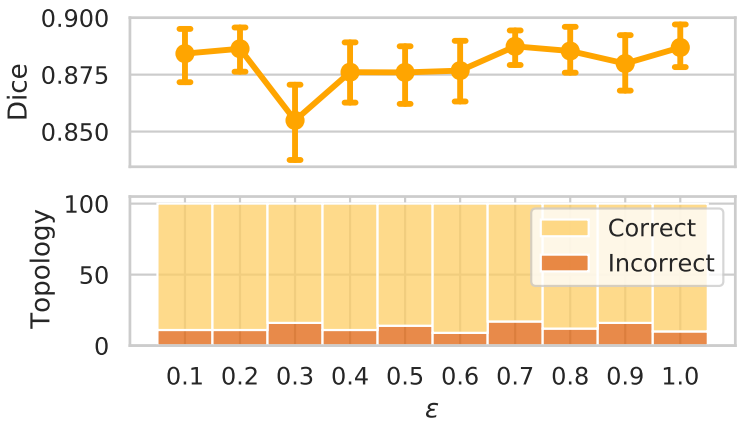


Figure A.2: The impact of the weighting factor, ϵ, between Dice overlap and TopoLoss in fine-tuning on the overlap segmentation performance.

B
Statistical Tests

When comparing the segmentation performance of M subjects, between N networks and using K different metrics, assessing whether one network statistically outperformed the other can be challenging. As this requires a large number of comparisons, using traditional multiple comparison corrections, such as Bonferroni [231], would not work so well. Therefore an alternative method is required. In this work, I use the classical multivariate test as implemented in FSL's PALM [229]. In this, the difference in performance between each individual (m) across each network (n) is found across the range of metrics.

Here I present those statistical tests across our different tasks.

ACDC Analysis

Figure B.1 shows the statistical results for the myocardium baseline performance, as reported in Chapter 4.6.4.

Liver Analysis

Figure B.2 shows the statistical results for the liver segmentation using both fully supervised and single slice supervision, as reported in Chapter 5.5.2.

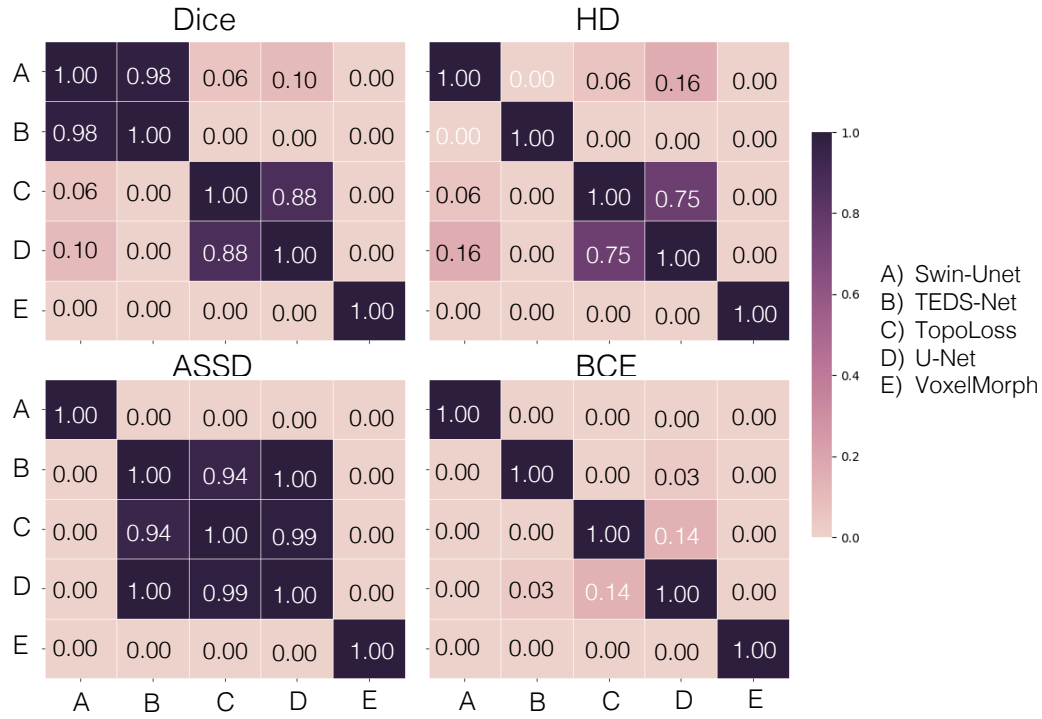


Figure B.1: Measure of statistically significant results for myocardium segmentation across four metrics (Dice, Hausdorff Distance (HD), Average Symmetrical Surface Distance (ASSD) and Binary Cross Entropy (BCE)) between the different network types. The heatmap shows the p values.

Cortical Plate Analysis

Figure B.3 shows the statistical results for the cortical plate segmentation, as reported in Chapter 6.4.2.2.

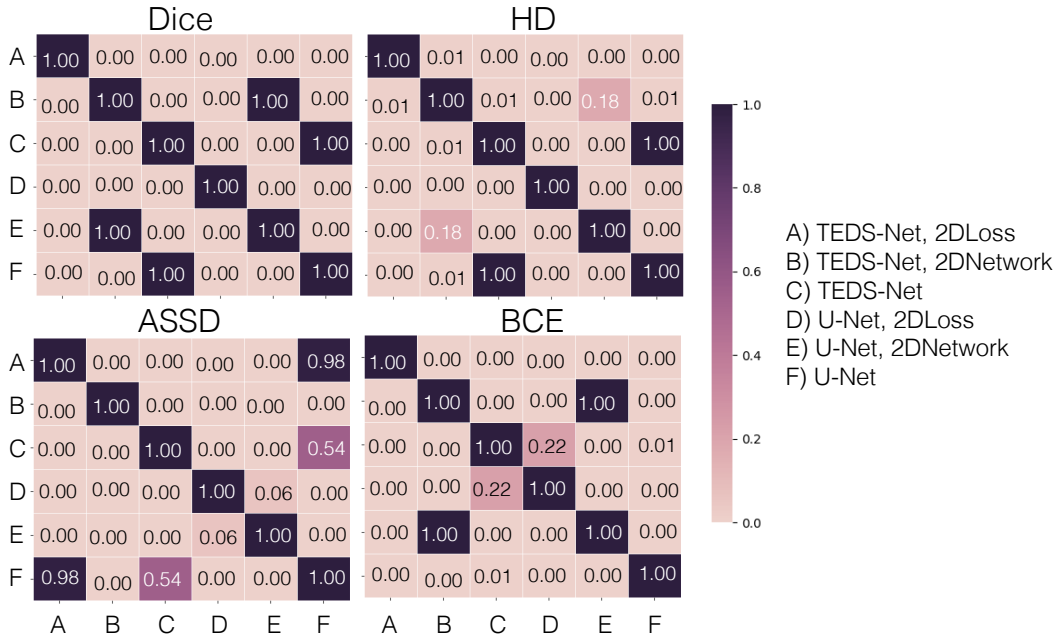


Figure B.2: Measure of statistically significant results for liver segmentation across four metrics (Dice, Hausdorff Distance (HD), Average Symmetrical Surface Distance (ASSD) and Binary Cross Entropy (BCE)) between the different network types. The heatmap shows the p values.

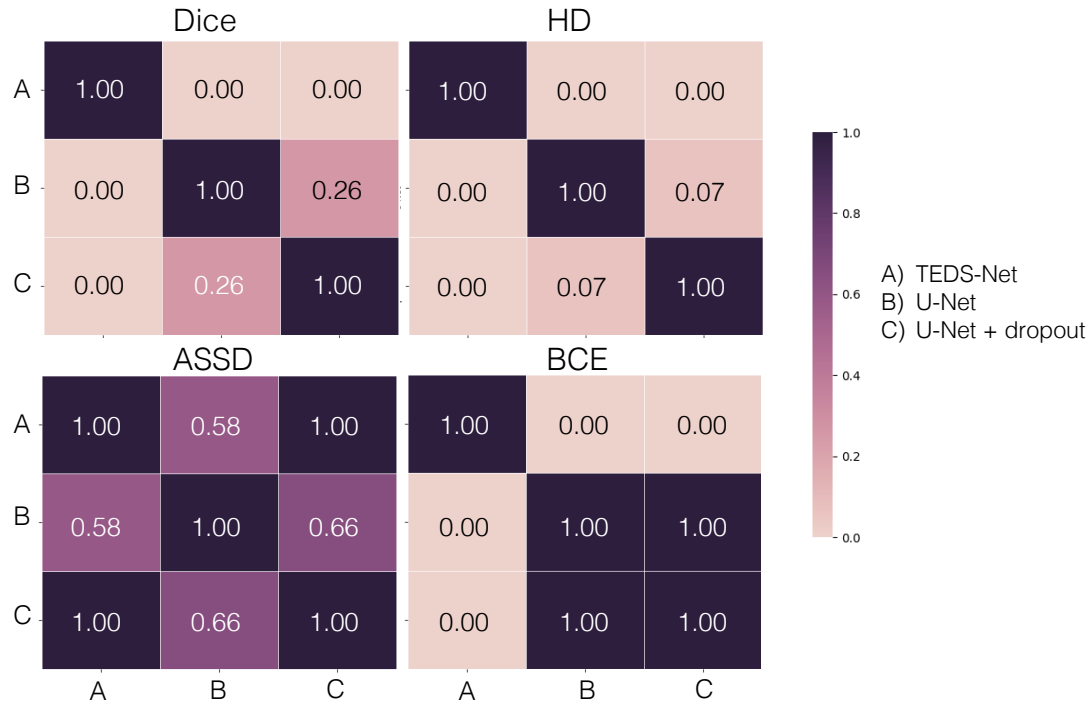


Figure B.3: Measure of statistically significant results for cortical plate segmentation across four metrics (Dice, Hausdorff Distance (HD), Average Symmetrical Surface Distance (ASSD) and Binary Cross Entropy (BCE)) between the different network types. The heatmap shows the p values.

C

Myocaridum Segmentation Examples

Qualitative examples of myocardium segmentation across the networks are shown in Figure C.1, Figure C.2 and Figure C.3, corresponding to the top, middle, and bottom five TEDS-Net segmentations in terms of Dice. On visual inspection, the segmentation quality across the baselines for the top (Figure C.1) and middle (Figure C.2) performing images seems consistent. However, when plotting Dice across the baselines for these images, as shown in Figure C.4, a slight increase in performance is seen for U-Net, TopoLoss and Swin-Net, consistent with the findings in Table 4.2. Conversely, in Figure C.3, the networks prediction heavily vary, which is reflected in Figure C.4. Interestingly, for the worse TEDS-Net segmentations perform on par with TopoLoss.

