

Delving deep into fetal neurosonography: an image analysis approach

Ruobing Huang

St.Peter's College
University of Oxford

*A thesis submitted for the degree of
Doctor of Philosophy*

Michaelmas 2017

Abstract

Ultrasound screening has been used for decades as the main modality to examine fetal brain development and to diagnose possible anomalies. However, basic clinical ultrasound examination of the fetal head is limited to axial planes of the brain and linear measurements which may have restrained its potential and efficacy. The recent introduction of three-dimensional (3D) ultrasound provides the opportunity to navigate to different anatomical planes and to evaluate structures in 3D within the developing brain. Regardless of acquisition methods, interpreting 2D/3D ultrasound fetal brain images require considerable skill and time.

In this thesis, a series of automatic image analysis algorithms are proposed that exploit the rich sonographic patterns captured by the scans and help to simplify clinical examination. The original contributions include:

- An original skull detection method for 3D ultrasound images, which achieves mean accuracy of $2.2 \pm 1.6mm$ compared to the ground truth (GT). In addition, the algorithm is utilised for accurate automated measurement of essential biometry in standard examinations: biparietal diameter (mean accuracy: $2.1 \pm 1.4mm$) and head circumference (mean accuracy: $4.5 \pm 3.7mm$).
- A plane detection algorithm. It automatically extracts mid-sagittal plane that provides visualization of midline structures, which are crucial to assess central nervous system malformations. The automated planes are in accordance with manual ones (within $3.0 \pm 3.5^\circ$).
- A general segmentation framework for delineating fetal brain structures in 2D images. The automatically generated predictions are found to be agreed with the manual delineations (mean dice-similarity coefficient: 0.79 ± 0.07). As a by-product, the algorithm generated automated biometry. The results might be further utilized for morphological evaluation in future research.
- An efficient localization model that is able to pinpoint the 3D locations of five key brain structures that are examined in a routine clinical examination. The predictions correlate with the ground truth: the average centre deviation is $1.8 \pm 1.4mm$, and the size difference between them is $1.9 \pm 1.5mm$. The application of this model may greatly reduce the time required for routine examination in clinical practice.

- A 3D affine registration pipeline. Leveraging the power of convolutional neural networks, the model takes raw 3D brain images as input and geometrically transforms fetal brains into a unified coordinate system (proposed as a Fetal Brain Talairach system).

The integration of these algorithms into computer-assisted analysis tools may greatly reduce the time and effort to evaluate 3D fetal neurosonography for clinicians. Furthermore, they will assist understanding of fetal brain maturation by distilling 2D/3D information directly from the uterus.

Delving deep into fetal neurosonography: an image analysis approach



Ruobing Huang
St.Peter's College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Michaelmas 2017

Acknowledgements

Personal

I am deeply grateful to my supervisor, Prof. Alison Noble, who has always been extremely helpful and patient with me. She provided me with unflinching encouragement whilst allowing me the room to work in my own way. It is her valuable guidance that lead me to the current state. I am grateful in every possible way.

My gratitude also goes to my collaborators, Dr. Ana Namburete and Weidi Xie, for valuable discussions and suggestions on my research. A special thank you to Dr. Ana Namburete, who changed my DPhil life. She has triggered and nourished my intellectual maturity that I will benefit from, for a long time to come. I would also like to thank Weidi Xie again, for not only be a knowledgeable collaborator, but also a dear friend, who always encouraged me and helped me with my code in my darkest hour.

I would also like to thank my colleagues at the Biomedical Image Analysis lab. In particular, Dr. Davis Vigneault, who helped me proofreading till the last minute and who recommended so many amazing podcasts to me that I am not able to finish without another PhD; Dr. Christopher Bridge, for being a walking Wikipedia and enlightened me so much about signal processing and British culture; Dr. Rui Jia, for the trip to MIUA we spent together, and for all the laughs and support you gave me. I would also like to thank Yuan Gao, Milar Omar, Huan Qi, Yifang Cai and Dr. Yipeng Hu in my lab. I am indebted to them for their help.

My sincerest thanks goes to my wonderful room-mate Di Shen, who was my closest companion during my DPhil and helped me go through all the nervous ‘breakdowns’ in the past two years. Meeting you has been one of the luckiest things that happened to me in Oxford. I would also like to thank my dearest friends Jingwen Cao, Geyuan Li, Yi Wang and Meiling Sun, who cannot always see me but always be there for me when I needed.

Finally, my deepest gratitude goes to my families. My parents, Mai Xu and Kun Huang, who might care more about my research progress than myself. It was with your unfailing support and selfless love that kept me going. I want to thank my grandparent Dongzhen Yang and Changhu Xu, who gave me my name made me what I am today.

Institutional

I would like to acknowledge St.Peter's College and Bill and Melinda Gates foundation for assisting my study at Oxford. I am also grateful to the INTERGROWTH 21-st project which provide valuable data and assistance for my research.

Abstract

Ultrasound screening has been used for decades as the main modality to examine fetal brain development and to diagnose possible anomalies. However, basic clinical ultrasound examination of the fetal head is limited to axial planes of the brain and linear measurements which may have restrained its potential and efficacy. The recent introduction of three-dimensional (3D) ultrasound provides the opportunity to navigate to different anatomical planes and to evaluate structures in 3D within the developing brain. Regardless of acquisition methods, interpreting 2D/3D ultrasound fetal brain images require considerable skill and time.

In this thesis, a series of automatic image analysis algorithms are proposed that exploit the rich sonographic patterns captured by the scans and help to simplify clinical examination. The original contributions include:

- An original skull detection method for 3D ultrasound images, which achieves mean accuracy of $2.2 \pm 1.6mm$ compared to the ground truth (GT). In addition, the algorithm is utilised for accurate automated measurement of essential biometry in standard examinations: biparietal diameter (mean accuracy: $2.1 \pm 1.4mm$) and head circumference (mean accuracy: $4.5 \pm 3.7mm$).
- A plane detection algorithm. It automatically extracts mid-sagittal plane that provides visualization of midline structures, which are crucial to assess central nervous system malformations. The automated planes are in accordance with manual ones (within $3.0 \pm 3.5^\circ$).
- A general segmentation framework for delineating fetal brain structures in 2D images. The automatically generated predictions are found to be agreed with the manual delineations (mean dice-similarity coefficient: 0.79 ± 0.07). As a by-product, the algorithm generated automated biometry. The results might be further utilized for morphological evaluation in future research.
- An efficient localization model that is able to pinpoint the 3D locations of five key brain structures that are examined in a routine clinical examination. The predictions correlate with the ground truth: the average centre deviation is $1.8 \pm 1.4mm$, and the size difference between them is $1.9 \pm 1.5mm$. The application of this model may greatly reduce the time required for routine examination in clinical practice.

- A 3D affine registration pipeline. Leveraging the power of convolutional neural networks, the model takes raw 3D brain images as input and geometrically transforms fetal brains into a unified coordinate system (proposed as a Fetal Brain Talairach system).

The integration of these algorithms into computer-assisted analysis tools may greatly reduce the time and effort to evaluate 3D fetal neurosonography for clinicians. Furthermore, they will assist understanding of fetal brain maturation by distilling 2D/3D information directly from the uterus.

Contents

List of Figures	xi
List of Abbreviations	xxi
1 Introduction	1
1.1 Clinical motivation	1
1.2 Contribution and Thesis Structure	3
1.3 Originality	4
1.4 List of Publications	7
1.4.1 Journal Articles	7
1.4.2 Peer-Reviewed Conference Proceedings	7
1.4.3 Conference Presentations	7
2 Review of fetal brain ultrasound image analysis	9
2.1 Skull stripping	10
2.1.1 Clinical Background	10
2.1.2 Related works	11
2.2 Mid-sagittal plane	13
2.2.1 Clinical background	13
2.2.2 Related works	14
2.3 Delineating brain structures in 2D images	15
2.3.1 Clinical background	15
2.3.2 Related works on ultrasound image analysis	19
2.4 Localizing multiple structures in 3D fetal neurosonography	21
2.4.1 Clinical background	21
2.4.2 Related works on ultrasound image analysis	24
2.5 Registering fetal brain in 3D neurosonography	27
2.6 Database	29
2.7 Conclusion	31

3	Visualization in fetal neurosonograms: automated skull-stripping and plane detection	33
3.1	Skull stripping in ultrasound images	34
3.1.1	Proposed method	34
3.2	Mid-sagittal plane detection	37
3.3	Experiments	41
3.3.1	Image dataset	41
3.3.2	Skull detection accuracy	41
3.3.3	Manual plane selection	42
3.3.4	Orientation deviation	43
3.3.5	Shape Analysis	43
3.4	Results and discussions	44
3.5	Conclusion	46
4	Segmentation of Key Clinical Anatomical Structures in 2D Ultrasound Images	49
4.1	Introduction	50
4.1.1	Proposed framework	51
4.2	Method	52
4.2.1	Target Fetal Brain Structures	52
4.2.2	Candidate Region Selection	53
4.2.3	Region Descriptor	54
4.2.4	Classification and voting mechanism	57
4.3	Implementation details	58
4.4	Datasets	59
4.5	Experiments	60
4.5.1	Classification and feature comparison	60
4.5.2	Segmentation accuracy	61
4.6	Results	61
4.6.1	Classification performance	61
4.6.2	Feature set comparison	62
4.6.3	Segmentation accuracy	66
4.7	Discussion	67
4.7.1	Challenging cases	68
4.7.2	Failure cases analysis	70
4.8	Summary	71

5	Localizing Multiple Brain Structures in 3D Ultrasound Images	73
5.1	Introduction	74
5.2	View-based Projection Network (VP-Net)	78
5.2.1	Preprocessing	78
5.2.2	Network architecture design	78
5.2.3	Detection as Segmentation	83
5.2.4	Visualizing the 3D Saliency Volume	85
5.2.5	Implementation details	85
5.3	Experiments	86
5.4	Results and Discussion	90
5.4.1	Qualitative Evaluation	90
5.4.2	Quantitative evaluations	91
5.4.3	Saliency Volume	98
5.4.4	Challenging cases	98
5.5	Conclusions	101
6	Automatic Similarity Registration of Fetal Brain : A Pilot Study	103
6.1	Introduction	104
6.2	Fetal Brain Talairach (FBT) coordinate system	105
6.2.1	Fetal Corpus Callosum	107
6.2.2	Establishing the unified coordinate system	108
6.3	Fetal brain similarity registration with 3D CNNs	110
6.3.1	Learning to align brains in the axial view	111
6.3.2	Learning rotations in three anatomical views	116
6.3.3	A full similarity 3D registration pipeline	123
6.4	Conclusion and future study	126
7	Conclusion	129
7.1	Contributions	129
7.2	Limitations and Future Work	132
7.3	Summary	135
Appendices		
References		143

List of Figures

2.1	Typical fetal head measurement in axial slice. The skull appears as a bright ellipse. The dashed line depicts the head circumference(HC), while the two headed arrows depicts the biparietal diameter (BPD).	10
2.2	The skull detection using morphological operations and Hough transform based on [18]. (a) shows the 2D axial plane in ultrasound image. (b) shows the image been classified into three class. (c) shows the extracted skeleton of bright object after morphological operations. (d) shows the process of randomized Hough transform. (e) shows the final ellipse pick by the framework. (f) shows the ellipse superimposed on the original axial slice.	12
2.3	Example of false skull detection using morphological operations and Hough transform. (a) shows the K-means algorithm pick out skull as well as the maternal tissues. (b) shows the skeleton of bright objects which contains distracting information. (c) shows the final ellipse detected by the framework which tried to fit the false skeleton and did not fit the skull properly.	13
2.4	Schematic of the mid-sagittal plane of a brain in 3D. The plane is plotted in transparent blue.	14
2.5	Fetal neurosonogram of a healthy fetus with gestation age of 21 weeks. (a) is a typical example of the mid-sagittal plane (MSP) observed in a ultrasound image. The corresponding location of skull and eye are pointed out. (b) and (c) are the MSP visualized in axial and coronal slice (as a bright thin line) respectively. Large portion of maternal tissue existed in all these views.	16
2.6	Two randomly selected examples of fetal brain ultrasound images. It can be seen that half of the brain has better visibility (pointed by white arrow).	16
2.7	Examples of normal brain in MRI and PET images. Both the brains exhibit good symmetry and there is not adjacent tissues outside the skull. Image credit to https://radiopaedia.org for the MRI image, and http://www.traumaticbraininjury.net for the PET image. . . .	17

2.8	Examples of MRI images of subjects with normal or abnormal corpus callosum (pointed by a white arrow). The images are listed in the order of the severity of the abnormality. The left-most image is a normal subject while the right-most has the most severe malformation: complete absence of corpus callosum. It can be seen there are various shades of abnormality. A simple identification of whether a structure exists might be inadequate to detect intermediate abnormality. Image credited to https://www.nimh.nih.gov , Elgamal et al. [43], and https://emedicine.medscape.com/article/407730-overview	18
2.9	Trans-ventricular plane of two randomly selected fetal ultrasound scans. The red dashed boxes highlight the anterior horn of lateral ventricle for each subject. Compare the two red boxes, it can be seen the appearance of the image patch differs dramatically in the same location of the same structure across different subjects. The yellow dashed boxes in the image of subject label a patch of brain tissues and a patch of maternal tissue outside the brain. The texture and intensity level of the two patches are visually similar.	22
2.10	Demonstration of key brain structures. The three sub-figures represent different anatomical views: axial, sagittal and coronal. The first column is the schematic of these structures: CSP (blue), LV (green), Tha (yellow), CB (red) and CM (purple). Two schematic of the coronal view are included for better visualization. The rest of columns are these structures observed in ultrasound images. Their locations are pointed out by white arrows.	23
2.11	A simple example of neural networks. It consists of three layers: input (red), hidden (blue) and output (green). Neurons are shown as circles and their connection (weights) are plotted as arrows. Image source: https://goo.gl/images/1RXTyH	26
2.12	An example of Convolutional neural network (CNN). An input is followed by interwoven convolutional and pooling layers. The final convolutional layer is flattened and passed to a fully-connected layer to generate output predictions. Image source: https://goo.gl/images/RaUtB9	27
2.13	Map of INTERGROWTH-21st research units and collaborating Institution. Research unit are labelled in red while collaborating centre are labelled in green.	30
3.1	Examples of ultrasound skull images. Subjects exhibit different contrast, skull shapes and orientations.	35

3.2	Challenging conditions present in identifying skull in fetal ultrasound images. The red lines point out large gap between skull segments, bright tissues that is closely adjacent of skull and different brightness of skull segments.	35
3.3	Skull detection process. The skull in each axial slice is detected through templates matching and the extracranial tissues are removed using the skull contour.	36
3.4	Schematic of the detection of the skull in 3D. The detected skull in each axial slice is plotted in red. The fitted ellipsoid is plotted in green.	38
3.5	Pipeline of the proposed tool for automated mid-sagittal plane selection. (a) Illustration of horizontal slices selection. (b) skull detection in an axial slice. (c) Detected midlines in a stack of axial slices. (d) Detected mid-sagittal plane.	38
3.6	Manual plane selection. The midline in an axial slice (yellow) and that of a coronal slice (blue) are identified to guide the localization of the mid-sagittal plane in 3D.	42
3.7	CC shape model. (a) The configuration of the fetal CC (b) Eight representative points (pink) are manually annotated on the boundary of the CC (green).	43
3.8	Planes' orientation deviation. The axis represents fetal GA versus the orientation deviation. Box plot showing median (red lines) and inliers within ± 2.7 standard deviation.	45
3.9	CC shape analysis. The red line depicts the CC in manually selected plane and the blue one depicts that in automatically selected plane.	46
4.1	Schematic flow chart of automated segmentation framework.	51
4.2	Trans-thalamic plane (first row) and mid-sagittal plane (second row) of the fetal brain. Schematics for CP and CC are shown in (a) and (e) respectively. Other sub-figures in each row are corresponding ultrasound image examples. The CP or CC is bounded by a yellow dashed box in each image.	52
4.3	Candidate region detection. (a) shows the original fetal ultrasound image of mid-sagittal plane. The CC, the spine and the nose of the fetus are pointed out. (b) displays the manual delineation of the CC in yellow. (c) presents the detected MSERs, where the positive example is plotted in red and the negative examples are randomly assigned with blue or green for comparison. It can be seen that the majority of MSERs belong to negative class.	55

4.4	Shape Descriptor 1: pre-processing procedure and feature extraction. Shape Descriptor 1 is derived from NCC values, and its value is proportional to the shape similarity between candidate region and templates. The column of templates shows six examples of CC templates used in the experiments.	56
4.5	Shape Descriptor 2: The area distribution of the candidate region on a polar coordinate system. The colour red represents a large number of pixels lies within the fan section, while blue represents a small number.	56
4.6	Receiver Operating Characteristic (ROC) curves for different classifiers, namely SVM, Adaboost and RUSBoost assigned with different SOI: CC (left) and CP (right) segmentation tasks. The blue solid line represents RUSBoost classifier, while the curve for AdaBoost is plotted using orange dashed line and yellow dotted line for SVM. .	62
4.7	The region classification accuracy (purple) and precision (yellow) are plotted with respect to fetal gestational age (GA). The number of volumes ranged from 15 to 25 per GA week.	63
4.8	The classification error using different feature sets are plotted against learning iterations (shape features: yellow, other feature: blue and the combination: orange).	63
4.9	Manual and automatic segmentation of CP. Original ultrasound image is shown in the far-left. Sub-image of Choroid Plexus (CP) is enlarged to visualize the details. Manual delineation (yellow) and automatic generated segmentation (red) of CP are also displayed. The two delineations are plotted in the right-most image for comparison.	63
4.10	CC automatic segmentation results for 5 different cases. Each row belong to one patient. The first column are the original ultrasound images, the corresponding automatic segmentations are presented in the second columns in light blue. The third column shows the CC length measurements carried out by two different human experts and those derived from automatic segmentation (expert a: red line, expert b: blue line, automatic result: green line).	64
4.11	Biometric measurements evaluation. (a) The boxplot of measurements deviation of CC length between observers and automated results. (b) Counterpart boxplot for CP diameter measuring.	65
4.12	CP automatic segmentation results. Each row displays a single subject. The CP region and the segmentation results are enlarged for easier examination. The manual segmentation is plotted in yellow while the automatic result is plotted in red. The two outlines (for each case) are then overlay on the same image for comparison. The value of the DSC and d_h is also displayed in the upper-right corner.	69

- 4.13 The two failure case of CP segmentation task within the test dataset. The CP image is enlarged for closer examination. The manual delineation is plotted in yellow. 70
- 5.1 (a) Schematic of the detected structures in the right-hemisphere of the brain in 3D. Each structure is plotted as: cavum septi pellucidi (CSP) in blue, thalami (Tha) in yellow, lateral ventricles (LV) in green, cerebellum (CE) in red and cisterna magna (CM) in purple. (b) 3D spatial configuration of the bounding boxes of the target brain structures. Each box is shown as a 3D cube whose color is in accord with that in sub-figure a. Notice that there is geometric overlap between CE and CM, as well as CSP and Tha. (c) Targeted structures in sagittal planes of a US volume. The structures is bounded by white dashed boxes. 75
- 5.2 Networks design details: (a) An overview of the CNN-based image analysis pipeline. A raw ultrasound volume is preprocessed using the skull-stripping tool (refer to Section 5.2.1). The masked brain image and two of its transformed (90° rotation) volume are passed, in parallel, to three independent VP-Nets. Each VP-Net outputs the projection mask of five structures (15 projected masks for 5 structures). The minimal rectangle that encloses the prediction silhouette is detected for each structure in each view, to generate the bounding box parameters. The 3D bounding box for a structure can be reconstructed with two obtained rectangle masks with backward projection. **A graphical demo for the VP-Net model is available at: <https://youtu.be/KVxkbqWYWxc>.** (b) Each individual VP-Net shares the same spirit as 2D U-Net [122] for full-resolution prediction. In VP-Net, convolution filters in all layers only scan along the 2D X-Y plane, while always penetrating along the Z-axis to capture contextual information along that dimension. . . . 77
- 5.3 **View-based Projections.** (a) Cerebellum (CE) in different views. Yellow boxes refer to the cross section of 3D human expert annotation. (b) Schematic of orthogonal projection. The cerebellum is highlighted in purple and bounded by a gray cube. Its orthogonal projections on axial plane (horizontal) and sagittal plane (vertical) are shown respectively. The combination of the two rectangles defines the 3D bounding box of the object. 79

- 5.4 Schematics of different convolutions. M, N, D represents height, width and depth respectively. k represents the kernel size in one dimension. (a) 2D convolutions. The process is calculated in 2 direction (x,y). The output is a 2D matrix. (b) 3D convolutions. The convolution is calculated in 3 direction (x,y,z). Kernel depth d satisfies $d < D$, and is usually set to the same value as k . The output is a 3D volume. (c) VP-Net convolutions. The kernel depth is equal to the input depth D . Note that the output is also a 2D matrix. 81
- 5.5 An example of 3D CNN architecture. Image source: Milletari et al. [129]. 82
- 5.6 Rectangle detection denotation. The prediction mask T is plotted in solid blue, and its contour is plotted in dashed blue line. The centroid $p(x_c, y_c)$, the orientation ϕ , the width $2da_m$ and the height $2db_m$ are labelled respectively. The black box denotes the detected minimal rectangle. 84
- 5.7 Back-ward projection. The red and blue are rectangle masks detected from two different views. The 3D bounding box can be obtained by intersection of their projection (dark purple). Note that here a rectangle mask is not stacked along the whole volume (as described in text) for a better visualisation of the intersection region. 85
- 5.8 Example of predicted bounding boxes of targeted structure in different anatomical views. Yellow box show the ground truth, and Red boxes are predictions. **First to last row:** Results from Random Forest, 3D U-Net, VP-Nets. **First to last column:** Results from Axial, Coronal, Sagittal views. Note that there are overlap between CE and CM in all the views (as shown in a and b). Tha coincides with CE and CM in coronal view as well. The results show the prediction of VP-Nets coincided better with ground truth than that of 3D U-nets and random forest. 93
- 5.9 Mean 2D & 3D IOU curves of 3D U-Nets (yellow) and VP-Nets (other colors). (b) shows that the 3D U-Nets performed similarly with other VP-Nets on coronal view in 2D IOU. All VP-Nets achieved higher accuracy when combine the three views together (e). Comparing different variants of VP-Nets, Model F(black) achieved the highest accuracy. It shows that larger feature channel and deeper layers can increase the network performance. 94

5.10	Visualization of a 3D saliency volume derived from CE detection in the axial view. The right-hand figure shows the points with top 2% gradient values in 3D. The color red represents higher gradient values, indicating a higher influence on the prediction outcome. Sub-figure (a)(b) shows a axial and a sagittal US slice with 3D saliency points projected on each view respectively. The images and the original US are zoomed in for detail visualization.	99
5.11	Challenging case of CE in test sets. (a) shows the schematic of anatomical configuration of CE in axial plane. The white eyeglass-shaped structure is the CE. (b) is the US slice of the discussed case. Half of the CE is blocked by acoustic shadows. (c) shows the ground truth bounding box of CE (overlaid in yellow). The confidence map predicted by VP-Nets F is shown as red in (d). In this case, IoU of the raw prediction map and the ground truth is 68%.	100
5.12	Challenging case of CC in test sets. (a) is the schematic of CC in mid-sagittal plane. The CC is shown as the blue comma-shape structure. The corresponding mid-sagittal slice of the US volume is given in (b). The ground truth is plotted as transparent yellow rectangle in (c). Similarly, the predicted confidence map is shown as red rectangle in (d).	101
5.13	Challenging case of LV in test sets. (a) is the schematic of LV in axial plane. (b), (d) show the corresponding axial image of the third and the fourth case respectively. No ground truth is displayed for (b) as the human expert reckoned LV is not visible in this case. (e) show the ground truth label of the LV for the fourth case. The predicted confidence map is superimposed on the corresponding ultrasound image in (c) and (f).	102
6.1	Montage of mid-axial slices from randomly selected subjects. All images is cropped in the image centre (a 150×150 pixel window) for better visualization.	106
6.2	Illustration of the anterior commissure (AC) and posterior commissure (PC) in the brain. The AC and PC are marked with orange circles. Image credited to Lilla Zöllei, slides for WhyNHow seminar 12/04/08.	107
6.3	The human corpus callosum. (a) The corpus callosum locates at mid-sagittal section of the brain. Image source: Life Science Databases from Anatomography via Wikipedia. (b) The human CC consists of three parts: the genu, the body and the splenium. Image resource: Pre AP Biology Study Guide with Mrsedward at Mountain View College.	107

- 6.4 Illustration of the spatial location of the corpus callosum (CC) and the cavum septic pellucidi (CSP) in the brain. The two are closely adjacent and superior to the cingulate gyrus. The thalamus located beneath them is also labelled. 108
- 6.5 The fetal corpus callosum in ultrasound images. (a) 2D ultrasound images of fetal CC, observed as a dark thin strip. (b) 3D ultrasound image demonstrates a single comma-shaped echogenic structure representing the CC-CSP complex. 109
- 6.6 The mid-sagittal plane of a 3D fetal brain ultrasound image. The reference landmarks for the Fetal Brain Talairach (FBT) coordinate system are marked in red cross. The line between the two landmarks (showed as white dashed line) defines the y-axis. The CC-CSP complex and the thalamus are highlighted by white arrows. 109
- 6.7 The established FBT system shown from four different angles. The vertical plane in each image is the mid-sagittal slice while the horizontal plane is the trans-thalamic plane. The skull boundary is represented in green ellipse frame. The reference landmarks (endpoints of the CC-CSP complex are marked in red. The x,y,z axis of the system is plotted in white dashed line. 111
- 6.8 An example of fetal thalamus observed in ultrasound slices. (a) In the mid-sagittal slice, the thalamus is the structure beneath the CC-CSP complex (both pointed out by a white arrow). The thalamus does not possess a clear boundary in the mid-plane and appears as an obscure blob. (b) In the axial view, the thalamus consists of two symmetrical dark spots that combined in a butterfly shape. It is easier to identify the thalamus in this view for this unique pattern. 112
- 6.9 Sketch of the proposed 3D model. The basic architecture is similar to a 3D U-Nets, and takes an ultrasound brain image as input (black). The output consists of three parts: the rotation deviation of the brain in the axial view, an image transformed by the STN, bounding box masks of 5 key brain structures. Different neural network blocks are plotted in different colours (shown in the right of the figure). . . . 114
- 6.10 Schematic of the 3D sphere mask created for reference landmarks in the mid-sagittal plan. The landmarks are plotted in transparent red circles respectively. 118

- 6.11 Schematic of the proposed models. (a) is the Half-Net A. It consists of several convolutional block (convolution $\wedge$ batch normalization $\wedge$ ReLu activation $\wedge$ Max-pooling layer), a STN module (pink) and prediction layer. The model takes an ultrasound brain image as input (black), and outputs the rotation angles of the image and an image transformed by the STN. (b) is the Half-Net B model. Note that the only difference is that the STN outputs three orientation estimations (highlighted by a red dashed box) instead of one. (c) is the Full-Net model. It is similar to the model shown in Fig.6.9, while the expansive stream only have one output: the CSP landmarks masks (the right most block in (c)). 120
- 6.12 Results of the Full-Net model. Each row shows results of each anatomical view: axial - blue, coronal - green, and sagittal - purple. The sub-figure in the left in each row shows the plot of angle deviation between prediction and ground truth. The x-axis is the ground truth deviation in degrees, while the y axis is the predicted angles in degrees. Points locates close to $y = x$ represent a good prediction. A straight line that best fits the scatters is calculated using logistic regression. The sub-figures in the right column are Bland–Altman plots, where the y-axis is the difference between the ground truth and the prediction. 121
- 6.13 Plots of the age of each subject (x axis) versus the absolute value of its mean angle deviation (y axis). The figure shows that the model performed similarly across the gestation age despite the contrast and structural variance observed among different age groups. Note that the y-axis limits is set as $0 \rightarrow 4^\circ$ to zoom in on the difference between subjects. 123
- 6.14 Comparison of typical example of ultrasound scans collected at different gestational age (GA). The trans-thalamic plane of fetal head are selected for anatomical consistency. (a) is a scan collected at the 18th gestational week while (b) is a scan collected at the 30th gestational week. It can be seen that the image of the younger fetus has higher contrast and the skull size is smaller. The image of older fetus has lower contrast and the anatomy has more complex structures due to brain maturation. This is a common pattern observed in fetal neurosonography. 123

- 6.15 Schematic of proposed pipeline. During training, an image is pre-aligned and processed with suitable orientation simulation and data-augmentation (e.g. rotation and cropping). The CNN module is trained to predict the orientation of the brain, segment the skull and detect the landmarks. During testing, an unseen image is cropped in the middle and passed into the trained CNN model. The image is subsequently rotated based on the predicted orientation angles and masked by the predicted skull mask. The output is a rigid aligned brain image based on the FBT standard. 125
- 6.16 Examples of the similarity registration results from the test set. The first row for each subject is the input image, the second row is the transformed image based on the prediction. Each column is the middle slice of the volume in each dimension. The input image of Example a has the largest axial rotation deviation ($GT = 17^\circ$) in the test set. The input of Example b has the largest mean angle deviation in the test set: 12° in axial, 18° in sagittal, and 1° in coronal. The prediction result demonstrates the model handles well on the subjects with large orientation and deviations in multiple views. . . 127
- 6.17 The case of the largest mean orientation prediction error in the test set. It might be explained that the image quality is not ideal thus obstructing structural visibility. 128

List of Abbreviations

2-D,3-D	Two- or three-dimensional, referring in this thesis to spatial dimensions in an image.
AC,PC	Anterior commissure and posterior commissure.
AUC,ROC . .	Classifier evaluation metrics/plot: i.e. area Under the Curve and receiver Operating Characteristic.
BPD,HC,CRL	Common fetal biometry, i.e. biparietal diameter, head circumference and crown-rump length measurement.
LMP	Common fetal biometry, i.e. biparietal diameter, head circumference, and the last menstrual period
CB,CC,CSP .	Fetal brain structures, i.e. cerebellum, corpus callosum, and cavum septi pellucidi.
CM,LV,Tha .	Fetal brain structures, i.e. cisterna magna, lateral ventricles, and thalamus.
CNN,FCN,STN	Deep learning architectures, i.e. convolution neural network, fully convoultion network, and spatial transformer network.
CNS	Central nervous system.
DSC	Dice similarity coefficient.
FASD	Fetal alcohol spectrum disorder.
FBT	A standard coordinate system for fetal brain: Fetal Brain Talarirach system.
GA, LMP . . .	Gestational age and and the last menstrual period.
GT	Ground truth.
IOU	Intersection over Union.
LSE	Least square estimation.
MRI,PET . . .	Common imaging modality, i.e. magnetic resonance imaging and positron-emission tomography.
MSER	Maximally Stable Extremal Region.

MSP	The Mid-sagittal plane.
NCC,RANSAC	Data processing algorithms, i.e. normalized cross-correlation and random sample consensus algorithm.
RF,RUSBoost	Machine learning algorithms, i.e. random Forest and Random Under-Sampling algorithm.
SOI	Structures of interest.
TC,TT,TV . .	Standard fetal neurosonography planes, i.e. trans-cerebellar, trans-thalamic, and trans-ventricular plane.
VP-Net	View-based projection network.