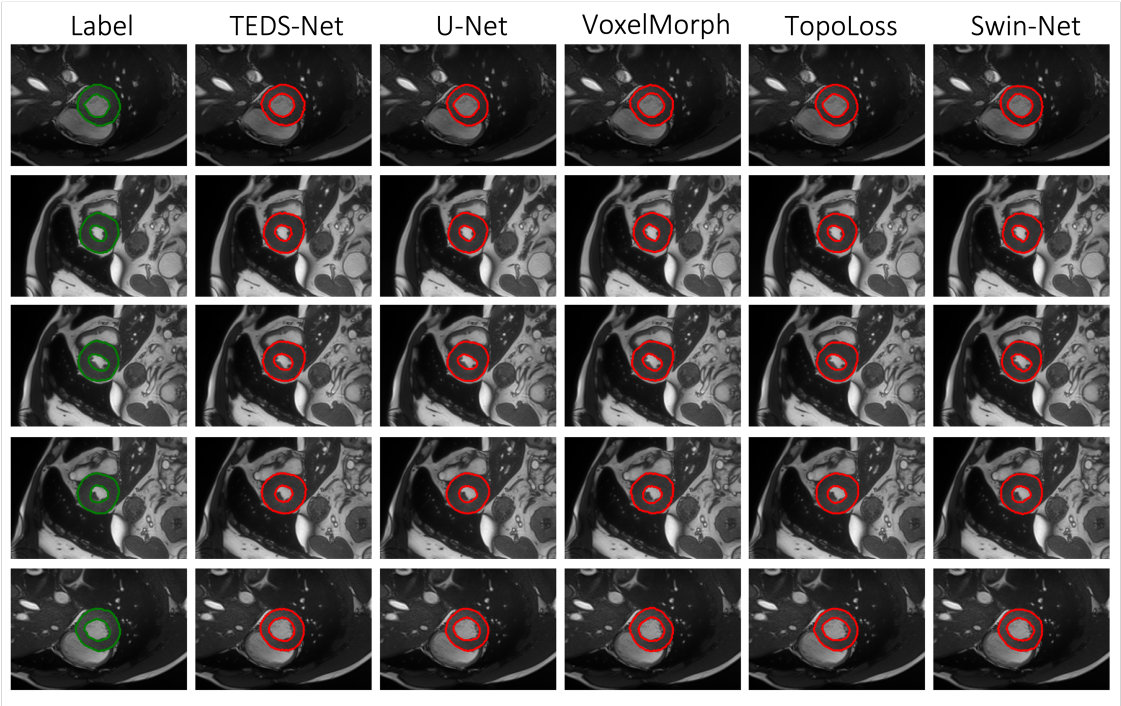


Figure C.1: The five best performing TEDS-Net segmentation in terms of Dice compared to the other baselines.

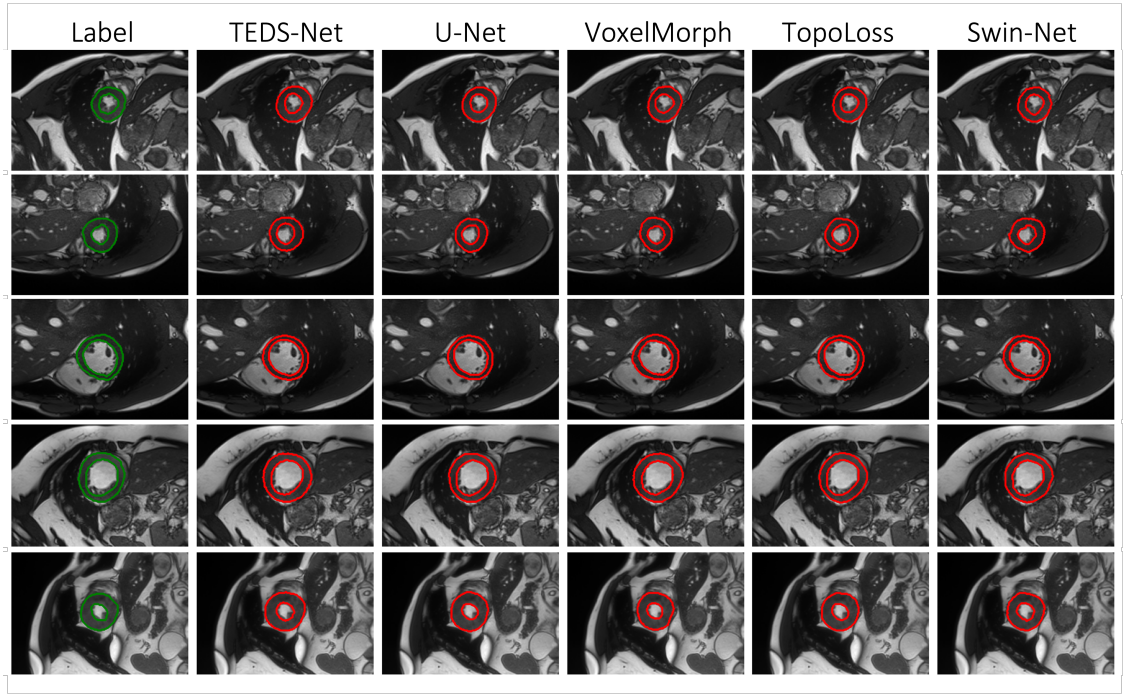


Figure C.2: The five mid-preforming TEDS-Net segmentation in terms of Dice compared, to the other baselines.

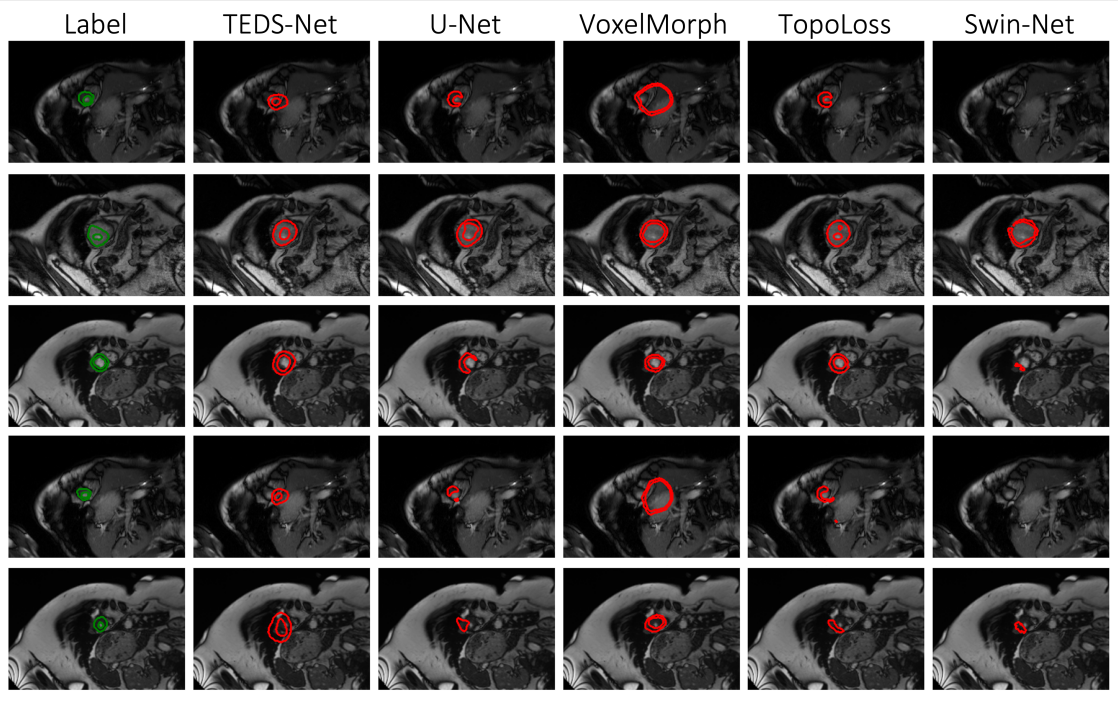


Figure C.3: The five worse performing TEDS-Net segmentations in terms of Dice, compared to the other baselines

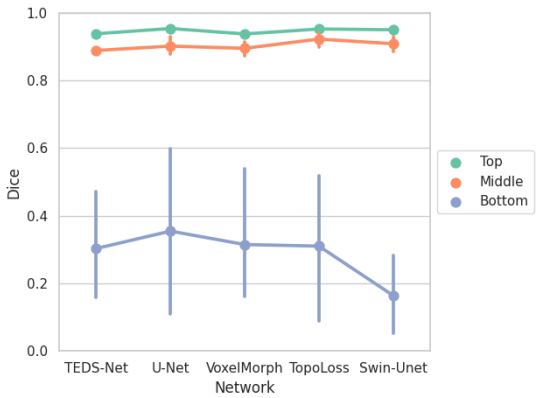


Figure C.4: The average Dice scores across the networks for the 3 groups displayed in Fig. C.1, Fig. C.2 and Fig. C.3.

D

Cortical Plate Segmentations

D.1 Initial Segmentation Methods

In my initial experiments, I tried to remove some of the topological holes within the U-Net predictions by using multi-task learning, with one decoder branch predicting a binary segmentation and the other a distance map to the segmentation, as shown in Figure D.1. I thought that tasking the network to learn the distance transform would mean the network had to intrinsically learn the higher-level structure of the cortex. However, this achieved a comparable Dice score and BCE to the U-Net but still resulted in topological errors.

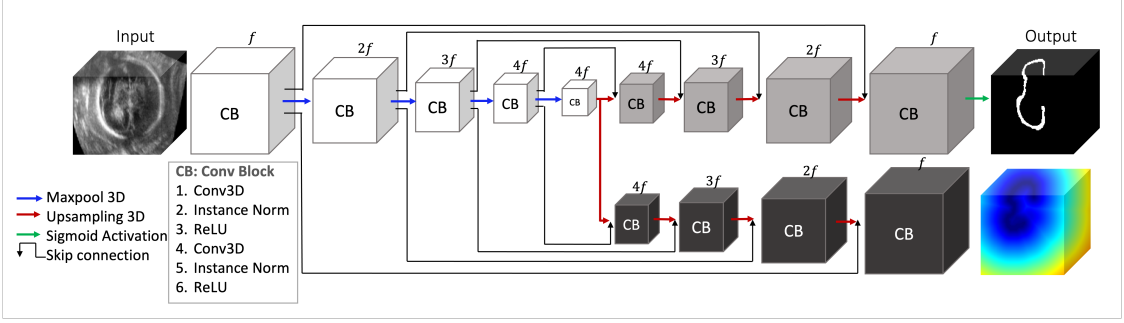


Figure D.1: The initial architecture used for cortical plate segmentation. Based of a U-Net with an additional decoder branch, tasked with learning the distance to the cortical plate.

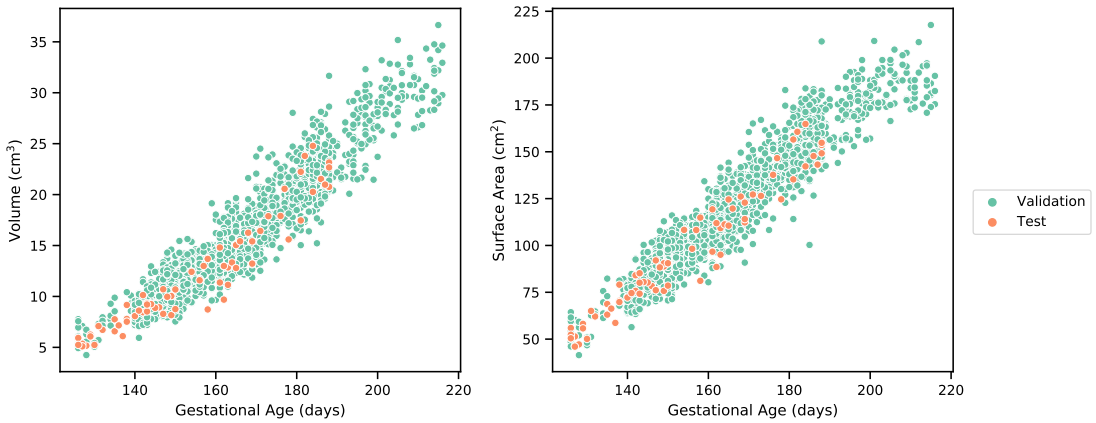


Figure D.2: Volume and surface area for a larger validation set of the INTERGROWTH-21st dataset, compared to the test set analysed in this chapter.

D.2 CP Growth on new data

As previously discussed, deep learning-based segmentation methods are advantageous over atlas-based propagation as no registrations are required to label new data. Figure D.2, shows the volume and surface area measures from a larger subset of aligned INTERGROWTH-21st dataset ($n = 1,643$), referred to as the validation set. In this figure, the validation set is compared to the test set, previously analysed in this chapter. The larger dataset closely follows the trend of the test set for both volume and surface area.

D.3 Domain Adaption

A drop in performance is often reported when deep learning techniques are applied to a new imaging domain (e.g. different scanners). To investigate TEDS-Net segmentation performance on new data, I computed the CP segmentation performance from $n = 233$ volumes collected from a new site, using a different scanning protocol and a different ultrasound probe. As manual labels were unavailable, the predicted annotations were inspected by an expert sonographer and graded using the following criteria:

1. High segmentation quality: Correct segmentation of the structure

2. Intermediate segmentation quality: Minor segmentation boundary errors ($< 20\%$ of the segmented structure is over- or under-segmented)

3. Low segmentation quality: Major segmentation boundary errors ($20 - 50\%$ of the segmented structure is over-or under-segmented)

4. Bad segmentation quality: Incorrect identification of the structure and/or $> 50\%$ of the structure over-or under-segmented.

The vast majority (78.5%, n=183) of the volumes were assigned the highest segmentation grade and 24 volumes (10.3%) were awarded the intermediate segmentation grade. In total 11.2% of the volumes was deemed "low" (n=19, 8.2%) or "bad" (n=7, 3.0%).

References

[1] JCL Rodrigues, T O'Regan, A Darekar, S Taylor, and V Goh. "Current pressure on the UK imaging workforce deters imaging research in the NHS and requires urgent attention". In: *Clinical Radiology* 77.12 (2022), pp. 913–919.

[2] NHS England and NHS Improvement. *Diagnostic Imaging Dataset Statistical Release*. `https://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2020/01/Provisional-Monthly-Diagnostic-Imaging-Dataset-Statistics-2020-01-23.pdf`. [Online; accessed 25-October-2022]. Jan. 2020.

[3] Debashis Ganguly, Srabonti Chakraborty, Maricel Balitanas, and Tai-hoon Kim. "Medical imaging: A review". In: *International Conference on Security-Enriched Urban Computing and Smart Grid*. Springer. 2010, pp. 504–516.

[4] Ramandeep Kaur and Jaswinder Kaur. "Current methods in medical image segmentation: a review". In: *Commun Comput Syst (ICCCS–2014)* 199 (2014).

[5] Dzung L Pham, Chenyang Xu, and Jerry L Prince. "A survey of current methods in medical image segmentation". In: *Annual review of biomedical engineering* 2.3 (2000), pp. 315–337.

[6] Yutian Chen, Wen Xie, Jiawei Zhang, Hailong Qiu, Dewen Zeng, Yiyu Shi, Haiyun Yuan, Jian Zhuang, Qianjun Jia, Yanchun Zhang, et al. "Myocardial Segmentation of Cardiac MRI Sequences With Temporal Consistency for Coronary Artery Disease Diagnosis". In: *Frontiers in Cardiovascular Medicine* 9 (2022).

[7] Elizabeth M Cespedes Feliciano, Karteek Popuri, Dana Cobzas, Vickie E Baracos, Mirza Faisal Beg, Arafat Dad Khan, Cydney Ma, Vincent Chow, Carla M Prado, Jingjie Xiao, et al. "Evaluation of automated computed tomography segmentation to assess body composition and mortality associations in cancer patients". In: *Journal of cachexia, sarcopenia and muscle* 11.5 (2020), pp. 1258–1269.

[8] Gregory Sharp, Karl D Fritscher, Vladimir Pekar, Marta Peroni, Nadya Shusharina, Harini Veeraraghavan, and Jinzhong Yang. "Vision 20/20: perspectives on automated image segmentation for radiotherapy". In: *Medical physics* 41.5 (2014), p. 050902.

[9] Bulat Ibragimov and Lei Xing. "Segmentation of organs-at-risks in head and neck CT images using convolutional neural networks". In: *Medical physics* 44.2 (2017), pp. 547–557.

[10] Marcel Toussaint, Raymond J Gilles, Noura Azzabou, Benjamin Marty, Alexandre Vignaud, Andreas Greiser, and Pierre G Carlier. "Characterization of benign myocarditis using quantitative delayed-enhancement imaging based on Molli T1 mapping". In: *Medicine* 94.43 (2015).

[11] Samer Alabed, Faisal Alandejani, Krit Dwivedi, Kavita Karunasaagarar, Michael Sharkey, Pankaj Garg, Patrick JH de Koning, Attila Tóth, Yousef Shahin, Christopher Johns, et al. "Validation of artificial intelligence cardiac MRI measurements: relationship to heart catheterization and mortality prediction". In: *Radiology* 305.1 (2022), pp. 68–79.

[12] The Royal College of Radiologists. *RCR Clinical Radiology Census Report 2021.* `https://www.rcr.ac.uk/sites/default/files/clinial_radiology_census_report_2021.pdf`. [Online; accessed 25-October-2022]. 2021.

[13] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "Imagenet classification with deep convolutional neural networks". In: *Communications of the ACM* 60.6 (2017), pp. 84–90.

[14] Karen Simonyan and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition". In: *arXiv preprint arXiv:1409.1556* (2014).

[15] Md Zahangir Alom, Tarek M Taha, Christopher Yakopcic, Stefan Westberg, Paheding Sidike, Mst Shamima Nasrin, Brian C Van Esesn, Abdul A S Awwal, and Vijayan K Asari. "The history began from AlexNet: A comprehensive survey on deep learning approaches". In: *arXiv preprint arXiv:1803.01164* (2018).

[16] DR Sarvamangala and Raghavendra V Kulkarni. "Convolutional neural networks in medical image understanding: a survey". In: *Evolutionary intelligence* (2021), pp. 1–22.

[17] James R Clough, Ilkay Oksuz, Nicholas Byrne, Julia A Schnabel, and Andrew P King. "Explicit topological priors for deep-learning based image segmentation using persistent homology". In: *International Conference on Information Processing in Medical Imaging*. Springer. 2019, pp. 16–28.

[18] Suprosanna Shit, Johannes C Paetzold, Anjany Sekuboyina, Ivan Ezhov, Alexander Unger, Andrey Zhylka, Josien PW Pluim, Ulrich Bauer, and Bjoern H Menze. "clDice-a Novel Topology-Preserving Loss Function for Tubular Structure Segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, pp. 16560–16569.

[19] Seung Yeon Shin, Sungwon Lee, Daniel Elton, James L Gulley, and Ronald M Summers. "Deep Small Bowel Segmentation with Cylindrical Topological Constraints". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2020, pp. 207–215.

[20] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, and Clara I Sánchez. "A survey on deep learning in medical image analysis". In: *Medical image analysis* 42 (2017), pp. 60–88.

[21] Francesco Galati, Sébastien Ourselin, and Maria A Zuluaga. "From Accuracy to Reliability and Robustness in Cardiac Magnetic Resonance Image Segmentation: A Review". In: *Applied Sciences* 12.8 (2022), p. 3936.

[22] C Bernard, P Droulle, F Didier, H Gerard, JC Larroche, F Plenat, F Bomsel, J Roland, and JC Hoeffel. "Echographic aspects of cerebral sulci in the ante-and perinatal period". In: *Journal de Radiologie* 69.8-9 (1988), pp. 521–532.

[23] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-Net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical image computing and computer-assisted intervention*. Springer. 2015, pp. 234–241.

[24] Hu Cao, Yueyue Wang, Joy Chen, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian, and Manning Wang. "Swin-unet: Unet-like pure transformer for medical image segmentation". In: *arXiv preprint arXiv:2105.05537* (2021).

[25] Adrian V Dalca, Guha Balakrishnan, John Guttag, and Mert R Sabuncu. "Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces". In: *Medical image analysis* 57 (2019), pp. 226–236.

[26] Xiaoling Hu, Fuxin Li, Dimitris Samaras, and Chao Chen. "Topology-preserving deep image segmentation". In: *Advances in Neural Information Processing Systems*. 2019, pp. 5658–5669.

[27] Yann LeCun. "The MNIST database of handwritten digits". In: *http://yann. lecun. com/exdb/mnist/* (1998).

[28] Neeraj Sharma, Lalit M Aggarwal, et al. "Automated medical image segmentation techniques". In: *Journal of medical physics* 35.1 (2010), p. 3.

[29] Prasanna K Sahoo, SAKC Soltani, and Andrew KC Wong. "A survey of thresholding techniques". In: *Computer vision, graphics, and image processing* 41.2 (1988), pp. 233–260.

[30] Peter Gibbs, David L Buckley, Stephen J Blackband, and Anthony Horsman. "Tumour volume determination from MR images by morphological segmentation". In: *Physics in Medicine & Biology* 41.11 (1996), p. 2437.

[31] Mariano Cabezas, Arnau Oliver, Xavier Lladó, Jordi Freixenet, and Meritxell Bach Cuadra. "A review of atlas-based segmentation for magnetic resonance brain images". In: *Computer methods and programs in biomedicine* 104.3 (2011), e158–e177.

[32] Torsten Rohlfing, Robert Brandt, Randolf Menzel, Daniel B Russakoff, and Calvin R Maurer. "Quo vadis, atlas-based segmentation?" In: *Handbook of biomedical image analysis*. Springer, 2005, pp. 435–486.

[33] M Bach Cuadra, Valérie Duay, and J-Ph Thiran. "Atlas-based segmentation". In: *Handbook of biomedical imaging*. Springer, 2015, pp. 221–244.

[34] Arthur W Toga and Paul M Thompson. "The role of image registration in brain mapping". In: *Image and vision computing* 19.1-2 (2001), pp. 3–24.

[35] Josien PW Pluim, JB Antoine Maintz, and Max A Viergever. "Mutual-information-based registration of medical images: a survey". In: *IEEE transactions on medical imaging* 22.8 (2003), pp. 986–1004.

[36] Derek LG Hill, Philipp G Batchelor, Mark Holden, and David J Hawkes. "Medical image registration". In: *Physics in medicine & biology* 46.3 (2001), R1.

[37] Hava Lester and Simon R Arridge. "A survey of hierarchical non-linear medical image registration". In: *Pattern recognition* 32.1 (1999), pp. 129–149.

[38] JB Antoine Maintz and Max A Viergever. "A survey of medical image registration". In: *Medical image analysis* 2.1 (1998), pp. 1–36.

[39] Ana IL Namburete, Weidi Xie, Mohammad Yaqub, Andrew Zisserman, and J Alison Noble. "Fully-automated alignment of 3D fetal brain ultrasound to a canonical reference space using multi-task learning". In: *Medical image analysis* 46 (2018), pp. 1–14.

[40] Adrian V Dalca, Guha Balakrishnan, John Guttag, and Mert R Sabuncu. "Unsupervised learning for fast probabilistic diffeomorphic registration". In: *MICCAI*. Springer. 2018, pp. 729–738.

[41] Piotr A Habas, Kio Kim, Francois Rousseau, Orit A Glenn, A James Barkovich, and Colin Studholme. "Atlas-based segmentation of the germinal matrix from in utero clinical MRI of the fetal brain". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2008, pp. 351–358.

[42] Ana IL Namburete, Raquel van Kampen, Aris T Papageorghiou, and Bartłomiej W Papież. "Multi-channel groupwise registration to construct an ultrasound-specific fetal brain atlas". In: *Data Driven Treatment Response Assessment and Preterm, Perinatal, and Paediatric Image Analysis*. Springer, 2018, pp. 76–86.

[43] Ali Gholipour, Judy A Estroff, Carol E Barnewolt, Susan A Connolly, and Simon K Warfield. "Fetal brain volumetry through MRI volumetric reconstruction and segmentation". In: *International journal of computer assisted radiology and surgery* 6.3 (2011), pp. 329–339.

[44] Michael Kass, Andrew Witkin, and Demetri Terzopoulos. "Snakes: Active contour models". In: *International journal of computer vision* 1.4 (1988), pp. 321–331.

[45] O Teytaud and D Sarrut. "Convergence speed of deformable models". In: *IJCNN'01. International Joint Conference on Neural Networks. Proceedings (Cat. No. 01CH37222)*. Vol. 4. IEEE. 2001, pp. 2850–2855.

[46] Kok F Lai and Roland T Chin. "On regularization, formulation and initialization of the active contour models (snakes)". In: *Asian Conference on Computer Vision*. Citeseer. 1993, pp. 542–545.

[47] Chenyang Xu and Jerry L Prince. "Snakes, shapes, and gradient vector flow". In: *IEEE Transactions on image processing* 7.3 (1998), pp. 359–369.