1

Introduction

Contents

1.1	Clinical motivation	1
1.2	Contribution and Thesis Structure	3
1.3	Originality	4
1.4	List of Publications	7
1.4.1	Journal Articles	7
1.4.2	Peer-Reviewed Conference Proceedings	7
1.4.3	Conference Presentations	7

1.1 Clinical motivation

The human brain is among the most complicated organs in a human body. Originating in the embryo from a simple neural tube, the brain gradually differentiates and extends to the final elaborate structure in late adolescence [1]. The fetal period is regarded as one of the most crucial developmental stages, during which the gross morphology of the developing brain undergoes striking changes. For decades, ultrasound has been the standard modality to assess brain health during gestation. It is economical, radiation-free, and allows real-time visualization [2]. Since its introduction in the late 1950s, it has progressively become an indispensable obstetric tool and plays an important role in neonatal care.

For fetal brain evaluation in particular, a basic ultrasound examination is usually performed around 20 weeks' gestation. Based on clinical guidance, it comprises both qualitative and quantitative evaluations [3]. Three axial planes: the trans-ventricular (TV), trans-cerebellar (TC), and trans-thalamic (TT) planes, are selected to identify the skull and key cerebral structures. Several important structures of note are: the lateral ventricles (LV), thalamus (Tha), cerebellum (CB), cisterna magna (CM), and cavum septi pellucidi (CSP). The head shape is assessed, and its diameter and circumference are measured as bio-markers of the subject's age and growth.

Although these are valuable sources of anatomical reference, we still lack knowledge regarding the three-dimensional (3D) volumes and shapes of structures within the developing human brain. The invention and clinical introduction of 3D ultrasound brings the possibility of visualizing planes that are difficult to obtain through 2D acquisition. It also allows surface rendering of anatomies, post-exam data processing and review [4]. In addition, 3D acquisition decreases the examination time and helps to reduce the inter- and intra-observer variations of biometric measurements, especially for less experienced sonographers [5].

However, navigating a 3D volume to find the desired planes or structures is time-consuming and tedious. The process requires a sonographer to recognize artifacts caused by 3D reconstruction and variable acquisition settings [6]. Specifically, the acquisition of desired planes and structures poses several challenges:

1. The ultrasound images intrinsically have low signal-to-noise ratio, speckles and other artifacts. The image quality varies further with gestation.
2. The anatomical structures in ultrasound images have large biological variations across the gestational period.
3. The position of the fetus is highly variable. The fetal head can be imaged from any direction in theory (though acquisition standards recommend certain angles of insonification). Furthermore, it is not uncommon to see fetus in a reverse position.
4. The acquisition of the correct plane and identification of certain features are heavily reliant on the skill of the sonographer [7].

To address the above issues, this thesis aims to use image analysis and machine learning techniques to exploit the rich sonographic information contained, but not yet fully exploited, in fetal neurosonograms. The main objectives seek to deepen the understanding of ultrasound images from the image processing perspective, and contribute in the long term to further promote the use of 3D ultrasound in monitoring fetal brain development.

1.2 Contribution and Thesis Structure

The main contribution of this thesis is a series of automatic image analysis tools to analyse fetal brain in ultrasound images.

Chapter 2 introduces the clinical motivation for each subsequently developed tool. It elaborates challenges and limitations that exist in current fetal neurosonography examinations. It then describes possible image analysis solutions and reviews progress made in related literatures.

Chapter 3 presents an original skull detection algorithm to remove extra-cranial tissues and a plane detection pipeline for visualising the mid-sagittal profile for 3D fetal neurosonograms. The exclusion of maternal tissues reduce the effect of background images in subsequent processing and analysis. This chapter describes an original detection approach based on template matching and a modified version by manipulating resolution. Validation experiments are designed and tested on 248 2D fetal ultrasound images. The algorithm is also utilized in a plane detection pipeline to extract mid-sagittal profile. The derived plane facilitates inspection of midline cerebral structures which are crucial in diagnosis of central nervous system disorders. The pipeline is evaluated on 62 3D scans collected from healthy fetuses. To further validate the proposed method, intra-/inter-observer annotations are calculated and compared accordingly. The results show that the automated plane detection pipeline has comparable error margins with that of intra-/inter-observer deviations. The pipeline also performed similarly across a broad age range.

Chapter 4 describes a general framework for automatic segmentation of fetal brain structures in 2D ultrasound images, simultaneously yielding an automated

biometry measurement. The approach is based on a novel region descriptor that characterizes the shape and local intensity context of different neurological structures without explicit models. The framework is validated by segmenting two fetal brain structures of clinical importance with different ultrasonic appearances - the corpus callosum (CC) and the choroid plexus (CP). Discussion of the results further reveals the difficulty and applicability of structure delineation in ultrasound images.

Chapter 5 proposes a novel convolution neural network (CNN) model: View-based Projection Networks (VP-Nets) localize multiple structures in 3D ultrasound images. The model uses projection to reduce the dimensionality, and thereby the memory consumption, of a 3D CNN model. As a result, it allows full-resolution predictions with fast inference during test. The tool is validated on a dataset of 285 3D ultrasound scans. The model is also compared with state-of-the-art techniques, i.e. Random Forest and 3D U-Nets. The effect of model depth and feature channels and the importance of the loss function is also investigated.

Chapter 3 - 5 have different objectives, but they face one common problem: the varying positions, orientations and scales of the fetal brain within the volumes. Instead of crafting features or designing algorithms that are robust to these variations, **Chapter 6** describes how to reducing this variation by "wrapping" brains into the same coordinate space. A standard coordinate system is introduced: the Fetal Brain Talarirach (FBT) coordinate system for fetal neorosonography. Three consecutive experiments are reported on fetal brain registration in the context of ultrasonography using CNN models. This pilot study reveals the ability of a CNN model to learn the geometric transformation and the importance of a balanced supervision in a multi-stream task.

The thesis concludes in **Chapter 7**. Aspects that need further investigation are discussed. Potential applications of the tool presented in this thesis are suggested.

1.3 Originality

I declare that I am the sole author of this thesis document, and I produced all the tables and figures unless otherwise specified. Critique and comments were provided

by my supervisor Prof. J. Alison Noble. I am the corresponding author of all image pre-processing operations, machine learning algorithm implementations, and data visualisation works presented here. Parts of the research was conducted in collaboration with others are specified as follows.

Chapter 3 The first skull template was provided by collaborator Dr. Mohammad Yaqub. Independently, I built a collection of skull templates with varying sizes and orientations based on this template. I then implemented the skull detection algorithm based on template matching using Matlab. The plane extraction pipeline was jointly designed with collaborator Dr. Ana Namburete. I independently implemented this pipeline and tested it on the ultrasound image dataset I collected from the INTERGROWTH-21st database. I also proposed and validated the idea to manipulate the resolution to accelerate the runtime and improve accuracy. Meanwhile, I extended the proposed skull-stripping algorithm on processing 3D volumes and reported preliminary results.

For the plane detection work, I manually labelled the mid-plane of all images at different instances (one month apart) as ground truth and calculated intra-observer deviation based on it. Dr. Ana Namburete helped to annotate same dataset and the results were used to calculate inter-observer variations. I independently wrote and debugged the scripts for slice selection, midline identification and plane fitting functions, and tested the framework on a dataset consisting of healthy subjects. I analysed the experimental results by comparing the plane parameters collected from different groups: intra-observer, inter-observer and automated results. To further test the model performance, I manually labelled the corpus callosum (CC) on the automated planes and manually extracted planes based on a point distribution model I designed. I then carried out validation experiments to compare the two groups of extracted CC.

Chapter 4 This chapter used two independent datasets. I independently created and annotated the CC dataset. The choroid plexus dataset and its annotation were provided by Dr. Ana Namburete. Independently, I tested and chose the candidate

region selection algorithm for this work. I then designed the shape descriptors and implemented the feature extraction algorithm in MATLAB. I implemented the segmentation framework using three different machine learning algorithms: support vector machine, AdaBoost and RUSBoost in MATLAB R2015b. I also designed the validation experiments, analysed the importance of different feature sets and discussed effects of subjects' age on the model performance.

Chapter 5 I built the 3D image dataset with the help of collaborator Dr. Ana Namburete. For preprocessing, I independently implemented a CNN model for skull segmentation. The ground truth skull masks were provided by Dr. Ana Namburete. I trained the model independently and carried out the validation experiments to test the accuracy.

I manually labelled all the investigated brain structures (i.e. CSP, LV, CE, CM, Tha) that are visible in the corresponding volumes. I independently proposed the design of the view-based projection network (VP-Net) model. The first model was built with the help of collaborator Weidi Xie. I built and trained all six variants of VP-Net model and evaluate the model on the test set. Independently, I implemented the comparison baselines: random forest and 3D U-Net in Python 3.5. I also did a five-fold cross validation experiment using the most sophisticate VP-Net model F to exam its generalisation ability. In addition, I created a custom loss function based on dice-coefficient metric, and trained a VP-Net model using this customized loss function on my own. The 3D saliency volume was retrieved with the help of Weidi Xie. I manually measured the biometry: diameter of CE and the width of CM and compared them with the automated results.

Chapter 6 I reviewed the literature on establishing MRI brain atlases and independently proposed the fetal brain talarirach (FBT) coordinate system as the standard space for registering fetal neurosonograms. I manually labelled the reference landmarks for the dataset collect from the INTERGROWTH-21st and the INTERBIO-21st studies. I designed and implemented consecutive experiments to investigate the ability of a 3D CNN model to learn the affine transformation in

the context of ultrasound images. Built on top of the obtained results, I created a fully automated similarity registration framework for 3D fetal brain images, which takes a raw ultrasound volume as input and outputs a brain image that is transformed into the FBT space.

1.4 List of Publications

1.4.1 Journal Articles

Ruobing Huang, Weidi Xie, J Alison Noble. "VP-Nets : Efficient Automatic Localization of Key Brain Structures in 3D Fetal Neurosonography". Medical Image Analysis (under second review).

Ruobing Huang, Ana IL Namburete, J Alison Noble. "Learning to segment key clinical anatomical structures in fetal neurosonography informed by a region-based descriptor." Journal of Medical Imaging 5, no. 1 (2018): 014007.

1.4.2 Peer-Reviewed Conference Proceedings

Ruobing Huang, Ana IL Namburete, Mohammad Yaqub, J Alison Noble. "Automated Mid-sagittal Plane Selection for Corpus Callosum visualization in 3D Fetal Ultrasound images". Proc. of MIUA, 2015 – Oral presentation.

1.4.3 Conference Presentations

Ruobing Huang, Weidi Xie, J Alison Noble. "Using 3D CNN to Localize Key Brain Structures in 3D Fetal Neurosonography", in the BioEngenuity Keynotes, 2017, Oxford, UK – Podium presentation.

Ruobing Huang, Ana IL Namburete, J Alison Noble. "Automatic segmentation of brain structures in Fetal Neurosonography based on a Region-based Descriptor, " in the Imaging Cluster seminar, 2016, Oxford, UK – Oral presentation.

Ruobing Huang, J Alison Noble. "Data Analytics Tools for Early Detection of Fetal Alcohol Syndrome using Fetal Ultrasound Images", in the St.Peter' s College Graduate Seminar, 2016, Oxford, UK – Oral presentation.

Ruobing Huang, Ana IL Namburete, Mohammad Yaqub, J Alison Noble. "Automated skull detection for ultrasound fetal brain images, " in the Oxford Imaging Festival, October 2015, Oxford, UK –Podium presentation.

2

Review of fetal brain ultrasound image analysis

Contents

2.1 Skull stripping	10
2.1.1 Clinical Background	10
2.1.2 Related works	11
2.2 Mid-sagittal plane	13
2.2.1 Clinical background	13
2.2.2 Related works	14
2.3 Delineating brain structures in 2D images	15
2.3.1 Clinical background	15
2.3.2 Related works on ultrasound image analysis	19
2.4 Localizing multiple structures in 3D fetal neurosonog-	
raphy	21
2.4.1 Clinical background	21
2.4.2 Related works on ultrasound image analysis	24
2.5 Registering fetal brain in 3D neurosonography	27
2.6 Database	29
2.7 Conclusion	31

This chapter outlines the key challenges and progress made in analysing fetal brains in ultrasound images. The order of discussion is arranged accordingly with the contribution chapters presented in this thesis. In each topic, clinical findings or needs are first described. Then, an image analytic tool is proposed to address

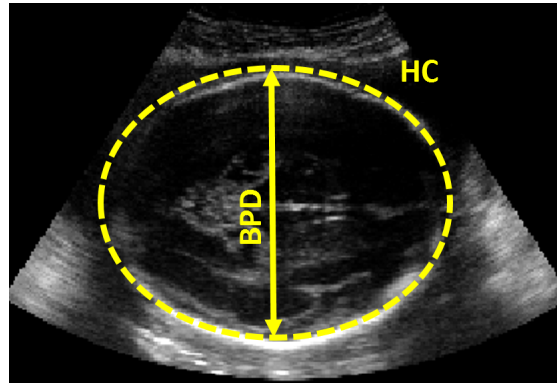


Figure 2.1: Typical fetal head measurement in axial slice. The skull appears as a bright ellipse. The dashed line depicts the head circumference(HC), while the two headed arrows depicts the biparietal diameter (BPD).

the problem and the strengths and weaknesses of established methods are reviewed. The review also briefly mentions works in a closely related field, i.e. neuro-image analysis technique for MRI data, to place the thesis contributions in context.

2.1 Skull stripping

2.1.1 Clinical Background

To help clinicians interpret ultrasound images and address the challenges discussed in Chapter 1, a substantial amount of research has been carried out. To analyse the fetal brain in ultrasound images, the first step is to detect the skull. The basic measurements of the skull, such as the biparietal diameter (BPD) and the head circumference (HC), shown in Figure 2.1, are important measurements for evaluating fetal growth, estimating the gestational age and diagnosing possible obstetric problems [8–11]. The BPD is the distance from the outer margin of the proximal skull to the inner margin of the distal skull. The HC is the circumference of the outer skull and is manually measured by tracing the skull or fitting an ellipse to the skull boundaries with digital callipers [12]. Successful detection of the skull helps to eliminate extra-cranial tissues and provide basic information about fetal health development. However, reports in the literature suggest that there are 1to7% intra- and inter-observer variations in head measurements due to the difficulty of interpreting ultrasound images [13, 14].

2.1.2 Related works

Several methods have been reported in the literature focused on detecting and segmenting the fetal skull. One of the most common early approaches employs morphological operations to enhance skull pixels followed by the Hough transform for ellipse detection. Zador et al. described a 2D method for detecting the skull using local thresholding and the Hough transform to find the centre of the head, for 74 of 75 images [14]. This method is relatively slow and requires manual selection of BPD and HC. Matsopoulos and Marshall [15] used morphological operations such as the grey-scale closing operation to connect the non-contiguous skull bones, and the fetal head area was filled by a iterative dilation algorithm. This approach does not guarantee an elliptical shape (no prior knowledge of shape). Therefore, might not generalize well to new data. Moreover, they only tested on one image. Hanna et al. [16] used thresholding to eliminate the small bright objects which do not belong to the skull, manually selected the rough head contour, and later used the Hough transform to achieve a more accurate elliptical contour. Chalana et al. [17] applied an active contour model to detect the inner and outer boundaries of the fetal skull. This method requires the user to specify a point near the true centre of a head.