[48] Nikos Paragios, Mikael Rousson, and Visvanathan Ramesh. "Knowledge-based registration & segmentation of the left ventricle: A level set approach". In: *Sixth IEEE Workshop on Applications of Computer Vision, 2002.(WACV 2002). Proceedings.* IEEE. 2002, pp. 37–42.

[49] Vicent Caselles, Francine Catté, Tomeu Coll, and Françoise Dibos. "A geometric model for active contours in image processing". In: *Numerische mathematik* 66.1 (1993), pp. 1–31.

[50] Suyu Dong, Gongning Luo, Kuanquan Wang, Shaodong Cao, Qince Li, and Henggui Zhang. "A combined fully convolutional networks and deformable model for automatic left ventricle segmentation based on 3D echocardiography". In: *BioMed research international* 2018 (2018).

[51] Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel. "Backpropagation applied to handwritten zip code recognition". In: *Neural computation* 1.4 (1989), pp. 541–551.

[52] Rafael C Gonzalez and Richard E Woods. *Digital image processing*. eng. Addison-Wesley world student series. Reading, Mass.: Addison-Wesley, 1992.

[53] Jonas Teuwen and Nikita Moriakov. "Convolutional neural networks". In: *Handbook of medical image computing and computer assisted intervention*. Elsevier, 2020, pp. 481–501.

[54] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. "Imagenet: A large-scale hierarchical image database". In: *2009 IEEE conference on computer vision and pattern recognition*. Ieee. 2009, pp. 248–255.

[55] Clement Zotti, Zhiming Luo, Alain Lalande, and Pierre-Marc Jodoin. "Convolutional neural network with shape prior applied to cardiac MRI segmentation". In: *IEEE journal of biomedical and health informatics* 23.3 (2018), pp. 1119–1128.

[56] Guofeng Tong, Yong Li, Huairong Chen, Qingchun Zhang, and Huiying Jiang. "Improved U-Net network for pulmonary nodules segmentation". In: *Optik* 174 (2018), pp. 460–469.

[57] Jiazhou Wang, Jiayu Lu, Gan Qin, Lijun Shen, Yiqun Sun, Hongmei Ying, Zhen Zhang, and Weigang Hu. "A deep learning-based autosegmentation of rectal tumors in MR images". In: *Medical physics* 45.6 (2018), pp. 2560–2564.

[58] Berk Norman, Valentina Pedoia, and Sharmila Majumdar. "Use of 2D U-Net convolutional neural networks for automated cartilage and meniscus segmentation of knee MR imaging data to determine relaxometry and morphometry". In: *Radiology* 288.1 (2018), p. 177.

[59] Simon Kohl, Bernardino Romera-Paredes, Clemens Meyer, Jeffrey De Fauw, Joseph R Ledsam, Klaus Maier-Hein, SM Eslami, Danilo Jimenez Rezende, and Olaf Ronneberger. "A probabilistic U-Net for segmentation of ambiguous images". In: *Advances in neural information processing systems* 31 (2018).

[60] Changlu Guo, Márton Szemenyei, Yang Pei, Yugen Yi, and Wei Zhou. "SD-UNet: A structured dropout U-Net for retinal vessel segmentation". In: *2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE)*. IEEE. 2019, pp. 439–444.

[61] Ozan Oktay, Enzo Ferrante, Konstantinos Kamnitsas, Mattias Heinrich, Wenjia Bai, Jose Caballero, Stuart A Cook, Antonio De Marvao, Timothy Dawes, Declan P O'Regan, et al. "Anatomically constrained neural networks (ACNNs): application to cardiac image enhancement and segmentation". In: *IEEE transactions on medical imaging* 37.2 (2017), pp. 384–395.

[62] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. "Unet++: A nested U-Net architecture for medical image segmentation". In: *Deep learning in medical image analysis and multimodal learning for clinical decision support*. Springer, 2018, pp. 3–11.

[63] Nahian Siddique, Sidike Paheding, Colin P Elkin, and Vijay Devabhaktuni. "U-Net and its variants for medical image segmentation: A review of theory and applications". In: *IEEE Access* (2021).

[64] Getao Du, Xu Cao, Jimin Liang, Xueli Chen, and Yonghua Zhan. "Medical image segmentation based on U-Net: A review". In: *Journal of Imaging Science and Technology* 64.2 (2020), pp. 20508–1.

[65] Houlong Zhao and Nongliang Sun. "Improved U-Net model for nerve segmentation". In: *International Conference on Image and Graphics*. Springer. 2017, pp. 496–504.

[66] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. "3D U-Net: learning dense volumetric segmentation from sparse annotation". In: *International conference on medical image computing and computer-assisted intervention*. Springer. 2016, pp. 424–432.

[67] 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)*. IEEE. 2016, pp. 565–571.

[68] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al. "Attention U-Net: Learning where to look for the pancreas". In: *arXiv preprint arXiv:1804.03999* (2018).

[69] Robin Strudel, Ricardo Garcia, Ivan Laptev, and Cordelia Schmid. "Segmenter: Transformer for semantic segmentation". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021, pp. 7262–7272.

[70] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. "Attention is all you need". In: *Advances in neural information processing systems* 30 (2017).

[71] Stéphane d'Ascoli, Hugo Touvron, Matthew L Leavitt, Ari S Morcos, Giulio Biroli, and Levent Sagun. "Convit: Improving vision transformers with soft convolutional inductive biases". In: *International Conference on Machine Learning*. PMLR. 2021, pp. 2286–2296.

[72] William Fedus, Barret Zoph, and Noam Shazeer. *Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity*. 2021.

[73] Varduhi Yeghiazaryan and Irina D Voiculescu. "Family of boundary overlap metrics for the evaluation of medical image segmentation". In: *Journal of Medical Imaging* 5.1 (2018), p. 015006.

[74] Shruti Jadon. "A survey of loss functions for semantic segmentation". In: *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*. IEEE. 2020, pp. 1–7.

[75] Lee R Dice. "Measures of the amount of ecologic association between species". In: *Ecology* 26.3 (1945), pp. 297–302.

[76] Daniel P Huttenlocher, Gregory A. Klanderman, and William J Rucklidge. "Comparing images using the Hausdorff distance". In: *IEEE Transactions on pattern analysis and machine intelligence* 15.9 (1993), pp. 850–863.

[77] Bradford Henry Arnold. *Intuitive concepts in elementary topology*. Courier Corporation, 2011.

[78] Priscille de Dumast, Hamza Kebiri, Chirine Atat, Vincent Dunet, Mériam Koob, and Meritxell Bach Cuadra. "Segmentation of the cortical plate in fetal brain MRI with a topological loss". In: *arXiv preprint arXiv:2010.12391* (2020).

[79] James Clough, Nicholas Byrne, Ilkay Oksuz, Veronika A Zimmer, Julia A Schnabel, and Andrew King. "A topological loss function for deep-learning based image segmentation using persistent homology". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2020).

[80] Agata Mosinska, Pablo Marquez-Neila, Mateusz Koziński, and Pascal Fua. "Beyond the pixel-wise loss for topology-aware delineation". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018, pp. 3136–3145.

[81] Jun Cao, Xiaoming Liu, Ying Zhang, and Man Wang. "A Multi-task Framework for Topology-guaranteed Retinal Layer Segmentation in OCT Images". In: *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE. 2020, pp. 3142–3147.

[82] Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al. "Spatial transformer networks". In: *Advances in neural information processing systems*. 2015, pp. 2017–2025.

[83] Nicola K Dinsdale, Mark Jenkinson, and Ana IL Namburete. "Spatial warping network for 3D segmentation of the hippocampus in mr images". In: *MICCAI*. Springer. 2019, pp. 284–291.

[84] Qi Zeng, Davood Karimi, Emily HT Pang, Shahed Mohammed, Caitlin Schneider, Mohammad Honarvar, and Septimiu E Salcudean. "Liver Segmentation in Magnetic Resonance Imaging via Mean Shape Fitting with Fully Convolutional Neural Networks". In: *MICCAI*. Springer. 2019, pp. 246–254.

[85] Udaranga Wickramasinghe, Graham Knott, and Pascal Fua. "Probabilistic Atlases to Enforce Topological Constraints". In: *MICCAI*. Springer. 2019, pp. 218–226.

[86] Matthew Chung Hai Lee, Kersten Petersen, Nick Pawlowski, Ben Glocker, and Michiel Schaap. "Tetris: Template transformer networks for image segmentation with shape priors". In: *IEEE transactions on medical imaging* 38.11 (2019), pp. 2596–2606.

[87] Bilge Karaçali and Christos Davatzikos. "Estimating topology preserving and smooth displacement fields". In: *IEEE Transactions on Medical Imaging* 23.7 (2004), pp. 868–880.

[88] Nicolas Passat, Phuc Ngo, Yukiko Kenmochi, and Hugues Talbot. "Homotopic affine transformations in the 2D Cartesian grid". In: *Journal of Mathematical Imaging and Vision* (2022), pp. 1–21.

[89] John Ashburner. "A fast diffeomorphic image registration algorithm". In: *Neuroimage* 38.1 (2007), pp. 95–113.

[90] Olivier Musse, Fabrice Heitz, and Jean-Paul Armspach. "Topology preserving deformable image matching using constrained hierarchical parametric models". In: *IEEE transactions on image processing* 10.7 (2001), pp. 1081–1093.

[91] Cleve Moler and Charles Van Loan. “Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later”. In: *SIAM review* 45.1 (2003), pp. 3–49.

[92] Vincent Arsigny, Olivier Commowick, Xavier Pennec, and Nicholas Ayache. “A log-euclidean polyaffine framework for locally rigid or affine registration”. In: *WBIR*. Springer. 2006, pp. 120–127.

[93] Tom Vercauteren, Xavier Pennec, Aymeric Perchant, and Nicholas Ayache. “Diffeomorphic demons: Efficient non-parametric image registration”. In: *NeuroImage* 45.1 (2009), S61–S72.

[94] M Faisal Beg, Michael I Miller, Alain Trouvé, and Laurent Younes. “Computing large deformation metric mappings via geodesic flows of diffeomorphisms”. In: *International journal of computer vision* 61.2 (2005), pp. 139–157.

[95] Yan Cao, Michael I Miller, Raimond L Winslow, and Laurent Younes. “Large deformation diffeomorphic metric mapping of vector fields”. In: *IEEE transactions on medical imaging* 24.9 (2005), pp. 1216–1230.

[96] Guha Balakrishnan, Amy Zhao, Mert R Sabuncu, John Guttag, and Adrian V Dalca. “Voxelmorph: a learning framework for deformable medical image registration”. In: *IEEE transactions on medical imaging* 38.8 (2019), pp. 1788–1800.

[97] Tony CW Mok and Albert Chung. “Fast symmetric diffeomorphic image registration with convolutional neural networks”. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020, pp. 4644–4653.

[98] Matthew Sinclair, Andreas Schuh, Karl Hahn, Kersten Petersen, Ying Bai, James Batten, Michiel Schaap, and Ben Glocker. “Atlas-ISTN: Joint Segmentation, Registration and Atlas Construction with Image-and-Spatial Transformer Networks”. In: *arXiv preprint arXiv:2012.10533* (2020).

[99] Phuc Ngo, Yukiko Kenmochi, Nicolas Passat, and Hugues Talbot. “Topology-preserving conditions for 2D digital images under rigid transformations”. In: *Journal of mathematical imaging and vision* 49.2 (2014), pp. 418–433.

[100] J-P Thirion. “Image matching as a diffusion process: an analogy with Maxwell’s demons”. In: *Medical image analysis* 2.3 (1998), pp. 243–260.

[101] Brian B Avants, Charles L Epstein, Murray Grossman, and James C Gee. “Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain”. In: *Medical image analysis* 12.1 (2008), pp. 26–41.

[102] Han Zhang and Lok Ming Lui. “Topology-Preserving Segmentation Network: A Deep Learning Segmentation Framework for Connected Component”. In: *arXiv preprint arXiv:2202.13331* (2022).

[103] Madeleine K. Wyburd, Nicola K. Dinsdale, Ana I. L. Namburete, and Mark Jenkinson. "TEDS-Net: Enforcing Diffeomorphisms in Spatial Transformers to Guarantee Topology Preservation in Segmentations". In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*. Ed. by Marleen de Bruijne, Philippe C. Cattin, Stéphane Cotin, Nicolas Padoy, Stefanie Speidel, Yefeng Zheng, and Caroline Essert. Cham: Springer International Publishing, 2021, pp. 250–260.

[104] Patricia E Hansen, Martha C Ballesteros, Kalevi Soila, Laura Garcia, and Joanne M Howard. "MR imaging of the developing human brain. Part 1. Prenatal development." In: *Radiographics* 13.1 (1993), pp. 21–36.

[105] Leonard B Radinsky. "A new approach to mammalian cranial analysis, illustrated by examples of prosimian primates". In: *Journal of Morphology* 124.2 (1968), pp. 167–179.

[106] Hans Elias and David Schwartz. "Surface areas of the cerebral cortex of mammals determined by stereological methods". In: *Science* 166.3901 (1969), pp. 111–113.

[107] Karl Zilles, Este Armstrong, Axel Schleicher, and Hans-Joachim Kretschmann. "The human pattern of gyrification in the cerebral cortex". In: *Anatomy and embryology* 179.2 (1988), pp. 173–179.

[108] Karl Zilles, Este Armstrong, Karl H Moser, Alex Schleicher, and Heinz Stephan. "Gyrification in the cerebral cortex of primates". In: *Brain, Behavior and Evolution* 34.3 (1989), pp. 143–150.

[109] Joan Stiles and Terry L Jernigan. "The basics of brain development". In: *Neuropsychology review* 20.4 (2010), pp. 327–348.

[110] Tao Sun and Robert F Hevner. "Growth and folding of the mammalian cerebral cortex: from molecules to malformations". In: *Nature Reviews Neuroscience* 15.4 (2014), pp. 217–232.

[111] Georg F. Striedter, Shyam Srinivasan, and Edwin S. Monuki. "Cortical Folding: When, Where, How, and Why?" In: *Annual Review of Neuroscience* 38.1 (July 2015), pp. 291–307.

[112] Cédric Clouchoux, Dimitri Kudelski, Ali Gholipour, Simon K Warfield, Sophie Viseur, Marine Bouyssi-Kobar, Jean-Luc Mari, Alan C Evans, Adre J Du Plessis, and Catherine Limperopoulos. "Quantitative in vivo MRI measurement of cortical development in the fetus". In: *Brain Structure and Function* 217.1 (2012), pp. 127–139.

[113] Yuning Zhang, Terrie E Inder, Jeffrey J Neil, Donna L Dierker, Dimitrios Alexopoulos, Peter J Anderson, and David C Van Essen. "Cortical structural abnormalities in very preterm children at 7 years of age". In: *Neuroimage* 109 (2015), pp. 469–479.

[114] C Clouchoux, AJ Du Plessis, M Bouyssi-Kobar, W Tworetzky, DB McElhinney, DW Brown, A Gholipour, D Kudelski, SK Warfield, RJ McCarter, et al. "Delayed cortical development in fetuses with complex congenital heart disease". In: *Cerebral cortex* 23.12 (2013), pp. 2932–2943.

[115] Julien Lefèvre, David Germanaud, Jessica Dubois, François Rousseau, Ines de Macedo Santos, Hugo Angleys, Jean-François Mangin, Petra S Hüppi, Nadine Girard, and François De Guio. "Are developmental trajectories of cortical folding comparable between cross-sectional datasets of fetuses and preterm newborns?" In: *Cerebral cortex* 26.7 (2015), pp. 3023–3035.

[116] Arthur W Toga, Paul M Thompson, and Elizabeth R Sowell. "Mapping brain maturation". In: *Focus* 29.3 (2006), pp. 148–390.

[117] M.Judaš I.Kostović. "Embryonic and Fetal Development of the Human Cerebral Cortex". In: *Brain mapping: An encyclopedic reference*. Ed. by Arthur W Toga. Vol. 2. Academic Press, 2015, pp. 167–175.

[118] Silvia Budday, Paul Steinmann III, and Ellen Kuhl. "Physical biology of human brain development". In: *Frontiers in cellular neuroscience* 9 (2015), p. 257.

[119] Maria E Barnes-Davis, Brady J Williamson, Stephanie L Merhar, Scott K Holland, and Darren S Kadis. "Extremely preterm children exhibit altered cortical thickness in language areas". In: *Scientific reports* 10.1 (2020), pp. 1–10.

[120] Stacy Goergen. *Case study: Normal brain fetal MRI - 22 weeks.* `https://radiopaedia.org/cases/50623/`. [Online; accessed 31-October-2022]. Feb. 2017.

[121] Ivica Kostovic and Pasko Rakic. "Developmental history of the transient subplate zone in the visual and somatosensory cortex of the macaque monkey and human brain". In: *Journal of Comparative Neurology* 297.3 (1990), pp. 441–470.

[122] David C Van Essen. "A tension-based theory of morphogenesis and compact wiring in the central nervous system". In: *Nature* 385.6614 (1997), pp. 313–318.

[123] Tuomas Tallinen, Jun Young Chung, François Rousseau, Nadine Girard, Julien Lefèvre, and Lakshminarayanan Mahadevan. "On the growth and form of cortical convolutions". In: *Nature Physics* 12.6 (2016), pp. 588–593.

[124] Gang Xu, Andrew K Knutsen, Krikor Dikranian, Christopher D Kroenke, Philip V Bayly, and Larry A Taber. "Axons pull on the brain, but tension does not drive cortical folding". In: *Journal of biomechanical engineering* 132.7 (2010).

[125] Catherine Garel, Emmanuel Chantrel, Hervé Brisse, Monique Elmaleh, Dominique Luton, Jean-François Oury, Guy Sebag, and Max Hassan. "Fetal cerebral cortex: normal gestational landmarks identified using prenatal MR imaging". In: *American Journal of Neuroradiology* 22.1 (2001), pp. 184–189.

[126] Je G Chi, Elizabeth C Dooling, and Floyd H Gilles. "Gyral development of the human brain". In: *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society* 1.1 (1977), pp. 86–93.

[127] Toai Bui et al. "Neuroanatomy, cerebral hemisphere". In: (2019).

[128] K Dorovini-Zis and CL Dolman. "Gestational development of brain." In: *Archives of pathology & laboratory medicine* 101.4 (1977), pp. 192–195.

[129] Deborah Levine and Patrick D Barnes. "Cortical maturation in normal and abnormal fetuses as assessed with prenatal MR imaging". In: *Radiology* 210.3 (1999), pp. 751–758.

[130] Hyuk Jin Yun, Hyun Ju Lee, Joo Young Lee, Tomo Tarui, Caitlin K Rollins, Cynthia M Ortinau, Henry A Feldman, P Ellen Grant, and Kiho Im. "Quantification of sulcal emergence timing and its variability in early fetal life: Hemispheric asymmetry and sex difference". In: *NeuroImage* 263 (2022), p. 119629.

[131] A Monteagudo and IE Timor-Tritsch. "Development of fetal gyri, sulci and fissures: a transvaginal sonographic study". In: *Ultrasound in Obstetrics and Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology* 9.4 (1997), pp. 222–228.

[132] A Toi, WS Lister, and KW Fong. "How early are fetal cerebral sulci visible at prenatal ultrasound and what is the normal pattern of early fetal sulcal development?" In: *Ultrasound in Obstetrics and Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology* 24.7 (2004), pp. 706–715.

[133] B Cohen-Sacher, T Lerman-Sagie, D Lev, and G Malinger. "Sonographic developmental milestones of the fetal cerebral cortex: a longitudinal study". In: *Ultrasound in Obstetrics and Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology* 27.5 (2006), pp. 494–502.

[134] Piotr A Habas, Julia A Scott, Ahmad Roosta, Vidya Rajagopalan, Kio Kim, Francois Rousseau, A James Barkovich, Orit A Glenn, and Colin Studholme. "Early folding patterns and asymmetries of the normal human brain detected from in utero MRI". In: *Cerebral cortex* 22.1 (2012), pp. 13–25.

[135] TP Naidich, JL Grant, N Altman, RA Zimmerman, SB Birchansky, B Braffman, and JL Daniel. "The developing cerebral surface. Preliminary report on the patterns of sulcal and gyral maturation–anatomy, ultrasound, and magnetic resonance imaging." In: *Neuroimaging Clinics of North America* 4.2 (1994), pp. 201–240.

[136] Cynthia M Ortinau, Caitlin K Rollins, Ali Gholipour, Hyuk Jin Yun, Mackenzie Marshall, Borjan Gagoski, Onur Afacan, Kevin Friedman, Wayne Tworetzky, Simon K Warfield, et al. "Early-emerging sulcal patterns are atypical in fetuses with congenital heart disease". In: *Cerebral Cortex* 29.8 (2019), pp. 3605–3616.

[137] E Quarello, J Stirnemann, Y Ville, and L Guibaud. "Assessment of fetal Sylvian fissure operculization between 22 and 32 weeks: a subjective approach". In: *Ultrasound in Obstetrics and Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology* 32.1 (2008), pp. 44–49.

[138] P Duret, F Samson, B Pinsard, EB Barbeau, A Boré, I Soulières, and L Mottron. "Gyrification changes are related to cognitive strengths in autism". In: *NeuroImage: Clinical* 20 (2018), pp. 415–423.

[139] David C Van Essen, Donna Dierker, AZ Snyder, Marcus E Raichle, Allan L Reiss, and Julie Korenberg. "Symmetry of cortical folding abnormalities in Williams syndrome revealed by surface-based analyses". In: *Journal of Neuroscience* 26.20 (2006), pp. 5470–5483.

[140] Lena Palaniyappan, Pavan Mallikarjun, Verghese Joseph, Thomas P White, and Peter F Liddle. "Folding of the prefrontal cortex in schizophrenia: regional differences in gyrification". In: *Biological psychiatry* 69.10 (2011), pp. 974–979.

[141] Virginia Fernández, Cristina Llinares-Benadero, and Vıctor Borrell. "Cerebral cortex expansion and folding: what have we learned?" In: *The EMBO journal* 35.10 (2016), pp. 1021–1044.

[142] Chloe A Stutterd and Richard J Leventer. "Polymicrogyria: a common and heterogeneous malformation of cortical development". In: 166.2 (2014), pp. 227–239.

[143] Dennis M Hedderich, Josef G Bäuml, Maria T Berndt, Aurore Menegaux, Lukas Scheef, Marcel Daamen, Claus Zimmer, Peter Bartmann, Henning Boecker, Dieter Wolke, et al. "Aberrant gyrification contributes to the link between gestational age and adult IQ after premature birth". In: *Brain* 142.5 (2019), pp. 1255–1269.

[144] Tamanna Moore, Enid M Hennessy, Jonathan Myles, Samantha J Johnson, Elizabeth S Draper, Kate L Costeloe, and Neil Marlow. "Neurological and developmental outcome in extremely preterm children born in England in 1995 and 2006: the EPICure studies". In: *BMJ* 345 (2012), e7961.

[145] Natalie Suff, Lisa Story, and Andrew Shennan. "The prediction of preterm delivery: What is new?" In: *Seminars in Fetal and Neonatal Medicine*. Vol. 24. 1. Elsevier. 2019, pp. 27–32.

[146] Hannah Blencowe, Simon Cousens, Doris Chou, Mikkel Oestergaard, Lale Say, Ann-Beth Moller, Mary Kinney, and Joy Lawn. "Born too soon: the global epidemiology of 15 million preterm births". In: *Reproductive health* 10.1 (2013), S2.