Lu et al. [18] used K-means clustering, morphological operations and a randomized Hough transform to automatically detect the skull. This technique is fully automated and the experimental procedure is described as follows:

1. A K-means algorithm was used to classify each pixel according to its intensity value into three groups: background, grey object and bright object.
2. Morphological operations (i.e. morphological binary area opening operation, morphological dilation) were performed to eliminate small bright objects (< 20 pixels), smooth the boundaries and extract the 'skeleton' of the bright object.
3. The iterative randomized Hough transform was performed by randomly choosing a set of pixels and solving a parametrically linear model of an ellipse, $c = [x_0, y_0, a, b, \phi]$, such that (x_0, y_0) are the centre coordinates, a and b are the long and short semi-axes, and ϕ is the angle of rotation. If a solution of c is valid for the curve type of interest (ellipse for their study), a vote will be cast on c . After

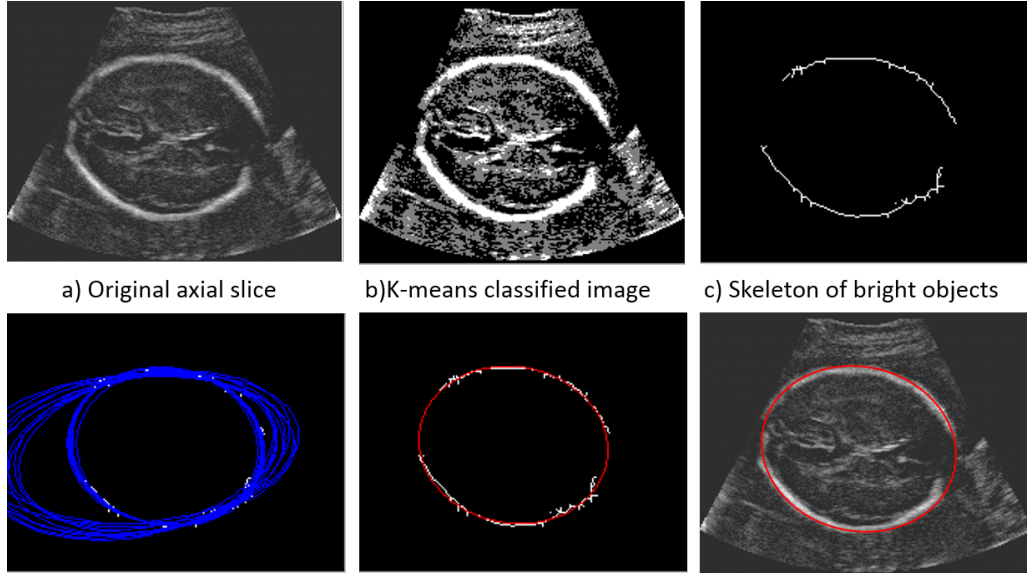


Figure 2.2: The skull detection using morphological operations and Hough transform based on [18]. (a) shows the 2D axial plane in ultrasound image. (b) shows the image been classified into three class. (c) shows the extracted skeleton of bright object after morphological operations. (d) shows the process of randomized Hough transform. (e) shows the final ellipse pick by the framework. (f) shows the ellipse superimposed on the original axial slice.

repeated for $K = 200$ iterations, the c with the maximum number of votes is chosen as the most likely representation of a of skull contour.

This technique was implemented to detect the skull on 6 images collected from fetuses at 22-23 gestational weeks. One example is shown in Figure 2.2. It can be seen that the K-means classifier roughly differentiates the skull from background and soft tissues by using the brightness of skull. The randomized Hough transform picks the ellipse that appears most frequently and fits the skull properly when superimposed on the skull. All the parameters for preprocessing and numbers of iterations were carefully tuned and chosen based on the testing image set.

However, when tested with images from a broader age range and more subjects from each gestational weeks (70 cases from 22 to 28 weeks old), the framework failed to detect the skull frequently. When carefully examined, their method tended to over-fit the testing image when choosing the parameters for pre-processing, which makes it difficult to generalize to new image data. In other words, given a random ultrasound image, it is possible to select the best parameters for picking out bright

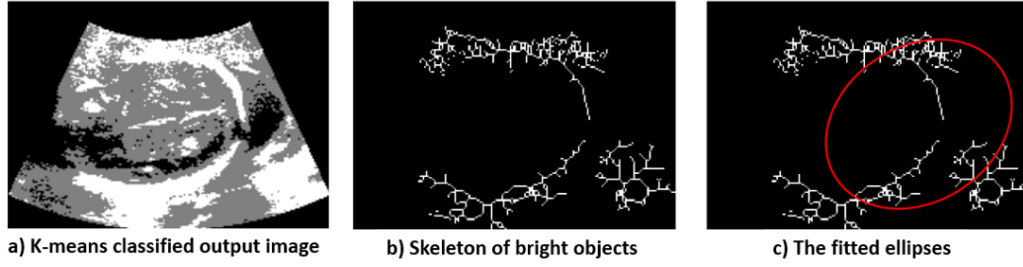


Figure 2.3: Example of false skull detection using morphological operations and Hough transform. (a) shows the K-means algorithm pick out skull as well as the maternal tissues. (b) shows the skeleton of bright objects which contains distracting information. (c) shows the final ellipse detected by the framework which tried to fit the false skeleton and did not fit the skull properly.

pixels, removing small bright objects and extracting the skeleton of bright pixel for this particular image. This selected set of parameters might be applicable on images with similar appearance but will fail on other images. Recall that ultrasound images have high variabilities in image quality and contrast across gestational ages and different subjects. Therefore, this approach has poor generalization ability. As illustrated in Figure 2.3, the preprocessing failed to exclude maternal tissues which contain huge amounts of bright pixels on a new image. Consequently, the ellipse chosen by the randomized Hough transform was "distracted" by the presence of bright extra-cranial pixels and failed to fit the skull properly.

In this case, to construct a fully automated framework, it is necessary to propose a new approach which has greater adaptability to challenging ultrasound images. It should make most use of the distinguishable features of skull in ultrasound images to reduce reliance on manual initialization or parameter selections. We look at this in Chapter 3.

2.2 Mid-sagittal plane

2.2.1 Clinical background

The brain of a healthy subject exhibits a rough bilateral symmetry with respect to the inter-hemispheric fissure that bisects the brain [19]. This fissure is commonly referred to as the anatomical mid-sagittal plane (MSP) (see Fig.2.4). The MSP is

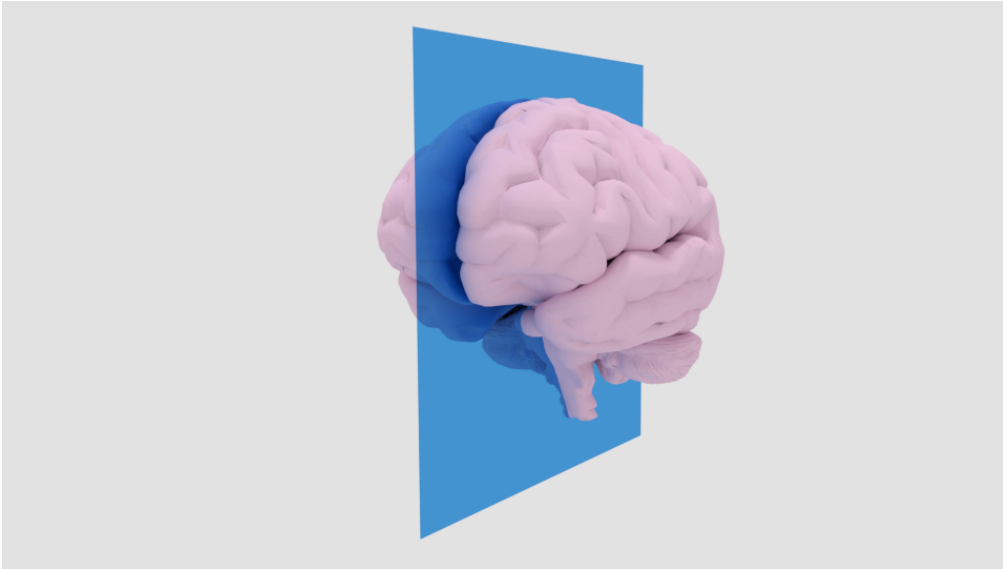


Figure 2.4: Schematic of the mid-sagittal plane of a brain in 3D. The plane is plotted in transparent blue.

essential for both clinical and technical analysis of brain images. The MSP can be helpful for detecting asymmetry of the brain, which in turn can assist diagnosis [20], i.e. fractures in CT images, lesions or bleedings in MR images, defects of perfusion in SPECT images. It also provides visualization of cerebral midline structures that are critical to the central nervous system [21]. The location of the plane can also be used to define the head orientation, usually the first step in normalization or registration of brain images, as aligning the MSP of brain images reduces the degrees of freedom of the transformation (number of unknown parameters) [22]. This point is investigated in Chapter 6.

2.2.2 Related works

Several papers in the medical imaging literature have previously considered the detection of MSP [23–27]. They can be divided into two categories: *symmetry-based* approaches and *feature-based* approaches.

The *symmetry-based* approaches take advantage of the natural similarity between two hemispheres and seek the partition that exhibits maximal symmetry. For example, Junck et al. developed a method for automatic detection of the line of symmetry in a transverse Positron-emission tomography (PET) image [23]. The

cross-correlation metric was used as a measure of similarity. The two parameters that specified the line of symmetry were found by an exhaustive search. Minoshima et al. presented a similar method which used a modified stochastic sign change criterion [24]. That work was also based on an exhaustive search of a small region of the parameter space for the optimal parameters. Ardekani et al. modified these approaches [23, 24] by incorporating different resolutions which reduce the time of parameter searching [26].

Feature-based methods focus on the intrinsic image appearance of the MSP itself. Brummer et al. proposed the extraction of edge maps of the MSP on 2D coronal Magnetic resonance imaging (MRI) scans [25]. Tuzikov et al. used snakes to extract fissure boundary and achieved plane detection by orthogonal regression [27].

Methods were developed for MRI [25, 27] and PET [23, 24] respectively and the literature is limited with respect to MSP detection in 3D fetal neurosonograms. Detecting the MSP in ultrasound images faces unique problems, e.g. varying image quality, large position and rotation variations and accompanied maternal tissues. Figure 2.5 shows a typical example of the MSP in a fetal neurosonogram.

Moreover, due to the signal refraction of skull bone, the two hemispheres of fetal brains have dramatically different visibilities. Figure 2.6) shows two examples of brains observed in ultrasound. Figure 2.7 shows the MRI and PET brain image. Comparing these figures, it can be seen that the appearance of brain does not exhibit the same symmetrical level of that witnessed in the MRI and PET modality. Therefore, the symmetry-based approaches are inapplicable in ultrasound brain images, while the feature-based methods might be more promising but need to be further tested. In sum, a specialised solution is needed to detect the MSP in fetal neurosonograms.

2.3 Delineating brain structures in 2D images

2.3.1 Clinical background

Current standard fetal neurosonography involves identification of several key brain structures and measurements of skull size [3]. These observations and metrics

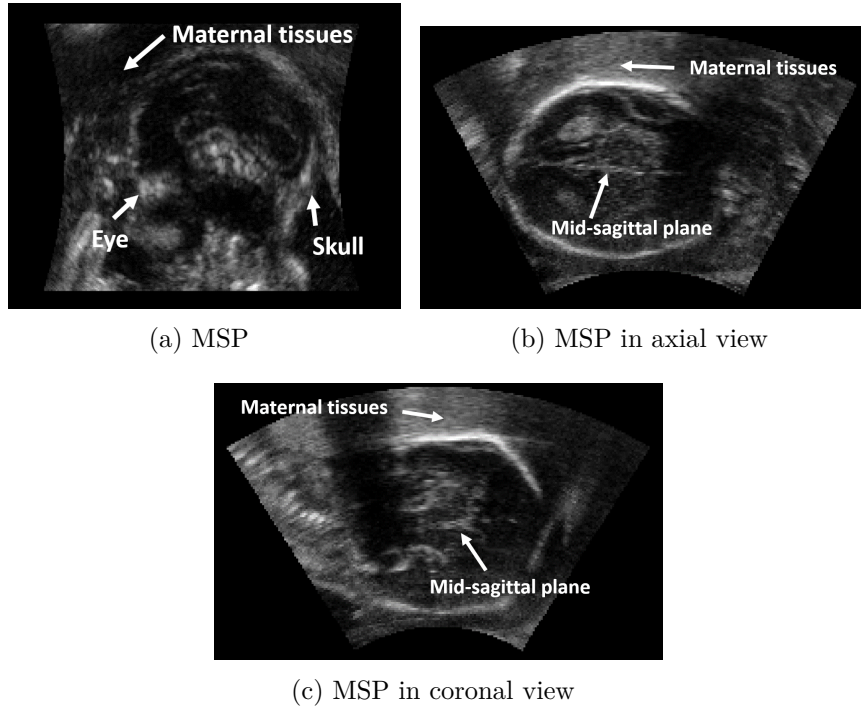


Figure 2.5: Fetal neurosonogram of a healthy fetus with gestation age of 21 weeks. (a) is a typical example of the mid-sagittal plane (MSP) observed in a ultrasound image. The corresponding location of skull and eye are pointed out. (b) and (c) are the MSP visualized in axial and coronal slice (as a bright thin line) respectively. Large portion of maternal tissue existed in all these views.

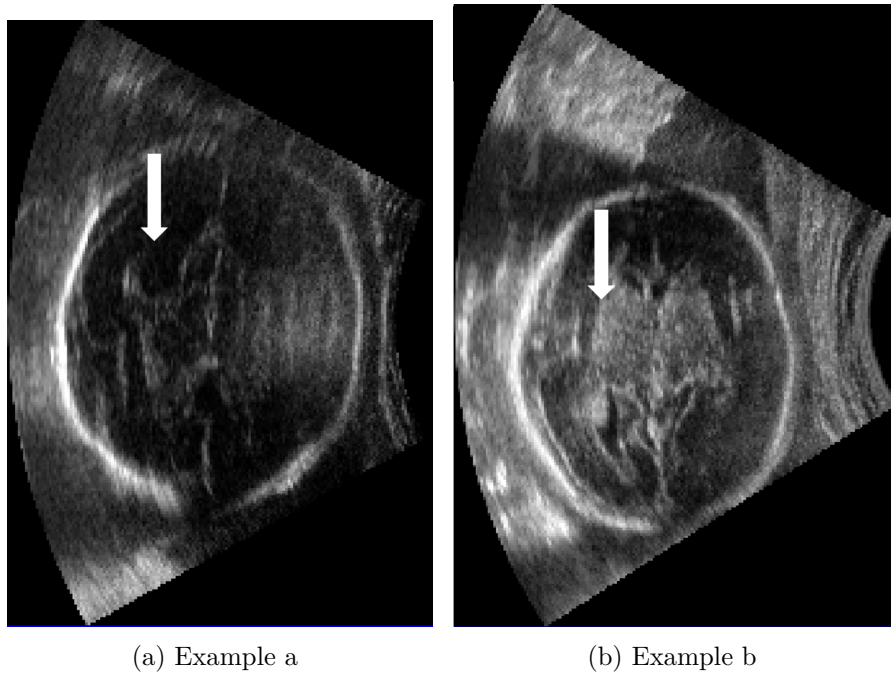


Figure 2.6: Two randomly selected examples of fetal brain ultrasound images. It can be seen that half of the brain has better visibility (pointed by white arrow).

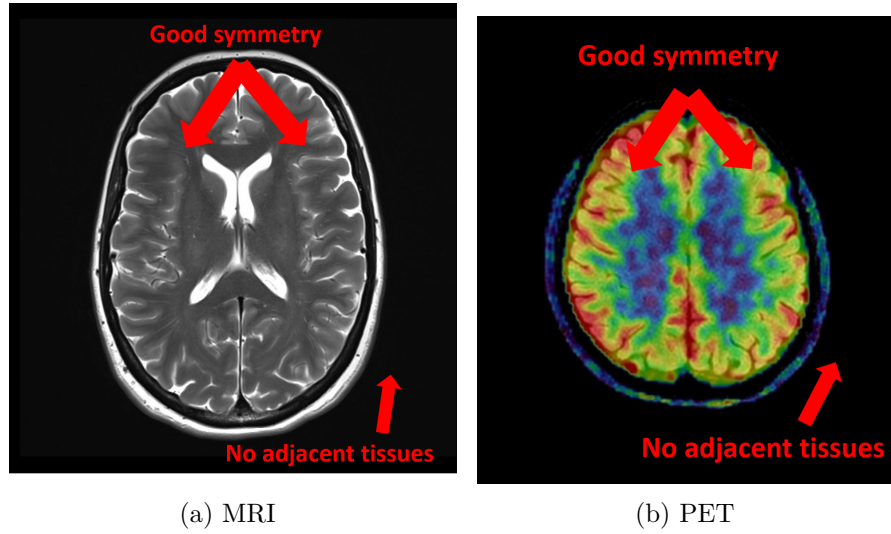


Figure 2.7: Examples of normal brain in MRI and PET images. Both the brains exhibit good symmetry and there is not adjacent tissues outside the skull. Image credit to <https://radiopaedia.org> for the MRI image, and <http://www.traumaticbraininjury.net> for the PET image.

offer a macroscopic evaluation of fetal brain health and can help to identify severe abnormalities e.g. complete absence of the corpus callosum [28, 29], microcephaly [30, 31], and macrocephaly [32]. However, the applicability of these gross evaluations is limited to conditions that exhibit subtle or mixed symptoms. For example, fetal alcohol spectrum disorder (FASD) is a constellation of birth defects resulting from maternal alcohol consumption during pregnancy [33, 34]. FASD can manifest in central nervous system damage which may result in lifelong intellectual disabilities [35]. A number of factors can influence the severity of disease, such as the amount of consumed alcohol, the period of effected gestation, and genetic background. As a result, FASD is not a binary concept but instead an umbrella term that covers a wide range of disorders, including fetal alcohol syndrome, partial fetal alcohol syndrome, and alcohol-related birth defect [36]. A patient with FASD can be identified with a cerebellum with reduced volume [37, 38] or a smaller or deformed corpus callosum [39, 40] and facial abnormalities as well [33]. Figure 2.8 displays various shades of abnormality of the corpus callosum (from normal to the most severe malformation) in MRI images. The intermediate abnormality cannot be detected by a simple identification of the structure (check whether this structure



Figure 2.8: Examples of MRI images of subjects with normal or abnormal corpus callosum (pointed by a white arrow). The images are listed in the order of the severity of the abnormality. The left-most image is a normal subject while the right-most has the most severe malformation: complete absence of corpus callosum. It can be seen there are various shades of abnormality. A simple identification of whether a structure exists might be inadequate to detect intermediate abnormality. Image credited to <https://www.nlm.nih.gov>, Elgamal et al. [43], and <https://emedicine.medscape.com/article/407730-overview>.

exists), but might be detectable by a measurement of its area. Furthermore, some clinical studies have related the biometry of other brain structures (e.g. the diameter of cisterna magna and the width of choroid plexus) to other congenital diseases [41, 42]. A more detailed evaluation of brain structures in ultrasound can reveal possible risks during pregnancy and allow early intervention.

Moreover, recording the biometry of brain structures in healthy subjects can benefit understanding of normal brain maturation and provide a useful reference for future diagnosis. In [44], Lee et al. suggested the diameter of the cerebellum is a useful indicator of fetal gestational age. Achiron et al. reported sonographic biometry of the fetal corpus callosum [45]. They found the maximum increase of its width occurs around 19 to 21 weeks' gestation while the length followed a linear growth. Serhatlioglu et al. did a similar study on other brain structures and pointed out the importance of collecting mean measurements from normal fetuses [46].

While these brain structures contain crucial developmental information, they can be difficult to identify and their measurements will increase the cost of time and labour. This is particularly true in late gestation when skull bone begins to calcify and reduces the visibility of structures. Segmentation of brain structures can generate automated measurements and provide shape information at the same time. An automatic algorithm dedicated to do this would benefit both clinical

practice and scientific research in evaluating fetal brains. We consider a method to achieve this in Chapter 4.

2.3.2 Related works on ultrasound image analysis

Prior work on ultrasound image analysis has primarily considered applying machine learning to automatic *detection* of fetal brain structures [47–49], rather than *segmentation*.

Detection aims to identify a structure, while segmentation provides additional information about a structure’s shape and appearance. Skull segmentation in fetal ultrasound has attracted most attention as a precursor to head biometry. Early work used morphological operations and the Hough transform [14, 15, 18]. Such approaches are sensitive to the quality of image acquisition and none of the early methods were validated on large datasets. Furthermore, segmentation of a brain structure can be more intractable, as the skull bone has the following characteristics:

- Its size is larger, i.e. a 25 gestational-week-old fetus has a head size around 230mm (HC) [50], while the diameter of its cerebellum is typically less than 30 mm [51];
- Strong echogenicity. The skull bones are highly hyperechoic (acoustic impedance = $7.8kg/(sec \cdot m^2)$) and thus appear bright on ultrasound, while brain structures are formed by less echogenic tissues and therefore are less distinct from the background (acoustic impedance = $1.6kg/(sec \cdot m^2)$) [52];
- Regular shape. The skull can be closely approximated mathematically as an ellipse. By contrast, intra-cranial structures usually have irregular and diverse shapes which make it laborious to design a geometric model for each structure.

Consequently, the established methods for fetal skull segmentation may not be applicable to normal brain structures. Novel approaches are required which

are able to handle problems such as a low foreground-to-background ratio and varying shapes of targets.

To the best of our knowledge there are only two prior publications on ultrasound *segmentation* of fetal brain structures. Yaqub et al. reported a semi-automatic approach to segment fetal brain structures in 3D ultrasound where they restricted the search region to the smallest cuboid that enclosed the structure [53]. Cuboid extraction was based on prior knowledge of the approximate location and the size of the structure of interest. Gutierrez-Becker et al. presented an algorithm to segment the fetal brain cerebellum using a statistical shape model [54]. While they reported promising results, validation was limited to only 20 subjects and it is unclear how that method could be readily extended to other brain structures. Furthermore, all images were acquired with a relatively strict protocol leading to the cerebellum appearing in a similar initial position for all the images. This suggests that the solution is not well-suited for general clinical imaging practice where the initial position of brain anatomy is uncertain.

Methods for segmenting ovarian follicles and fetal heart chambers have also been reported [55, 56]. Here the anatomical configurations are different from the intra-cranial anatomy setting, where the presence of skull bone affects the visibility of other tissues. Furthermore, fetal brain structures can have large variations in terms of shape, texture and size due to gestational age (GA), which further complicates the problem. Hence, it is hard to generalize these existing methods to segment brain structures.

In MRI studies, approaches such as statistical atlas models [57, 58], and patch-based techniques [59, 60] have been proposed for fetal/adult brain segmentation. In particular, Habas et al. presented a spatiotemporal atlas of fetal brain created using MRI images collected from fetuses with a gestational age range from 20 to 24 weeks [57]. The main advantage of this method was that it can correctly capture growth-related change, as subjects were grouped into different age cohorts to create age-specific atlases [57]. Gholipour et al. used a multi-atlas approach that fuses labels from age-specific atlases and reduce segmentation errors [58]. Tu et al.

trained a classifier on local image patches and used the resulting confidence maps to train a second classifier for extraction of high-level information [59]. Coupe et al. achieved the labelling by finding the most similar patch among training set [61]. Eskildsen et al. adapted this method by incorporating images of different resolutions which resulted in better accuracy [60]. In [62], Wright et al. analysed the pros and cons of an atlas-based expectation-maximization framework and path-based segmentation. Their results showed that the atlas based framework performed superiorly in segmenting tissues with strong intensity boundaries while the latter outperformed the former method when classifying pixels with similar intensities.

Prior works in MRI shed light on the study of segmentation in ultrasound images. Nevertheless, it is important to note that the appearance of structures in MRI images is quite different from the ultrasound case. Fetal sonography does not lend itself well to anatomical atlas-based techniques as anatomical structures have varied appearance depending on object tissue properties / echogeneity, and the pose of the object with respect to the transducer (fetal motion makes this a practical challenge to control). Appearance of anatomical boundaries is typically incomplete and strong acoustic shadows can obscure structures. A patch-based method assumes that the appearance or the intensity level of certain brain regions is consistent across subjects. This assumption does not hold true in ultrasound. Figure 2.9 displays the trans-ventricular plane of two randomly selected fetal ultrasound scans. The image patches in the same location of the lateral ventricle (circled by red box) have different appearance while the brain tissue can exhibit similar visual feature with extra-cranial tissue. This suggests segmenting brain structures in fetal neurosonography might require a ultrasound-specific approach to address relevant issues.

2.4 Localizing multiple structures in 3D fetal neurosonography

2.4.1 Clinical background

Rather than investigating individual structures as previous section, this section considers localizing multiple structures in 3D brain images. Chapter 1 listed the

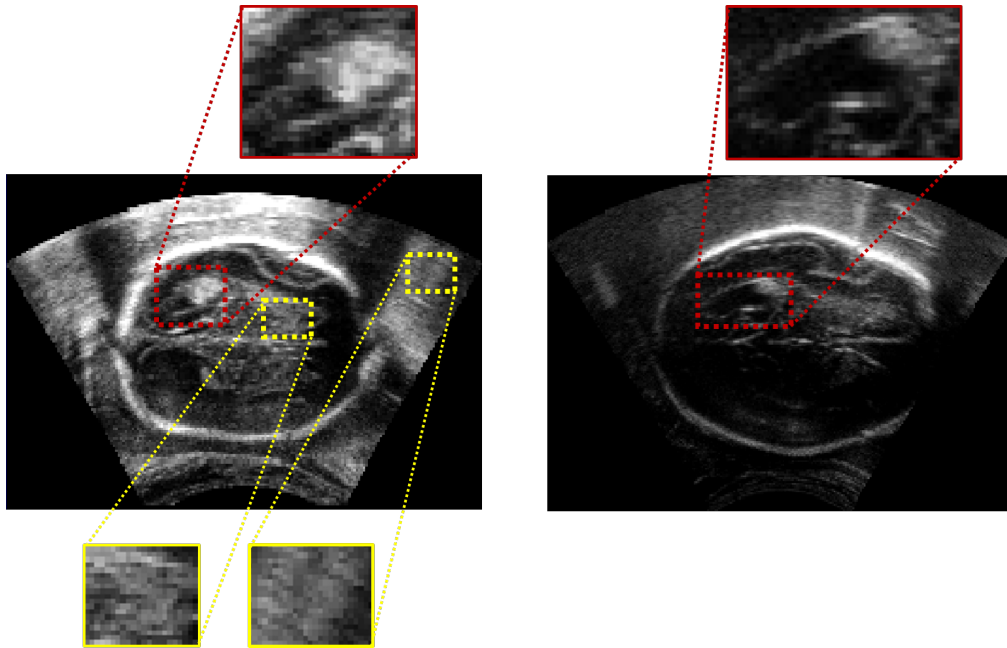


Figure 2.9: Trans-ventricular plane of two randomly selected fetal ultrasound scans. The red dashed boxes highlight the anterior horn of lateral ventricle for each subject. Compare the two red boxes, it can be seen the appearance of the image patch differs dramatically in the same location of the same structure across different subjects. The yellow dashed boxes in the image of subject label a patch of brain tissues and a patch of maternal tissue outside the brain. The texture and intensity level of the two patches are visually similar.

key brain anatomies included in the standard examination, i.e. the lateral ventricles (LV), cavum septi pellucidi (CSP), thalami (Tha), cerebellum (CE), and cisterna magna (CM). Figure 2.10 offers a detail schematic of these structure and their visualization in ultrasound. In practice, localizing these structures is typically achieved by first manipulating the probe to obtain a suitable 2D plane and then identifying the structure of interest. Navigating the brain in 3D space is non-trivial due to the presence of speckle, skull calcification, varying fetal brain positions, sizes and orientations within volumes. To correlate a real structure with the anatomical pattern seen on the 2D screen of an ultrasound machine, a sonographer/clinician must use spatial understanding and a good grasp of anatomical knowledge to mentally reconstruct these structures in 3D.

For example, the CM is a vital indicator of trisomy 18 and other chromosomal abnormalities [63]. It is located between the CB and the dorsal surface of medulla

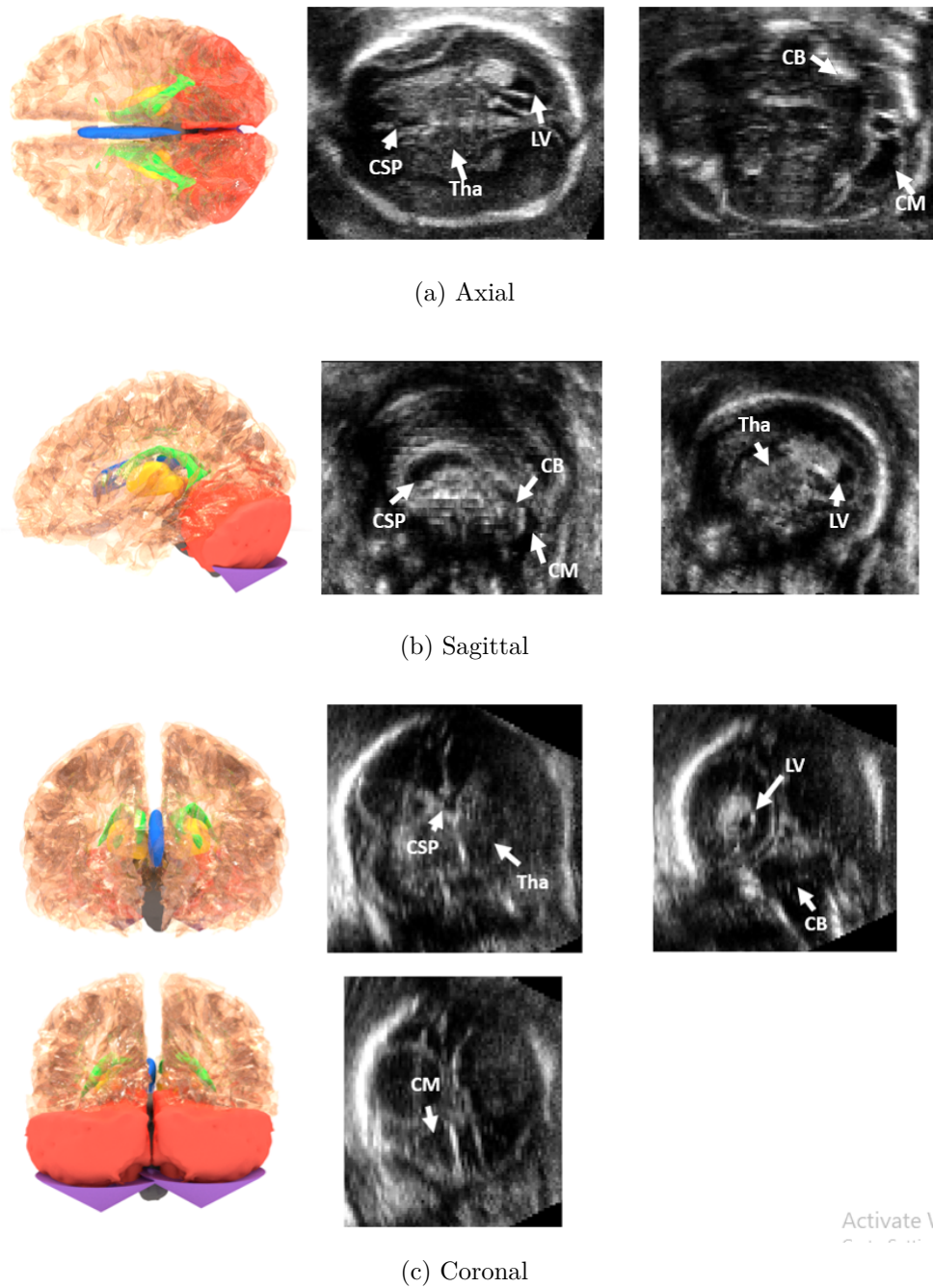


Figure 2.10: Demonstration of key brain structures. The three sub-figures represent different anatomical views: axial, sagittal and coronal. The first column is the schematic of these structures: CSP (blue), LV (green), Tha (yellow), CB (red) and CM (purple). Two schematic of the coronal view are included for better visualization. The rest of columns are these structures observed in ultrasound images. Their locations are pointed out by white arrows.

oblongata (purple in Fig.2.10). In ultrasound scans, it is visualized as a ‘M’-shaped dark blob in axial view (top-right) and a pair of symmetrical triangles in coronal view (bottom-right). It might be difficult to relate these two dissimilar patterns to the same anatomical structures without certain level of experience and expertise. With the increasing availability of 3D ultrasound machines, it is possible to develop an image analysis tool to automate and accelerate this process. With the assistance of such a tool, sonographers only need to sweep the probe to collect a 3D fetal brain scan and the algorithm would automatically localize the target structures for further examination. The 3D localization of a target can also inform its 3D location and its size. This biometry may assist the study of dislocation and volume deviation of cerebral structures for a large population of fetuses in the future.