[147] Robert L Goldenberg, Jay D Iams, Anita Das, Brian M Mercer, Paul J Meis, Atef H Moawad, Menachem Miodovnik, J Peter VanDorsten, Steve N Caritis, Gary R Thurnau, et al. "The Preterm Prediction Study: sequential cervical length and fetal fibronectin testing for the prediction of spontaneous preterm birth". In: *American journal of obstetrics and gynecology* 182.3 (2000), pp. 636–643.

[148] Dennis M Hedderich, Josef G Bäuml, Aurore Menegaux, Mihai Avram, Marcel Daamen, Claus Zimmer, Peter Bartmann, Lukas Scheef, Henning Boecker, Dieter Wolke, et al. "An analysis of MRI derived cortical complexity in premature-born adults: Regional patterns, risk factors, and potential significance". In: *NeuroImage* 208 (2020), p. 116438.

[149] Nancy Medley, Borna Poljak, Silvia Mammarella, and Zarko Alfirevic. "Clinical guidelines for prevention and management of preterm birth: a systematic review". In: *BJOG: An International Journal of Obstetrics & Gynaecology* 125.11 (2018), pp. 1361–1369.

[150] Moriah E Thomason, Dustin Scheinost, Janessa H Manning, Lauren E Grove, Jasmine Hect, Narcis Marshall, Edgar Hernandez-Andrade, Susan Berman, Athina Pappas, Lami Yeo, et al. "Weak functional connectivity in the human fetal brain prior to preterm birth". In: *Scientific reports* 7.1 (2017), pp. 1–10.

[151] Paul D Morton, Nobuyuki Ishibashi, and Richard A Jonas. "Neurodevelopmental abnormalities and congenital heart disease: insights into altered brain maturation". In: *Circulation research* 120.6 (2017), pp. 960–977.

[152] Mallory Owen, Michael Shevell, Mary Donofrio, Annette Majnemer, Robert McCarter, Gilbert Vezina, Marine Bouyssi-Kobar, Iordanis Evangelou, Dena Freeman, Neil Weisenfeld, et al. "Brain volume and neurobehavior in newborns with complex congenital heart defects". In: *The Journal of pediatrics* 164.5 (2014), pp. 1121–1127.

[153] William I Norwood Jr. "Hypoplastic left heart syndrome". In: *The annals of thoracic surgery* 52.3 (1991), pp. 688–695.

[154] David J Barron, Mark D Kilby, Ben Davies, John GC Wright, Timothy J Jones, and William J Brawn. "Hypoplastic left heart syndrome". In: *The Lancet* 374.9689 (2009), pp. 551–564.

[155] G Malinger, D Lev, and T Lerman-Sagie. "Normal and abnormal fetal brain development during the third trimester as demonstrated by neurosonography". In: *European journal of radiology* 57.2 (2006), pp. 226–232.

[156] Dario Paladini, Gustavo Malinger, Ana Monteagudo, Gianluigi Pilu, Ilan Timor-Tritsch, and Ants Toi. "Sonographic examination of the fetal central nervous system: guidelines for performing the 'basic examination' and the 'fetal neurosonogram'". In: *Ultrasound in Obstetrics and Gynecology* 29.1 (2007), pp. 109–116.

[157] LJ Salomon, Z Alfirevic, V Berghella, C Bilardo, E Hernandez-Andrade, SL Johnsen, K Kalache, K-Y Leung, G Malinger, H Munoz, et al. "Practice guidelines for performance of the routine mid-trimester fetal ultrasound scan". In: *Ultrasound in Obstetrics & Gynecology* 37.1 (2011), pp. 116–126.

[158] G Masselli, MR Vaccaro Notte, A Zacharzewska-Gondek, F Laghi, L Manganaro, and R Brunelli. "Fetal MRI of CNS abnormalities". In: *Clinical Radiology* 75.8 (2020), 640–e1.

[159] International Society of Ultrasound in Obstetrics and Gynecology Education Committee and others. "Sonographic examination of the fetal central nervous system: guidelines for performing the 'basic examination' and the fetal neurosonogram". In: *Ultrasound in obstetrics & gynecology: the official journal of the International Society of Ultrasound in Obstetrics and Gynecology* 29.1 (2007), pp. 109–116.

[160] Andrew CG Breeze, Peta MA Alexander, Edile M Murdoch, Hannah H Missfelder-Lobos, Gerald A Hackett, and Christoph C Lees. "Obstetric and neonatal outcomes in severe fetal ventriculomegaly". In: *Prenatal Diagnosis: Published in Affiliation With the International Society for Prenatal Diagnosis* 27.2 (2007), pp. 124–129.

[161] Michael Shepard and Roy A Filly. "A standardized plane for biparietal diameter measurement." In: *Journal of Ultrasound in medicine* 1.4 (1982), pp. 145–150.

[162] Raheli Dror, Gustavo Malinger, Liat Ben-Sira, Dorit Lev, Chaim G Pick, and Tally Lerman-Sagie. "Developmental outcome of children with enlargement of the cisterna magna identified in utero". In: *Journal of child neurology* 24.12 (2009), pp. 1486–1492.

[163] Venkatraman Bhat and Varun Bhat. "Neonatal neurosonography: A pictorial essay". In: *The Indian journal of radiology & imaging* 24.4 (2014), p. 389.

[164] Marıa J Rodriguez-Sibaja, José Villar, Eric O Ohuma, Raffaele Napolitano, Sabrina Heyl, Maria Carvalho, Yasmin A Jaffer, J Alison Noble, Manuela Oberto, M Purwar, et al. "Fetal cerebellar growth and Sylvian fissure maturation: international standards from Fetal Growth Longitudinal Study of INTERGROWTH-21st Project". In: *Ultrasound in Obstetrics & Gynecology* 57.4 (2021), pp. 614–623.

[165] Ana Monteagudo and Ilan E Timor-Tritsch. "Normal sonographic development of the central nervous system from the second trimester onwards using 2D, 3D and transvaginal sonography". In: *Prenatal Diagnosis: Published in Affiliation With the International Society for Prenatal Diagnosis* 29.4 (2009), pp. 326–339.

[166] Suguru Yasutomi, Tatsuya Arakaki, Ryu Matsuoka, Akira Sakai, Reina Komatsu, Kanto Shozu, Ai Dozen, Hidenori Machino, Ken Asada, Syuzo Kaneko, et al. "Shadow estimation for ultrasound images using auto-encoding structures and synthetic shadows". In: *Applied Sciences* 11.3 (2021), p. 1127.

[167] Elisenda Eixarch Eduard Gratacós. *Fetal Brain Care: Therapies for brain neurodevelopment in fetal growth restriction.* `https://clinicaltrials.gov/ProvidedDocs/62/NCT05038462/Prot_SAP_000.pdf`. [Online; accessed 16-November-2022]. July 2021.

[168] LR Pistorius, Ph Stoutenbeek, F Groenendaal, L De Vries, G Manten, E Mulder, and G Visser. "Grade and symmetry of normal fetal cortical development: a longitudinal two-and three-dimensional ultrasound study". In: *Ultrasound in obstetrics & gynecology* 36.6 (2010), pp. 700–708.

[169] Raffaele Napolitano, Malid Molloholli, Vera Donadono, Eric O Ohuma, SZ Wanyonyi, Bryn Kemp, Mohammad K Yaqub, Stephen Ash, Fernando C Barros, Maria Carvalho, et al. "International standards for fetal brain structures based on serial ultrasound measurements from Fetal Growth Longitudinal Study of INTERGROWTH-21st Project". In: *Ultrasound in Obstetrics & Gynecology* 56.3 (2020), pp. 359–370.

[170] I Alonso, M Borenstein, G Grant, I Narbona, and G Azumendi. "Depth of brain fissures in normal fetuses by prenatal ultrasound between 19 and 30 weeks of gestation". In: *Ultrasound in obstetrics & gynecology* 36.6 (2010), pp. 693–699.

[171] Sofie C Husen, Irene V Koning, Attie TJI Go, Anne W van Graafeiland, Sten P Willemsen, Irene AL Groenenberg, and Regine PM Steegers-Theunissen. "Three-dimensional ultrasound imaging of fetal brain fissures in the growth restricted fetus". In: *PloS one* 14.5 (2019).

[172] Xi Chen, Sheng-Li Li, Guo-Yang Luo, Errol R Norwitz, Shu-Yuan Ouyang, Hua-Xuan Wen, Ying Yuan, Xiao-Xian Tian, and Jia-Min He. "Ultrasonographic characteristics of cortical sulcus development in the human fetus between 18 and 41 weeks of gestation". In: *Chinese medical journal* 130.8 (2017), p. 920.

[173] LC Poon, DS Sahota, P Chaemsaithong, T Nakamura, M Machida, K Naruse, YM Wah, TY Leung, and RK Pooh. "Transvaginal three-dimensional ultrasound assessment of Sylvian fissures at 18–30 weeks' gestation". In: *Ultrasound in Obstetrics and Gynecology* 54.2 (2019), pp. 190–198.

[174] MA Coelho Neto, P Roncato, Carolina Oliveira Nastri, and Wellington de Paula Martins. "True Reproducibility of UltraSound Techniques (TRUST): systematic review of reliability studies in obstetrics and gynecology". In: *Ultrasound in Obstetrics & Gynecology* 46.1 (2015), pp. 14–20.

[175] Julia A Scott, Piotr A Habas, Kio Kim, Vidya Rajagopalan, Kia S Hamzelou, James M Corbett-Detig, A James Barkovich, Orit A Glenn, and Colin Studholme. "Growth trajectories of the human fetal brain tissues estimated from 3D reconstructed in utero MRI". In: *International Journal of Developmental Neuroscience* 29.5 (2011), pp. 529–536.

[176] Jing Xia, Fan Wang, Oualid M Benkarim, Gerard Sanroma, Gemma Piella, Miguel A González Ballester, Nadine Hahner, Elisenda Eixarch, Caiming Zhang, Dinggang Shen, et al. "Fetal cortical surface atlas parcellation based on growth patterns". In: *Human brain mapping* 40.13 (2019), pp. 3881–3899.

[177] Philipp G. Batchelor, AD Castellano Smith, Derek L. G. Hill, David J. Hawkes, Tim C. S. Cox, and AF Dean. "Measures of folding applied to the development of the human fetal brain". In: *IEEE transactions on medical imaging* 21.8 (2002), pp. 953–965.

[178] Haoran Dou, Davood Karimi, Caitlin K Rollins, Cynthia M Ortinau, Lana Vasung, Clemente Velasco-Annis, Abdelhakim Ouaalam, Xin Yang, Dong Ni, and Ali Gholipour. "A Deep Attentive Convolutional Neural Network for Automatic Cortical Plate Segmentation in Fetal MRI". In: *arXiv preprint arXiv:2004.12847* (2020).

[179] Nadieh Khalili, Nikolas Lessmann, Elise Turk, N Claessens, Roel de Heus, Tessel Kolk, Max A Viergever, Manon JNL Benders, and Ivana Išgum. "Automatic brain tissue segmentation in fetal MRI using convolutional neural networks". In: *Magnetic resonance imaging* 64 (2019), pp. 77–89.

[180] Ahmed E Fetit, Amir Alansary, Lucilio Cordero-Grande, John Cupitt, Alice B Davidson, A David Edwards, Joseph V Hajnal, Emer Hughes, Konstantinos Kamnitsas, Vanessa Kyriakopoulou, et al. "A deep learning approach to segmentation of the developing cortex in fetal brain MRI with minimal manual labeling". In: *Medical Imaging with Deep Learning*. 2020.

[181] Benoıt Caldairou, Nicolas Passat, Piotr Habas, Colin Studholme, Mériam Koob, J-L Dietemann, and François Rousseau. "Segmentation of the cortex in fetal MRI using a topological model". In: *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. IEEE. 2011, pp. 2045–2048.

[182] M Bach Cuadra, M Schaer, A Andre, L Guibaud, S Eliez, and J-Ph Thiran. "Brain tissue segmentation of fetal MR images". In: *Workshop on Image Analysis for Developing Brain, in 12th International Conference on Medical Image Computing and Computer Assisted Intervention*. CONF. 2009.