2.4.2 Related works on ultrasound image analysis

Before reviewing specific articles, the definition of some terminology is first discussed. Object detection, localization and segmentation are used frequently in medical image processing and in the broader field of computer vision. While segmentation generally involves assigning class labels to individual pixels, both detection and localization require putting bounding boxes around a target. These two terms are used interchangeably within some literature and the difference between them is relatively ill-defined. Here, these two terms are differentiated based on the meaning of their root word. Detection is defined as discovering the existence of an object. It focuses on determining whether an object appears in an image and cares less about its location. Localization is defined as restricting an object to a particular place, which prioritizes finding the exact location and size of a target. In this work, the main goal is localization as it emphasizes the spatial location and the size of an anatomy within a image.

Classical machine learning approaches

A rich literature can be found on localization in medical image analysis. The subject of discussion here is restrained to 3D images and 2D ultrasound images.

Object localization is essentially a high-level abstraction task which maps a whole 3D volume to several digits. Researchers have designed specialized features and complex processing systems for this task [49, 64–68]. Some works learn the mapping from feature representations to raw coordinates of the targeted bounding boxes [49, 64, 69]. For instance, Criminisi et al use a multi-class random regression forest to localize structures in 3D CT scans [64]. Based on the intuition that distinct voxel clusters outside an organ may predict the position of the target with high confidence (context information), they extract pairs of sub-volumes at random sampled locations. The intensity difference between each pair is calculated as one feature. They then trained a regression forest to select the most informative features and generate bounding box coordinates. Carneiro et al. presented a detection method using a probabilistic model, whereby the posterior classifiers were trained by probabilistic boosting trees [47]. They used Haar features to represent the appearance of an image region, which was originally proposed for face detection [70]. In subsequent work of [47], Sofka and colleagues used similar features to characterize a single structure while incorporating the spatial relationships between multiple targets through sequential sampling [49]. The posterior and transition distributions of each target and their connections are derived as simple Gaussian distributions. In [48], Namburete et al. argued that the Nakagami distribution is a more suitable distribution for ultrasound images as it is able to distinguish between scatterers of different concentrations within the tissue [71]. They then describe a method to detect the presence of the choroid plexus in 2D ultrasound using AdaBoost based on the Nakagami distribution. These methods relied on the quality of hand-crafted features or specially selected probability distributions.

Deep learning approaches

Most of the state-of-the-art localization algorithms in computer vision involves deep learning technology [72–74]. Deep learning is a class of machine learning algorithms that has gained enormous popularity recently [75, 76]. It uses cascading multiple layers of non-linear processing units for feature extraction and transformation.

These layers are called neural networks, while the non-linear processing units are called neurons as they are loosely based on the interpretation of a biological nervous system. A simple neural network consists of an input, an output and a hidden layer (see Fig. 2.11). Each layer has a dozen of neurons which are connected by network parameters, sometimes are called weights.

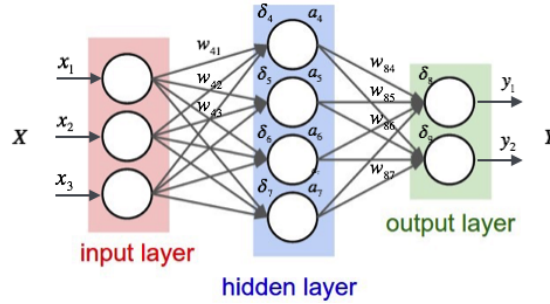


Figure 2.11: A simple example of neural networks. It consists of three layers: input (red), hidden (blue) and output (green). Neurons are shown as circles and their connection (weights) are plotted as arrows. Image source: <https://goo.gl/images/1RXTyH>.

Convolutional neural network (CNN) is one type of deep, feed-forward neural networks that has been widely used in computer vision [77–79]. In contrast to fully connected networks, the connections between hidden layers in a CNN do not have independent weights. Instead, weights are shared across the same layer. This design reduces the number of parameters and allows the networks to grow deeper, which helps to boost model performance [78]. Typically, each convolutional layer is followed by a pooling layer to increase receptive field and to obtain global features. The convolutional layer with the largest receptive field are usually flattened and connected with a fully-connected layer to output the final prediction (see Fig. 2.12).

In the field of medical image analysis, efforts on 3D localization using CNNs mainly focused on solving the problem in a lower dimension, i.e. individual 2D slices [65–68]. Zhou et al. presented a method for 3D CT images. It first detects the 2D partial appearances of a 3D organ from multiple directions using ensemble learning. The 2D results are subsequently combined through a majority voting mechanism [65]. In [66], Lu et al. used a CNN to classify an image slice and a fully convolutional network to classify local patches. Similarly, the score maps

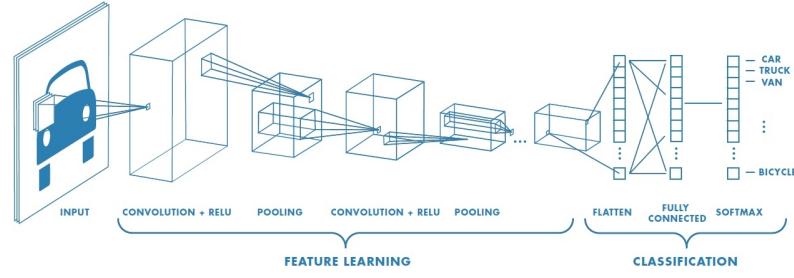


Figure 2.12: An example of Convolutional neural network (CNN). An input is followed by interwoven convolutional and pooling layers. The final convolutional layer is flattened and passed to a fully-connected layer to generate output predictions. Image source: <https://goo.gl/images/RaUtB9>.

collected from different slices/orientations are fused by majority voting or sum rule. De Vos et al. treated each slice of a 3D CT image separately and used a single CNN to find multiple targets in three anatomical views [67]. They also tested several different CNN architectures. The voxels that have been classified as positive in all three views are selected. The largest 3D connected component of them is retained as the output bounding box. These methods are simple to implement and have fast run-time due to dimension reduction. However, these methods all strongly depend on correct detection in 2D as they treat each slice separately in the first organ detection module. This may not be robust near the boundary of targets where only part of the organ is visible.

In summary, some methods have been established for structure localization in 3D medical images, but there is room for improvement. A data-driven technology (e.g. CNN) might be useful to reduce the reliance on hand design features. We investigate this problem in Chapter 5.

2.5 Registering fetal brain in 3D neurosonography

Image registration is the process of aligning two or more images from different subjects, or taken at different times or by different sensors. It is widely used in computer vision and medical imaging to combine various data sources for image fusion, change detection and image restoration [80]. The majority of registration

methods consists of four basic steps: feature detection, feature matching, transform model estimation and image transformation [81]. They can be explained as follows: (1). Feature detection. Features usually consist of distinctive objects (e.g. edges, contours, corners and closed-boundary regions) that are easily detectable [82–85]. This step requires selecting discriminated information that is present in all conditions considered by a specific task. Ideally, the extracted feature should be robust to situations such as image degradation, object occlusion and lighting condition variations. (2). Feature matching. This step establishes the correspondence between the non-aligned image and the reference image based on the detected features. Various feature descriptors and similarity measures have been proposed to set up accurate accordance [86–88]. (3). Transform model estimation. Different transformation models, including rigid, affine, projective, and deformable models, are selected for each task based on the application and prior knowledge [89–91]. These models are also known as mapping functions, whose parameters are computed based on feature matching results. (4). Image transformation. This step is actual process of employing the estimated mapping function on the non-align images based on selected interpolation method.

A key challenge in analysis of fetal brain is rapid changes in brain morphology and morphometry occurring over very short time-scales. Habas et al. used group-wise registration of manual segmentations and voxel-wise nonlinear modelling to create spatiotemporal fetal brain atlas [92]. They excluded temporally inconsistent tissue boundaries and rely on basic tissue or structures such as gray matter, ventricles, skull, etc. Rousseau et al. propose an iterative slice-to-volume registration process to reconstruct a single 3D volume from a stack of acquired 2D MR slice [93]. They first globally aligned the slices using rigid transformation the corrected slice motion by registering a slice to the reconstructed volume. They used normalized mutual information [94] as similarity measure. Murgasova et al. use intensity matching and robust statistics to register 2D slice to a fetal MR volume [95]. In specific, they classified voxels and slices into inliers and outliers using the

expectation–maximization (EM) algorithm to remove artifacts caused by motion corruption and misregistration.

Most literature on fetal brain registration focuses on MR images or volumes. To the best of my knowledge, no previous work has been done to align fetal brain in neurosonography. One key factor need to be considered is that the fetal brain can appear in reverse position in a 3D ultrasound volume (180° orientation deviation in axial and sagittal view). This particular challenge differs from most applications as the target and the reference usually have similar orientations (e.g. adult brain in MRI, kidney in PET). Classical registration methods, such as correlation-like methods, might fail in this scenario as the difference between the target and the reference is too. Therefore, it might be more practical to align fetal brains to a standard coordinate system as pre-processing before non-rigid registration. A linear transformation can be decomposed into several parameters each represents translation, rotation, scaling etc. It can thus be translated as a regression problem that can take advantage of the power CNN in efficiently extracting features from data. The state-of-the-art CNN-based registration algorithms usually trained on simulated deformations [96] as supervision or use certain similarity measure as the network loss function [97, 98]. Chapter 6 borrows the idea of using simulated deformation to train CNNs and explores, for the first time, registering fetal brain in 3D neurosonography.

2.6 Database

All data used in this thesis was sampled from the database built by the INTERGROWTH-21st project [99]. The International Fetal and Newborn Growth project for the 21st Century (INTERGROWTH-st) was a multicentre, multi-ethnic, population-based study. Its establishment involved more than 300 researchers and clinicians from 27 institutions in 18 countries worldwide, coordinated from the University of Oxford (see Fig.2.13). Its primary aim was to study growth, health, nutrition and neuro-development from 14 weeks gestation to 2 years of age, to produce international prescriptive standards for pregnancy dating, maternal weight gain, fetal growth, newborn size, the postnatal growth of preterm infants and cognitive

development at 2 years of age. Recruitment for the study took place between April 27, 2009 and March 2, 2014 in eight study sites: the cities of Pelotas (Brazil), Turin (Italy), Muscat (Oman), Oxford (UK), Seattle (USA), Shunyi County, Beijing (China), the central area of Nagpur (India) and the Parklands suburb of Nairobi (Kenya). Women were recruited based on a strict standard involving age, height, weight, medical history etc. to ensure healthy pregnancy. The recruited women were monitored from early pregnancy onwards to capture very detailed information about fetal growth patterns. Subjects were excluded if certain risk factors were identified e.g. intrauterine growth restriction (IUGR), alcohol use, or smoking during gestation. Multiple pregnancies were also excluded from the study. At study entry, a first-trimester scan was collected for the referential crown-rump length (CRL) measurement, along with the last menstrual period (LMP, if known) for age estimation. The 2D and 3D US image data were obtained using a Philips HD9 linear probe at a transducer frequency of 2-5MHz from over 4000 healthy pregnant women that underwent up to six scanning sessions throughout pregnancy [100].

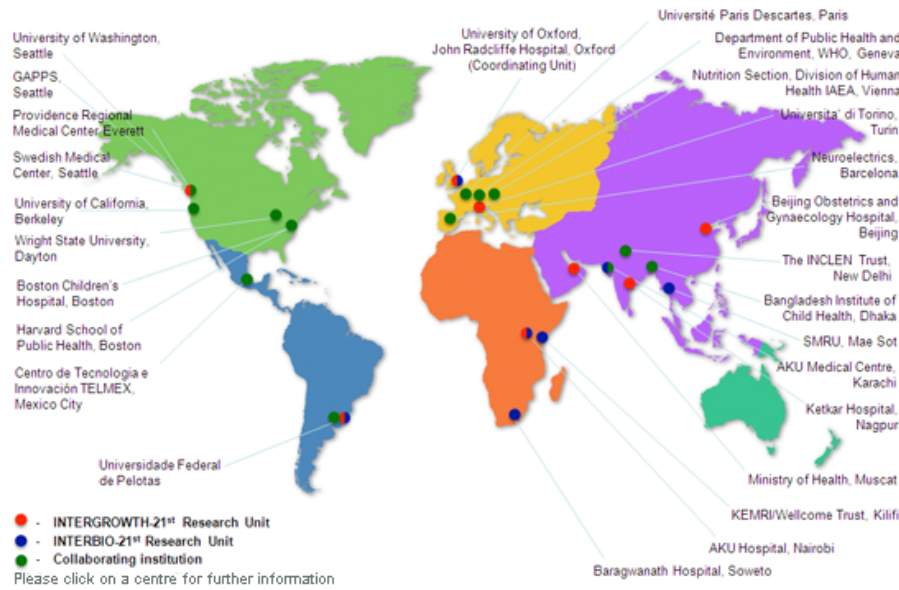


Figure 2.13: Map of INTERGROWTH-21st research units and collaborating Institution. Research unit are labelled in red while collaborating centre are labelled in green.

2.7 Conclusion

This chapter discussed several limitations of the current standard fetal neurosonography examination and discussed how image analysis tools can help to solve these problems or simplify the process. Section 2.1 focused on the fetal skull. It also reviewed literature about skull detection in ultrasound images. Section 2.2 described the importance of the mid-sagittal plane, which provides visualization of midline cerebral structures. As this plane is not included in a routine scan. I proposed to build a plane extraction tool and discussed several plane extraction methods for MRI or PET images. Section 2.3 showed how the segmentation of brain structures could benefit prenatal diagnosis and understanding brain maturation. It raised the need to develop segmentation tools for ultrasound scans and reviewed relevant methods in the literature on this problem. Section 2.4 analysed the obstacles in localizing structures in 3D fetal brain images. It then explored the merits and limitation of the state-of-the-art automated localization algorithms. Finally, the chapter provided details about the database used in this thesis.

3

Visualization in fetal neurosonograms: automated skull-stripping and plane detection

Contents

3.1 Skull stripping in ultrasound images	34
3.1.1 Proposed method	34
3.2 Mid-sagittal plane detection	37
3.3 Experiments	41
3.3.1 Image dataset	41
3.3.2 Skull detection accuracy	41
3.3.3 Manual plane selection	42
3.3.4 Orientation deviation	43
3.3.5 Shape Analysis	43
3.4 Results and discussions	44
3.5 Conclusion	46

This chapter presents a plane detection pipeline for visualizing the mid-sagittal profile for 3D fetal neurosonograms. The most challenging part, and the core of this pipeline, is the skull stripping algorithm. The removal of extra-cranial background frees the viewer or subsequent processing from the distraction of maternal tissues. It serves as a preliminary and fundamental step in analysing ultrasound images of the fetal brain. This chapter offers a detailed description of the original proposed

skull stripping algorithm and its improved version. Validation experiments and their results are presented in this chapter and discussed. Early version of the work presented here have been published in:

Ruobing Huang, Ana IL Namburete, Mohammad Yaqub, J Alison Noble. Automated Mid-sagittal Plane Selection for Corpus Callosum visualization in 3D Fetal Ultrasound images. Proc. of MIUA, 2015 – Oral presentation.

3.1 Skull stripping in ultrasound images

During standard acquisition of 3D ultrasound fetal head images, the ultrasound probe is positioned perpendicular to the mid-sagittal plane while allowing for a small angle deviation towards the midline. As a result, on the axial slice of brain images, the skulls have varying rotation towards the image axis. Meanwhile, due to the natural biological differences between fetuses and gestational age, skull sizes also vary (shown in Figure 3.1). Despite the difference in size and rotation angle, it can be seen that the skull is bright and its shape approximates an ellipse in every case. However, the proximal skull segment (the part closest to the ultrasound probe) often exhibits a different brightness the distal one (shown in Figure 3.2). There can also be other bright tissues adjacent to the head skull and moderately large gaps exist between skull segments (Fig.3.2). Therefore, designing a tool to differentiate the skull from background in ultrasound images that could function well over a broad age range is complex and requires deliberate design.