[183] Piotr A Habas, Kio Kim, Francois Rousseau, Orit A Glenn, A James Barkovich, and Colin Studholme. "Atlas-based segmentation of developing tissues in the human brain with quantitative validation in young fetuses". In: *Human brain mapping* 31.9 (2010), pp. 1348–1358.

[184] Jessica Dubois, Manon Benders, Arnaud Cachia, Francois Lazeyras, R Ha-Vinh Leuchter, Stéphane V Sizonenko, Christina Borradori-Tolsa, Jean-François Mangin, and Petra Susan Hüppi. "Mapping the early cortical folding process in the preterm newborn brain". In: *Cerebral Cortex* 18.6 (2008), pp. 1444–1454.

[185] J-F Mangin, Denis Riviere, Arnaud Cachia, Edouard Duchesnay, Yves Cointepas, Dimitri Papadopoulos-Orfanos, Paola Scifo, T Ochiai, Francis Brunelle, and Jean Régis. "A framework to study the cortical folding patterns". In: *Neuroimage* 23 (2004), S129–S138.

[186] A Serag, V Kyriakopoulou, MA Rutherford, AD Edwards, JV Hajnal, P Aljabar, SJ Counsell, J Boardman, and D Rueckert. "A multi-channel 4D probabilistic atlas of the developing brain: application to fetuses and neonates". In: *Annals of the BMVA* 2012.3 (2012), pp. 1–14.

[187] Ali Gholipour, Caitlin K Rollins, Clemente Velasco-Annis, Abdelhakim Ouaalam, Alireza Akhondi-Asl, Onur Afacan, Cynthia M Ortinau, Sean Clancy, Catherine Limperopoulos, Edward Yang, et al. "A normative spatiotemporal MRI atlas of the fetal brain for automatic segmentation and analysis of early brain growth". In: *Scientific reports* 7.1 (2017), pp. 1–13.

[188] Vidya Rajagopalan, Julia Scott, Piotr A Habas, Kio Kim, James Corbett-Detig, Francois Rousseau, A James Barkovich, Orit A Glenn, and Colin Studholme. "Local tissue growth patterns underlying normal fetal human brain gyrification quantified in utero". In: *Journal of neuroscience* 31.8 (2011), pp. 2878–2887.

[189] Priscille de Dumast, Hamza Kebiri, Vincent Dunet, Mériam Koob, and Meritxell Bach Cuadra. "Multi-dimensional topological loss for cortical plate segmentation in fetal brain MRI". In: *arXiv preprint arXiv:2208.07566* (2022).

[190] Madeleine K Wyburd, Mark Jenkinson, and Ana IL Namburete. "Cortical plate segmentation using CNNs in 3D fetal ultrasound". In: *Annual Conference on Medical Image Understanding and Analysis*. Springer. 2020, pp. 56–68.

[191] Anders M Dale, Bruce Fischl, and Martin I Sereno. "Cortical surface-based analysis: I. Segmentation and surface reconstruction". In: *Neuroimage* 9.2 (1999), pp. 179–194.

[192] Mitsuhiro Nishida, Nikolaos Makris, David N Kennedy, Mark Vangel, Bruce Fischl, Kalpathy S Krishnamoorthy, Verne S Caviness, and P Ellen Grant. "Detailed semiautomated MRI based morphometry of the neonatal brain: preliminary results". In: *Neuroimage* 32.3 (2006), pp. 1041–1049.

[193] Jérôme Dubois, M Benders, C Borradori-Tolsa, A Cachia, François Lazeyras, R Ha-Vinh Leuchter, SV Sizonenko, SK Warfield, JF Mangin, and Petra Susan Hüppi. "Primary cortical folding in the human newborn: an early marker of later functional development". In: *Brain* 131.8 (2008), pp. 2028–2041.

[194] N Guizard, C Lepage, V Fonov, H Hakyemez, A Evans, and C Limperopoulos. "Development of fetus brain atlas from multi-axial MR acquisitions". In: *Proceedings of the Sixteenth Annual Meeting of the International Society for Magnetic Resonance in Medicine*. Vol. 672. 2008, p. 132.

[195] Zeynettin Akkus, Alfiia Galimzianova, Assaf Hoogi, Daniel L Rubin, and Bradley J Erickson. "Deep learning for brain MRI segmentation: state of the art and future directions". In: *Journal of digital imaging* 30.4 (2017), pp. 449–459.

[196] Konstantinos Kamnitsas, Christian Ledig, Virginia FJ 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". In: *Medical image analysis* 36 (2017), pp. 61–78.

[197] Pim Moeskops, Max A Viergever, Adriënne M Mendrik, Linda S De Vries, Manon JNL Benders, and Ivana Išgum. "Automatic segmentation of MR brain images with a convolutional neural network". In: *IEEE transactions on medical imaging* 35.5 (2016), pp. 1252–1261.

[198] Martin Rajchl, Matthew CH Lee, Ozan Oktay, Konstantinos Kamnitsas, Jonathan Passerat-Palmbach, Wenjia Bai, Mellisa Damodaram, Mary A Rutherford, Joseph V Hajnal, Bernhard Kainz, et al. "Deepcut: Object segmentation from bounding box annotations using convolutional neural networks". In: *IEEE transactions on medical imaging* 36.2 (2016), pp. 674–683.

[199] Lorenzo Venturini, Aris T Papageorghiou, J Alison Noble, and Ana IL Namburete. "Multi-task CNN for Structural Semantic Segmentation in 3D Fetal Brain Ultrasound". In: *Annual Conference on Medical Image Understanding and Analysis*. Springer. 2019, pp. 164–173.

[200] Linde S Hesse, Moska Aliasi, Felipe Moser, Monique C Haak, Weidi Xie, Mark Jenkinson, Ana IL Namburete, INTERGROWTH-21st Consortium, et al. "Subcortical segmentation of the fetal brain in 3D ultrasound using deep learning". In: *NeuroImage* 254 (2022), p. 119117.

[201] Tatsat R Patel, Nikhil Paliwal, Prakhar Jaiswal, Muhammad Waqas, Maxim Mokin, Adnan H Siddiqui, Hui Meng, Rahul Rai, and Vincent Tutino. "Multi-resolution CNN for brain vessel segmentation from cerebrovascular images of intracranial aneurysm: a comparison of U-Net and DeepMedic". In: *Medical Imaging 2020: Computer-Aided Diagnosis*. Vol. 11314. SPIE. 2020, pp. 677–685.

[202] Magnus Ziegler, Jesper Alfraeus, Mariana Bustamante, Elin Good, Jan Engvall, Ebo de Muinck, and Petter Dyverfeldt. "Automated segmentation of the individual branches of the carotid arteries in contrast-enhanced MR angiography using DeepMedic". In: *BMC medical imaging* 21.1 (2021), pp. 1–10.

[203] Priscille de Dumast, Hamza Kebiri, Chirine Atat, Vincent Dunet, Mériam Koob, and Meritxell Bach Cuadra. "Segmentation of the cortical plate in fetal brain MRI with a topological loss". In: *Uncertainty for Safe Utilization of Machine Learning in Medical Imaging, and Perinatal Imaging, Placental and Preterm Image Analysis*. Springer, 2021, pp. 200–209.

[204] Xiao Han, Chenyang Xu, Ulisses Braga-Neto, and Jerry L Prince. "Topology correction in brain cortex segmentation using a multiscale, graph-based algorithm". In: *IEEE Transactions on Medical Imaging* 21.2 (2002), pp. 109–121.

[205] Florent Ségonne, Eric Grimson, and Bruce Fischl. "A genetic algorithm for the topology correction of cortical surfaces". In: *Biennial International Conference on Information Processing in Medical Imaging*. Springer. 2005, pp. 393–405.

[206] Rachel Aine Yotter, Robert Dahnke, Paul M Thompson, and Christian Gaser. "Topological correction of brain surface meshes using spherical harmonics". In: *Human brain mapping* 32.7 (2011), pp. 1109–1124.

[207] J Corbett-Detig, PA Habas, Julia A Scott, K Kim, V Rajagopalan, PS McQuillen, AJ Barkovich, Orit A Glenn, and C Studholme. "3D global and regional patterns of human fetal subplate growth determined in utero". In: *Brain Structure and Function* 215.3-4 (2011), pp. 255–263.

[208] Jean-François Mangin, Vincent Frouin, Isabelle Bloch, Jean Régis, and Jaime López-Krahe. "From 3D magnetic resonance images to structural representations of the cortex topography using topology preserving deformations". In: *Journal of Mathematical imaging and Vision* 5.4 (1995), pp. 297–318.

[209] Vanessa Kyriakopoulou, Deniz Vatansever, Alice Davidson, Prachi Patkee, Samia Elkommos, Andrew Chew, Miriam Martinez-Biarge, Bibbi Hagberg, Mellisa Damodaram, Joanna Allsop, et al. "Normative biometry of the fetal brain using magnetic resonance imaging". In: *Brain Structure and Function* 222.5 (2017), pp. 2295–2307.

[210] Hui-Hsin Hu, Wan-Yuo Guo, Hui-Yun Chen, Po-Shan Wang, Chih-I Hung, Jen-Chuen Hsieh, and Yu-Te Wu. "Morphological regionalization using fetal magnetic resonance images of normal developing brains". In: *European Journal of Neuroscience* 29.8 (2009), pp. 1560–1567.

[211] Anjali Sadhwani, David Wypij, Valerie Rofeberg, Ali Gholipour, Maggie Mittleman, Julia Rohde, Clemente Velasco-Annis, Johanna Calderon, Kevin G Friedman, Wayne Tworetzky, et al. "Fetal brain volume predicts neurodevelopment in congenital heart disease". In: *Circulation* 145.15 (2022), pp. 1108–1119.

[212] Feifei Xu, Mengting Liu, Sharon Y Kim, Xinting Ge, Zhonghe Zhang, Yuchun Tang, Xiangtao Lin, Arthur W Toga, Shuwei Liu, and Hosung Kim. "Morphological development trajectory and structural covariance network of the human fetal cortical plate during the early second trimester". In: *Cerebral Cortex* 31.10 (2021), pp. 4794–4807.

[213] Stephen E Jones, Bradley R Buchbinder, and Itzhak Aharon. "Three-dimensional mapping of cortical thickness using Laplace's equation". In: *Human brain mapping* 11.1 (2000), pp. 12–32.

[214] Bruce Fischl and Anders M Dale. "Measuring the thickness of the human cerebral cortex from magnetic resonance images". In: *Proceedings of the National Academy of Sciences* 97.20 (2000), pp. 11050–11055.

[215] Gang Li, Li Wang, Feng Shi, John H Gilmore, Weili Lin, and Dinggang Shen. "Construction of 4D high-definition cortical surface atlases of infants: Methods and applications". In: *Medical image analysis* 25.1 (2015), pp. 22–36.

[216] Joshua S Shimony, Christopher D Smyser, Graham Wideman, Dimitrios Alexopoulos, Jason Hill, John Harwell, Donna Dierker, David C Van Essen, Terrie E Inder, and Jeffrey J Neil. "Comparison of cortical folding measures for evaluation of developing human brain". In: *Neuroimage* 125 (2016), pp. 780–790.

[217] Eileen Luders, Paul M Thompson, KL Narr, Arthur W Toga, L Jancke, and Christian Gaser. "A curvature-based approach to estimate local gyrification on the cortical surface". In: *Neuroimage* 29.4 (2006), pp. 1224–1230.

[218] Suyash P Awate, Lawrence Win, Paul Yushkevich, Robert T Schultz, and James C Gee. "3D cerebral cortical morphometry in autism: increased folding in children and adolescents in frontal, parietal, and temporal lobes". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2008, pp. 559–567.

[219] Nicolaos Kapouleas. "Complete constant mean curvature surfaces in Euclidean three-space". In: *Annals of Mathematics* 131.2 (1990), pp. 239–330.