3.1.1 Proposed method

In summary, the skull is bright and its shape approximates an ellipse despite the gaps and different brightness between skull segments. The skull can be considered as an elliptic bright object in the 2D axial view of a 3D US image. Its contour can be detected through template matching with a binary image containing an ellipse contour. Template matching is essentially sliding a template around the image looking for a location where the values of the template and that of the image are best

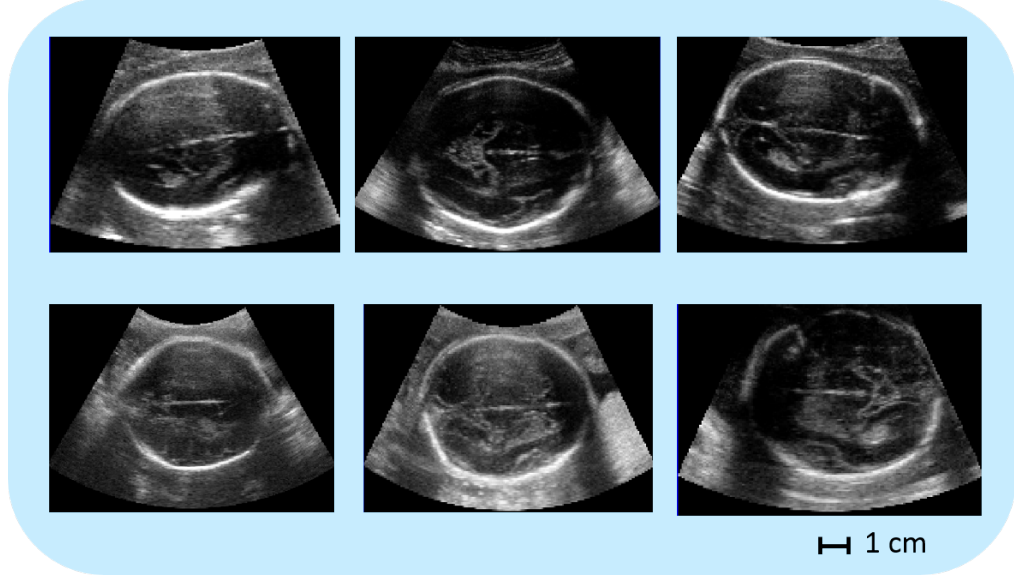


Figure 3.1: Examples of ultrasound skull images. Subjects exhibit different contrast, skull shapes and orientations.

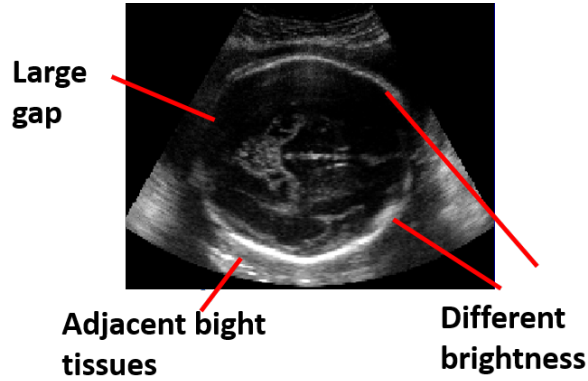


Figure 3.2: Challenging conditions present in identifying skull in fetal ultrasound images. The red lines point out large gap between skull segments, bright tissues that is closely adjacent of skull and different brightness of skull segments.

aligned. A simple way to achieve this is to measure the sum of square differences between the template and the image region where they overlap. It can be written as:

$$\begin{aligned}
 d_{f,t}^2(u, v) &= \sum_{x,y} (f(x, y) - t(x - u, y - v))^2 \\
 &= \sum_{x,y} (f^2(x, y) + t^2(x - u, y - v) - 2f(x, y)t(x - u, y - v)) \quad (3.1)
 \end{aligned}$$

where t denotes the template, f denotes the image, the sum is over x, y under the window containing the feature t in u, v . As shown, the first and the second parts are constant. The third part is the negative of the cross-correlation value

between f and t times two. It can be seen that the Euclidean distance between the template and the image increases with the correlation. This provides some intuition to use cross-correlation in template matching. However, one weakness of calculating correlation in this way is that its value can be high in locations where image intensity $f(x, y)$ is high but do not necessarily match well with the template [101]. Therefore, the normalized cross-correlations (NCC) is introduced as:

$$N(u, v) = \frac{\sum_{x,y} [f(x, y) - \bar{f}_{u,v}] [t(x - u, y - v) - \bar{t}]}{\{\sum_{x,y} [f(x, y) - \bar{f}_{u,v}]^2 \sum_{x,y} [t(x - u, y - v) - \bar{t}]^2\}^{0.5}} \quad (3.2)$$

, where $\bar{t}$ is the mean of the template and $\bar{f}_{u,v}$ is the mean of f in the region under the template. NCC measures the similarity of two series as a function of displacement, but it is invariant to the scaling of the template intensity or the that of the computed image region. This feature is particularly useful in ultrasound image processing as the subjects are commonly observed with inconsistent intensity levels.

Therefore, skull detection is achieved by calculating the NCC between the original image and ellipse templates of 5 different sizes (major axis 85 ± 15 mm, skull thickness 2.9 ± 0.6 mm) and 6 different orientations (ranging $\pm 30^\circ$). The template corresponding to the maximum cross-correlation value is then selected to create a mask image to remove the extra-cranial and maternal tissues (Fig.3.3).

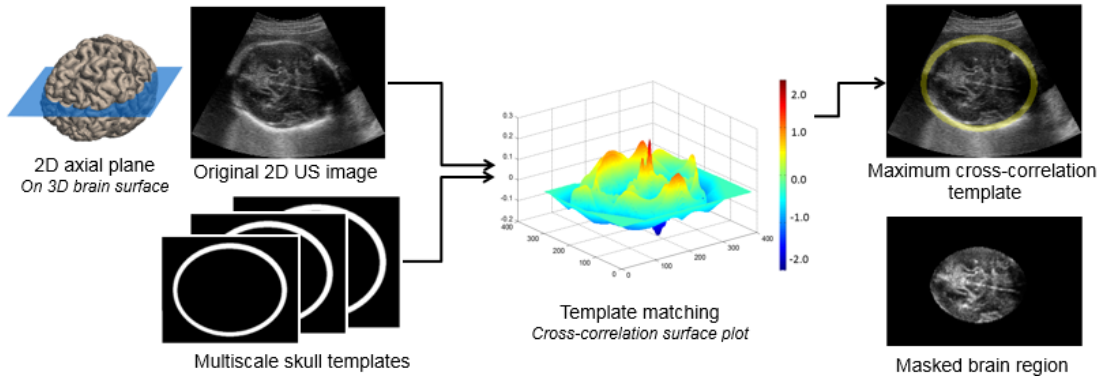


Figure 3.3: Skull detection process. The skull in each axial slice is detected through templates matching and the extracranial tissues are removed using the skull contour.

Resolution Reduction It was found empirically that, as the skull is the largest and the most distinct object in the ultrasound image, one can identify the skull contour in images with lower resolution. In other words, one can reduce the resolution of the target ultrasound images and the templates simultaneously and achieve similar accuracy in detecting the skull. It can reduce the computation time considerably as the NCC is calculated in a sliding-window manner. The tool is targeted at subjects from a broad age range (22 to 28 gestational weeks), whose skull sizes have large biomedical variations. Normalizing the image size can reduce the number of templates the algorithm requires to achieve a good accuracy. Moreover, the skull stripping algorithm is performed on a stack of axial slices in the later introduced plane detection pipeline. The speed of it has a great impact on the overall speed of the pipeline. In practice, the y-axis of a 2D ultrasound image (typical size = 210 pixels) is resized while its x-axis is reduced accordingly to keep the same aspect ratio. This was found to reduce the computation time, on average, from 3.2s to 0.6s to process per image.

Skull detection in 3D The skull stripping algorithm discussed above focused on detecting the skull in 2D ultrasound images of the fetal brain. It is natural to leverage this algorithm for skull detection in 3D. To do this, one can apply it on a stack of axial slices sampled from a 3D ultrasound image and use the detected skull in each slice to fit a 3D ellipsoid that encloses the skull. Figure 3.4 gives one example of 3D skull detection. The ellipsoid is fitted using a least square algorithm [102].

3.2 Mid-sagittal plane detection

The mid-sagittal plane splits the human brain into two hemispheres which usually exhibit a rough bilateral symmetry. The detection of the mid-sagittal plane is often the first step in spatial normalization of brain images [103].

To identify the mid-sagittal plane, 2D horizontal slices of the fetal head are automatically selected and the skull boundaries are detected to isolate the brain region. In each of the slices, the midline is automatically detected to generate a

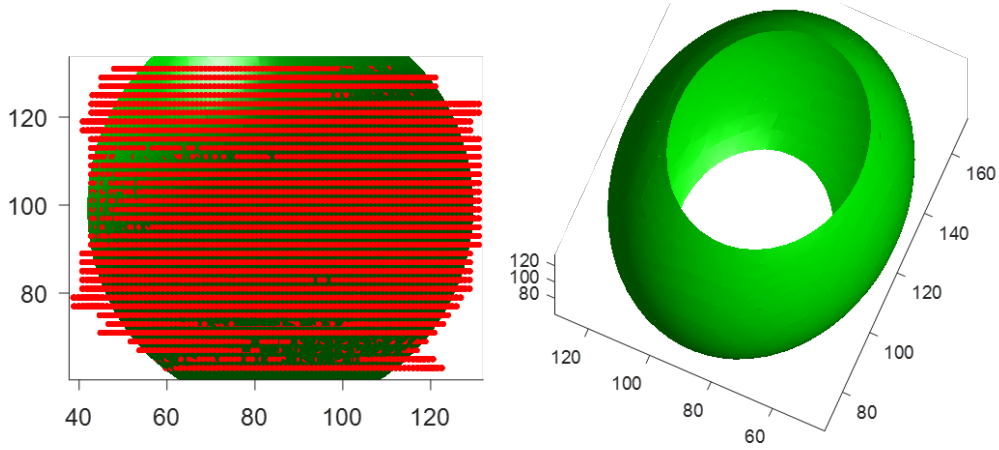


Figure 3.4: Schematic of the detection of the skull in 3D. The detected skull in each axial slice is plotted in red. The fitted ellipsoid is plotted in green.

set of lines which are later used to estimate the mid-sagittal plane. This procedure is summarized in Figure 3.5.

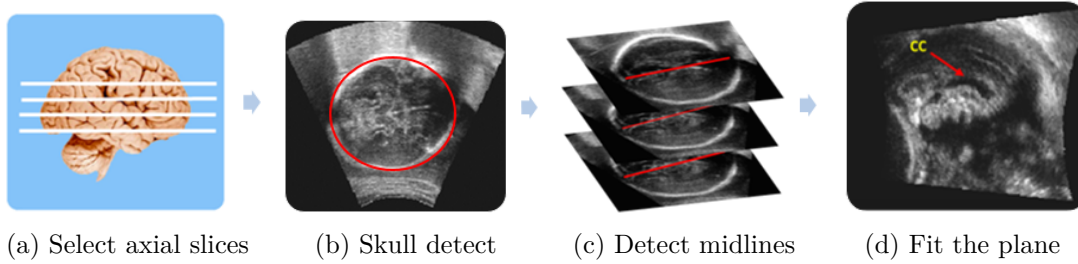


Figure 3.5: Pipeline of the proposed tool for automated mid-sagittal plane selection. (a) Illustration of horizontal slices selection. (b) skull detection in an axial slice. (c) Detected midlines in a stack of axial slices. (d) Detected mid-sagittal plane.

Selection of axial slices The mid-sagittal plane (MSP) divides the brain into two cerebral hemispheres. In ultrasound images, the MSP appears on the axial plane in the form of a bright line, i.e. the midline. Therefore, fitting a plane to these midlines will approximate the MSP. Firstly, a stack of 2D horizontal slices is sampled spanning the supratentorial region¹ at regular intervals ($s=1.2\text{mm}$) from each 3D US image. These slices are then used for skull detection.

¹In anatomy, the supratentorial region of the brain is the area located above the tentorium cerebelli, which contains the cerebrum.

Skull detection The next step of the pipeline is skull detection. The detail of the algorithm has been described in the Sect.3.1. The detected skull is used to make a mask image to remove the extra-cranial tissues and the skull.

Midline Identification Having successfully masked the brain regions, the midline is identified in each 2D image slice. It is expected that the best fit plane will intersect the midlines in all 2D axial planes within the brain volume. The midline of the brain is the longest connected structure of width $2.2 \pm 0.4\text{mm}$. Signed feature symmetry is used to enhance the ridge-like midline [104]. The main idea behind the feature symmetry is the difference of symmetric and asymmetric frequency components in a signal can quantify the extent to which a structure is locally symmetric. Signed feature symmetry retains the sign of the symmetric frequency components, which gives information on whether the region is a ‘peak’ or a ‘valley’. As the MSP always appears as bright, symmetric structure in axial slice, it is intuitive to use signed feature symmetry to enhance the midline. It is defined as:

$$S(x) = \frac{\max(0, (f_e(x) - f_o(x) - T))}{A(x) + \epsilon} \cdot \text{sgn}(f_e(x)) \quad (3.3)$$

, where $f_e(x)$ and $f_o(x)$ are the symmetric and asymmetric frequency components, T is a sensitivity threshold, $A(x)$ is the local amplitude, ϵ is a small number that avoids division by zero, and $\text{sgn}(\cdot)$ is the signum function. Specifically, the signed feature symmetry is calculated for the masked brain image using a log-Gabor radial filter [105], defined as,

$$G(\omega) = \exp\left(-\frac{\log\left(\frac{\omega}{\omega_0}\right)^2}{2\log(\sigma_0)^2}\right), \quad (3.4)$$

where $G(\cdot)$ represents the log-Gabor filter. Here, ω_0 denotes the centre-frequency of the passband and governs the scales of structures that are picked out by the filter. It can be transformed to the wavelengths of structures with: $\lambda_0 = 2\pi/\omega_0$. σ_0 denotes the bandwidth. This controls how selective the filter is. In this work, the value is chosen as $\lambda_0 = 22.8\text{mm}$, bandwidth $\sigma_0 = 0.6\text{mm}$.

The obtained signed feature symmetry is the binarized with a threshold of zero, produces a binary image. Two bright points in this binary image are randomly selected to generate a line and the bright pixels it crosses cast a vote on this pair of points. The process is repeated iteratively and the pair of points with the maximum number of votes in every 2D axial slice is stored. The detail description of the algorithm is given as:

Given:

Set S of n points: $p_1, \dots, p_n$.

Number of iterations, T .

1. Initialize the parameter for a 2D line as:

$l_{max} : y = a_{max} * x + b_{max}$, where $a_{max} = 0$, $b_{max} = 0$, and $C_{max} = 0$.

2. Do for $t = 1, 2, \dots, T$:

a. Randomly selected two points p_{t1}, p_{t2} .

b. Fit a line $l_t : y = a_t * x + b_t$ to these two points.

c. Calculate the number of bright pixels on this line as C_t .

d. If $C_t > C_{max} : a_{max} = a_t, b_{max} = b_t, C_{max} = C_t$.