[220] Hao Huang, Rong Xue, Jiangyang Zhang, Tianbo Ren, Linda J Richards, Paul Yarowsky, Michael I Miller, and Susumu Mori. "Anatomical characterization of human fetal brain development with diffusion tensor magnetic resonance imaging". In: *Journal of Neuroscience* 29.13 (2009), pp. 4263–4273.

[221] Colin Studholme. "Mapping fetal brain development in utero using MRI: the big bang of brain mapping". In: *Annual review of biomedical engineering* 13 (2011), p. 345.

[222] Olivier Bernard, Alain Lalande, Clement Zotti, Frederick Cervenansky, Xin Yang, Pheng-Ann Heng, Irem Cetin, Karim Lekadir, Oscar Camara, Miguel Angel Gonzalez Ballester, et al. "Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved?" In: *IEEE transactions on medical imaging* 37.11 (2018), pp. 2514–2525.

[223] Nadine Kawel, Evrim B Turkbey, J Jeffrey Carr, John Eng, Antoinette S Gomes, W Gregory Hundley, Craig Johnson, Sofia C Masri, Martin R Prince, Rob J van der Geest, et al. "Normal left ventricular myocardial thickness for middle-aged and older subjects with steady-state free precession cardiac magnetic resonance: the multi-ethnic study of atherosclerosis". In: *Circulation: Cardiovascular Imaging* 5.4 (2012), pp. 500–508.

[224] Johann Benedikt Listing. *Vorstudien zur topologie*. Vandenhoeck und Ruprecht, 1848.

[225] Reinhard Klette and Azriel Rosenfeld. *Digital geometry: Geometric methods for digital picture analysis*. Morgan Kaufmann, 2004.

[226] Linda G Shapiro. "Connected component labeling and adjacency graph construction". In: *Machine intelligence and pattern recognition*. Vol. 19. Elsevier, 1996, pp. 1–30.

[227] Azriel Rosenfeld. "Connectivity in digital pictures". In: *Journal of the ACM (JACM)* 17.1 (1970), pp. 146–160.

[228] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. "Instance normalization: The missing ingredient for fast stylization". In: *arXiv preprint arXiv:1607.08022* (2016).

[229] Anderson M Winkler, Matthew A Webster, Jonathan C Brooks, Irene Tracey, Stephen M Smith, and Thomas E Nichols. "Non-parametric combination and related permutation tests for neuroimaging". In: *Human brain mapping* 37.4 (2016), pp. 1486–1511.

[230] Stefan Van der Walt, Johannes L Schönberger, Juan Nunez-Iglesias, François Boulogne, Joshua D Warner, Neil Yager, Emmanuelle Gouillart, and Tony Yu. "scikit-image: image processing in Python". In: *PeerJ* 2 (2014), e453.

[231] Carlo E Bonferroni. "Il calcolo delle assicurazioni su gruppi di teste". In: *Studi in onore del professore salvatore ortu carboni* (1935), pp. 13–60.

[232] George Papandreou, Liang-Chieh Chen, Kevin P Murphy, and Alan L Yuille. "Weakly-and semi-supervised learning of a deep convolutional network for semantic image segmentation". In: *Proceedings of the IEEE international conference on computer vision*. 2015, pp. 1742–1750.

[233] Qiuhui Chen and Yi Hong. "Scribble2D5: Weakly-Supervised Volumetric Image Segmentation via Scribble Annotations". In: *arXiv preprint arXiv:2205.06779* (2022).

[234] Pak-Hei Yeung, Ana IL Namburete, and Weidi Xie. "Sli2vol: Annotate a 3D volume from a single slice with self-supervised learning". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2021, pp. 69–79.

[235] Michela Antonelli, Annika Reinke, Spyridon Bakas, Keyvan Farahani, Annette Kopp-Schneider, Bennett A Landman, Geert Litjens, Bjoern Menze, Olaf Ronneberger, Ronald M Summers, et al. "The medical segmentation decathlon". In: *Nature communications* 13.1 (2022), pp. 1–13.

[236] Irfan Zeb, Dong Li, Khurram Nasir, Ronit Katz, Vahid N Larijani, and Matthew J Budoff. "Computed tomography scans in the evaluation of fatty liver disease in a population based study: the multi-ethnic study of atherosclerosis". In: *Academic radiology* 19.7 (2012), pp. 811–818.

[237] Rodrigo Santa Cruz, Léo Lebrat, Darren Fu, Pierrick Bourgeat, Jurgen Fripp, Clinton Fookes, and Olivier Salvado. "CorticalFlow++: Boosting Cortical Surface Reconstruction Accuracy, Regularity, and Interoperability". In: *arXiv preprint arXiv:2206.06598* (2022).

[238] Léo Lebrat, Rodrigo Santa Cruz, Frédéric de Gournay, Darren Fu, Pierrick Bourgeat, Jurgen Fripp, Clinton Fookes, and Olivier Salvado. "CorticalFlow: A Diffeomorphic Mesh Deformation Module for Cortical Surface Reconstruction". In: *arXiv preprint arXiv:2206.02374* (2022).

[239] Yu Sun Chung, Christopher J Hyatt, and Michael C Stevens. "Adolescent maturation of the relationship between cortical gyrification and cognitive ability". In: *NeuroImage* 158 (2017), pp. 319–331.

[240] Lisa Story and Mary Rutherford. "Advances and applications in fetal magnetic resonance imaging". In: *The Obstetrician & Gynaecologist* 17.3 (2015), pp. 189–199.

[241] Aris T Papageorghiou, Eric O Ohuma, Douglas G Altman, Tullia Todros, Leila Cheikh Ismail, Ann Lambert, Yasmin A Jaffer, Enrico Bertino, Michael G Gravett, Manorama Purwar, et al. "International standards for fetal growth based on serial ultrasound measurements: the Fetal Growth Longitudinal Study of the INTERGROWTH-21st Project". In: *The Lancet* 384.9946 (2014), pp. 869–879.

[242] J Villar, DG Altman, M Purwar, JA Noble, HE Knight, P Ruyan, L Cheikh Ismail, FC Barros, A Lambert, AT Papageorghiou, et al. "The objectives, design and implementation of the INTERGROWTH-21st Project". In: *BJOG: An International Journal of Obstetrics & Gynaecology* 120 (2013), pp. 9–26.

[243] Elizabeth Murray, Michelle Fernandes, Charles RJ Newton, Amina Abubakar, Stephen H Kennedy, Jose Villar, and Alan Stein. "Evaluation of the INTERGROWTH-21st Neurodevelopment Assessment (INTER-NDA) in 2 year-old children". In: *PLoS One* 13.2 (2018), e0193406.

[244] Michelle Fernandes, José Villar, Alan Stein, Eleonora Staines Urias, Cutberto Garza, Cesar G Victora, Fernando C Barros, Enrico Bertino, Manorama Purwar, Maria Carvalho, et al. "INTERGROWTH-21st Project international INTER-NDA standards for child development at 2 years of age: an international prospective population-based study". In: *BMJ open* 10.6 (2020), e035258.

[245] Xiujuan Geng, Gary E Christensen, Hong Gu, Thomas J Ross, and Yihong Yang. "Implicit reference-based group-wise image registration and its application to structural and functional MRI". In: *Neuroimage* 47.4 (2009), pp. 1341–1351.

[246] Ron Kikinis, Steve D Pieper, and Kirby G Vosburgh. "3D Slicer: a platform for subject-specific image analysis, visualization, and clinical support". In: *Intraoperative imaging and image-guided therapy*. Springer, 2014, pp. 277–289.

[247] Jinwoo Hong, Hyuk Jin Yun, Gilsoon Park, Seonggyu Kim, Cynthia T Laurentys, Leticia C Siqueira, Tomo Tarui, Caitlin K Rollins, Cynthia M Ortinau, P Ellen Grant, et al. "Fetal cortical plate segmentation using fully convolutional networks with multiple plane aggregation". In: *Frontiers in neuroscience* (2020), p. 1226.

[248] Davood Karimi, Serge Didenko Vasylechko, and Ali Gholipour. "Convolution-free medical image segmentation using transformers". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2021, pp. 78–88.

[249] Jonathan Z Bakdash and Laura R Marusich. "Repeated measures correlation". In: *Frontiers in psychology* 8 (2017), p. 456.

[250] Thomas Lewiner, Hélio Lopes, Antônio Wilson Vieira, and Geovan Tavares. "Efficient implementation of marching cubes' cases with topological guarantees". In: *Journal of graphics tools* 8.2 (2003), pp. 1–15.

[251] Anastasija Arechvo, Göran Lingman, Lars Thurn, Tomas Jansson, and Ligita Jokubkiene. "Fusion imaging in brain structure measurements on a fetus phantom, combining real-time ultrasound with magnetic resonance imaging". In: *Australasian Journal of Ultrasound in Medicine* 24.3 (2021), pp. 161–172.

[252] Anderson M Winkler, Peter Kochunov, John Blangero, Laura Almasy, Karl Zilles, Peter T Fox, Ravindranath Duggirala, and David C Glahn. "Cortical thickness or grey matter volume? The importance of selecting the phenotype for imaging genetics studies". In: *Neuroimage* 53.3 (2010), pp. 1135–1146.

[253] Zvi Kivilevitch, Reuven Achiron, and Yaron Zalel. "Fetal brain asymmetry: in utero sonographic study of normal fetuses". In: *American journal of obstetrics and gynecology* 202.4 (2010), 359–e1.

[254] John H Gilmore, Weili Lin, Marcel W Prastawa, Christopher B Looney, Y Sampath K Vetsa, Rebecca C Knickmeyer, Dianne D Evans, J Keith Smith, Robert M Hamer, Jeffrey A Lieberman, et al. "Regional gray matter growth, sexual dimorphism, and cerebral asymmetry in the neonatal brain". In: *Journal of Neuroscience* 27.6 (2007), pp. 1255–1260.

[255] Gregor Kasprian, Georg Langs, Peter C Brugger, Mario Bittner, Michael Weber, Mavilde Arantes, and Daniela Prayer. "The prenatal origin of hemispheric asymmetry: an in utero neuroimaging study". In: *Cerebral cortex* 21.5 (2011), pp. 1076–1083.

[256] R Hering-Hanit, R Achiron, S Lipitz, and A Achiron. "Asymmetry of fetal cerebral hemispheres: in utero ultrasound study". In: *Archives of Disease in Childhood-Fetal and Neonatal Edition* 85.3 (2001), F194–F196.

[257] Christine Chiarello, David Vazquez, Adam Felton, and Alessandra McDowell. "Structural asymmetry of the human cerebral cortex: Regional and between-subject variability of surface area, cortical thickness, and local gyrification". In: *Neuropsychologia* 93 (2016), pp. 365–379.

[258] Daniel John Cunningham. *Contribution to the surface anatomy of the cerebral hemispheres*. 7. Academy House, 1892. Chap. The Sylvian Fissure and the Island of Reil.

[259] Albert M Galaburda, Marjorie LeMay, Thomas L Kemper, and Norman Geschwind. "Right-Left Asymmetries in the Brain: Structural differences between the hemispheres may underlie cerebral dominance." In: *Science* 199.4331 (1978), pp. 852–856.

[260] Felipe Moser, Ruobing Huang, Bartłomiej W Papież, Ana IL Namburete, INTERGROWTH-21st Consortium, et al. "BEAN: Brain Extraction and Alignment Network for 3D Fetal Neurosonography". In: *NeuroImage* 258 (2022), p. 119341.

[261] Nicola K Dinsdale, Emma Bluemke, Vaanathi Sundaresan, Mark Jenkinson, Stephen M Smith, and Ana IL Namburete. "Challenges for machine learning in clinical translation of big data imaging studies". In: *Neuron* (2022).

[262] John Guttag and Adrian V Dalca. "HyperMorph: Amortized Hyperparameter Learning for Image Registration". In: *Information Processing in Medical Imaging: 27th International Conference, IPMI 2021, Virtual Event, June 28–June 30, 2021, Proceedings*. Vol. 12729. Springer Nature. 2021, p. 3.