3. Output the final mid-line parameter:

$l_{max} : y = a_{max} * x + b_{max}$

Fitting the mid-sagittal plane The stored points are processed with the Random sample consensus (RANSAC) algorithm, an iterative method to estimate parameters of a mathematical model from a set of observed data [106]. One of the key assumptions of this algorithm is that the data set is not ‘clean’. In other words, it consists of some ‘inliers’, whose distribution can be explained by the true model parameters, and some ‘outliers’ that may result from noise and thus do not fit the model. This feature is particularly favourable in this application as the fetal brains present complicated anatomy patterns and certain structures (i.e. ventricles) can generate misleading bright lines in some of the axial slices. Instead of using all the detected midlines to fit a plane using the least square estimation (LSE), the exclusion of a subset of data points add robustness to the algorithm and lead to better prediction. The RANSAC algorithm can be summarized as the repetition of the following steps:

1. Select a random subset from the observed point set as *inliers*.
2. Fit a plane to these *inliers* using LSE.
3. Test the rest of the data against this model, i.e. calculate the distance d between the point and the fitted plane. A point is considered as part of the consensus set if d is smaller than a threshold t .
4. The consensus set with the largest number of data points is recorded as the final inliers set.

In particular, a point will be considered as an outlier if it does not fit the model within the distance threshold $t=0.6$ mm. Subsequently, the mid-sagittal plane is achieved by fitting a plane using the output inliers.

3.3 Experiments

3.3.1 Image dataset

The skull stripping algorithm is tested on 248 randomly-selected 2D ultrasound fetal brain images from 25 healthy fetuses age range from 20 to 30 weeks. The typical size of a image is 210×190 pixels, with the resolution of 0.6×0.6 mm.

The plane detection algorithm is tested on 3D fetal neurosonography data collected from 62 healthy fetuses, ranging from 22 to 28 gestational weeks. The typical size of each image is $220 \times 210 \times 170$ voxels, with each isotropic voxel measuring 0.6 mm in size. The average processing time for each volume was 9.3 seconds on an Intel T1700 3.5 GHz processor. Three experiments were conducted: (a) comparison of the centre location and the size of the skulls that are automatic detected and manually delineated, (b) comparison between the orientation of manually and automatically extracted mid-sagittal plane, (c) ability of this model to recover the shape of a clinical important brain structures: corpus callosum (CC) (i.e. the shape of the CC visualized in automatic derived mid-sagittal slices, comparable to that of the CC in manually-selected 2D planes).

3.3.2 Skull detection accuracy

The skull in each of the sampled 2D images was manually segmented by one human expert. Subsequently, each 2D image was processed using the proposed skull stripping algorithm at different resolutions, i.e. high resolution y-axis =

126 mm, median resolution y-axis = 75 mm, low resolution y-axis = 50 mm. The skull detection error was evaluated in terms of centre deviation and size deviation. Specifically, the centre deviation is the distance between the centre of the automatically detected skull and that of the manual delineated skull. The size deviation is the difference in terms of the BPD (length of the minor axis, see Sect.3.1) and the OFD (length of the major axis) measurements of the fetal skull between the automatic detection and the manual delineation. The algorithm performance with the original resolution and the reduced resolution is reported and discussed in Sect.3.4.

3.3.3 Manual plane selection

In a mid-sagittal view of the fetal brain, the CC appears as a dark crescent in an US image ². To assess the accuracy in plane selection of the proposed system, the mid-sagittal planes were manually selected in a multiview graphical user interface by identifying the central line of the plane in both axial and coronal views, and using the 3D configurations of these two indicators to guide the localization of the mid-sagittal plane (Figure 3.6).

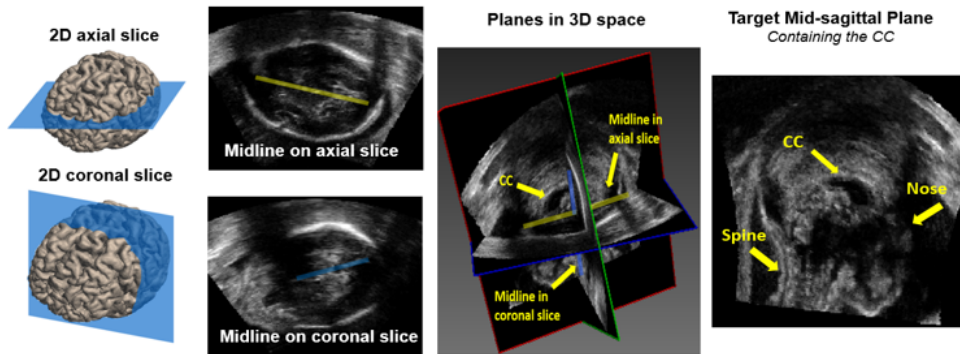


Figure 3.6: Manual plane selection. The midline in an axial slice (yellow) and that of a coronal slice (blue) are identified to guide the localization of the mid-sagittal plane in 3D.

²The detail discussion of this structure and its visual characteristic in ultrasound images can be found in Chapter 6

3.3.4 Orientation deviation

The accuracy of the automated mid-sagittal plane detection was evaluated by calculating its angular deviation from the manually selected mid-sagittal plane. The angular deviations (i.e. dihedral angles) between the normal vector of an automatically selected plane and the corresponding manual plane was calculated. The intra/inter-observer variability in mid-sagittal selection was also assessed by comparing the plane orientation deviation of the results selected by the same examiner at different instances and by two independent examiners. Since the minimal distance between the manual and estimated planes would only be informative when the two planes are strictly parallel, it is then decided to compare the visualization of the structures of interest, namely the CC, on the manual and estimated planes.

3.3.5 Shape Analysis

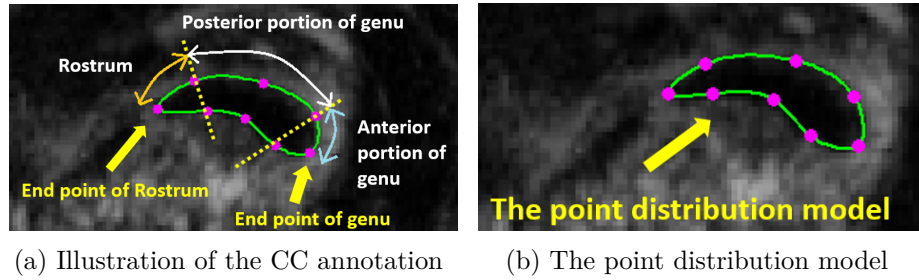


Figure 3.7: CC shape model. (a) The configuration of the fetal CC (b) Eight representative points (pink) are manually annotated on the boundary of the CC (green).

In clinical practice, the sonographer selects a suitable plane from the 3D brain space on the basis of presence, shape, and size of anatomical structures of interest [107]. Analysing the shape similarity and variation of the CC on the automatically/manually selected planes is helpful to assess clinical viability of the plane selection framework. To depict the shape of the CC on a mid-sagittal plane, a point distribution model is established by choosing a set of 8 representative points on the CC's boundary (Figure 3.7(a)). The shape instances on the automatically selected planes are then aligned as closely as possible to each other with the Procrustes analysis [108]. The mean shape of the automatically derived CC is

defined by taking the arithmetic mean of the whole set of registered shape instances. The same procedure is performed on the CC in the manually selected planes. Finally, the two mean shapes of the CC were registered together to compare their variation and overlap.

3.4 Results and discussions

Skull detection accuracy Table 3.1 reports the performance of the skull detection algorithm at different spatial resolutions. It shows the algorithm achieves good accuracy in finding the location (the second column), and the size (the third and the fourth column) of the skull in 2D ultrasound images. Compare the first two rows with the second row and the third row, this shows that reducing the resolution not only reduces processing time, but also increases detection accuracy. One interpretation of this is that reducing resolution also decrease the effect of bright intra-cranial tissues. It is noteworthy that excessive reduction of spatial resolution has a negative effect: the Median resolution outperforms the Low resolution in detection accuracy. It can be explained as the reduction of resolution inevitably loses image information. Moreover, as the image size is nearly halved in this case, the detection prediction accuracy is limited by pixel size. The fourth row in Tab.3.1 lists the inter-observer variations of measuring BPD and HC recorded in the study of [18]. The error rate is slightly smaller but comparable to the automatic detection results using Median resolution. However, it is can be inferred that the detection accuracy can be further improved using a richer template set (more sizes and orientations) at the price of longer computation time. It is important to note that the main aim of the skull stripping algorithm is to remove the extra-cranial tissues efficiently. For the purpose of this work, a suitable balance between algorithm efficiency and detection accuracy is of high priority.

Plane extraction accuracy As shown in Figure 3.8, the automated versus manual errors are comparable to those of the intra/inter-observer error margins, which confirms the validity this algorithm. Though the intra-observer errors have

Table 3.1: Results of skull detection experiments. The centre deviation (Cen Dev) and the size deviation (in terms of BPD and HC measurements) are reported. Each row shows experiment results with different resolution (See Sect.3.3.2). The fourth row of the table lists the inter-observer disagreement reported in [18], where measurements by 8 different observers are compared.

Resolution	Y-axis size(mm)	Cen Dev(mm)	BPD(mm)	HC (mm)
High	126	3.4 ± 2.0	3.4 ± 2.7	6.7 ± 5.6
Median	75	2.2 ± 1.6	2.1 ± 1.4	4.5 ± 3.7
Low	50	2.6 ± 1.8	2.6 ± 1.7	5.1 ± 4.2
Inter-observer	—	—	1.8 ± 1.2	4.4 ± 3.5

relatively the lowest deviation, the automated-versus-manual and inter-observer variability share similar deviation. More than 75% of the automated-versus-manual angle deviations are lower than 5° in all age groups, as well as the intra-observer deviation. This shows that this system is functional over a broad GA range, and is robust to image variability related to GA.

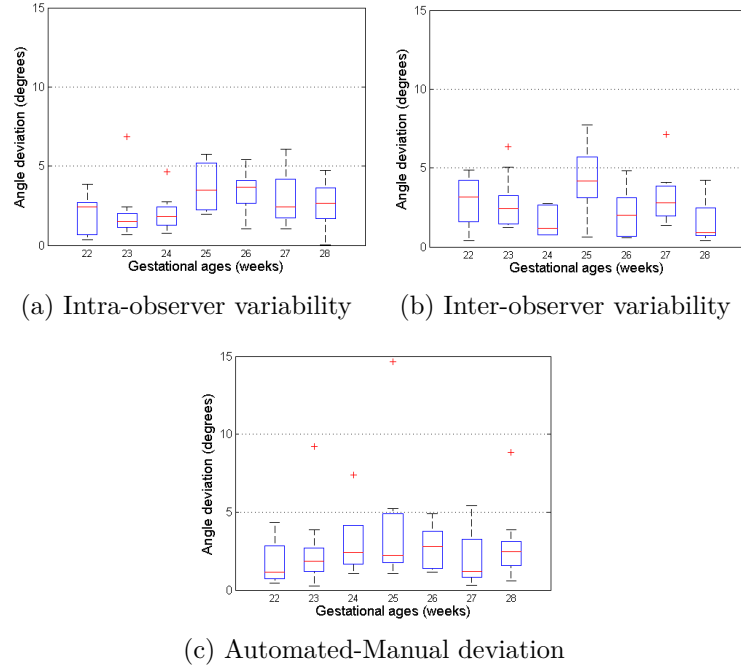


Figure 3.8: Planes' orientation deviation. The axis represents fetal GA versus the orientation deviation. Box plot showing median (red lines) and inliers within ± 2.7 standard deviation.

The mean shapes of the CC from the automatically/manually selected planes

are shown in Figure 3.9(a). It is clear that there is considerable overlap between the recovered CC shapes in the automatically (blue) and manually (red) selected planes. To further assess the proximity of the two shapes, the distance between the boundaries was calculated. The spatial deviation was computed on 90 points selected on the CC from the manually selected plane at 1 mm intervals and their corresponding points from the automatically selected planes. The mean distance between pairs of points is 0.32 ± 0.25 mm, takes up 1.19% of the perimeter of the mean shape of the CC (27mm). This validates the shape consistency of the CC on the automatically selected planes and that of the manually selected ones. Figure 3.9(b) shows the ± 1 standard variation of the CC shape in automatically selected plane. The shape variation of the CC in automatically selected plane is within the acceptable range, suggesting the clinical viability of this model for recovering the CC for growth assessment.

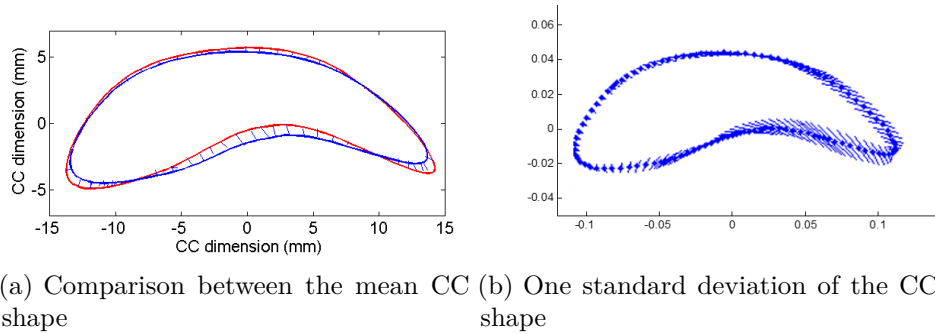


Figure 3.9: CC shape analysis. The red line depicts the CC in manually selected plane and the blue one depicts that in automatically selected plane.

3.5 Conclusion

This chapter first presented an original skull-stripping algorithm which was designed to overcome practical issues of traditional skull detection methods (see Chapter 2 for details). Secondly, the chapter demonstrated a plane detection algorithm to identify the mid-sagittal plane from 3D fetal neurosonography to visualize the CC. This tool is fully automated, and requires no user-intervention during processing. The similarity between the automatically and manually selected planes was validated

by comparing their angular deviation and the CC shape overlap. It has shown that the angular deviations between the two sets are within the intra-observer plane selection errors boundary and they share similar variance with inter-observer variability. Furthermore, the mean shape of the CC recovered from the automatically selected planes is highly similar to that of manually selected planes, while the shape variation of the CC lies within an acceptable range. In the next chapter, we will look at individual brain structure on 2D ultrasound planes in more details, and the mid-sagittal plane is among one of the investigated planes.

4

Segmentation of Key Clinical Anatomical Structures in 2D Ultrasound Images

Contents

4.1	Introduction	50
4.1.1	Proposed framework	51
4.2	Method	52
4.2.1	Target Fetal Brain Structures	52
4.2.2	Candidate Region Selection	53
4.2.3	Region Descriptor	54
4.2.4	Classification and voting mechanism	57
4.3	Implementation details	58
4.4	Datasets	59
4.5	Experiments	60
4.5.1	Classification and feature comparison	60
4.5.2	Segmentation accuracy	61
4.6	Results	61
4.6.1	Classification performance	61
4.6.2	Feature set comparison	62
4.6.3	Segmentation accuracy	66
4.7	Discussion	67
4.7.1	Challenging cases	68
4.7.2	Failure cases analysis	70
4.8	Summary	71

The previous chapter described visualization techniques as the first step of interpreting and understanding an fetal ultrasound brain image. This chapter now

turns to look at extraction of individual anatomical structures. Specifically, this chapter describes a general framework for automatic segmentation of fetal brain structures in 2D ultrasound images which simultaneously yields automated biometry measurement. The approach is based on a novel region descriptor that characterizes the shape and local intensity features of different neurological structures without explicit models. Validation experiments are presented to segment two fetal brain structures of clinical importance that have quite different ultrasonic appearances - the corpus callosum (CC) and the choroid plexus (CP). The next chapter will extend this work by investigating multiple brain structures in 3D space. The research presented in this chapter is under journal review:

Ruobing Huang, Ana IL Namburete, J Alison Noble. Learning to Segment Key Clinical Anatomical Structures in Fetal Neurosonography informed by a Region-based Descriptor. Journal of Medical Imaging (under second review).

4.1 Introduction

As described in current clinical guidelines [3], standard fetal neurosonogram examination involves identification of several key brain structures and measurement of skull size. These observations and metrics offer a macroscopic and coarse evaluation of fetal brain health. It would help to identify severe abnormalities e.g. complete absence of corpus callosum, microcephaly and macrocephaly. However, the applicability to conditions with subtle symptoms is limited.

Some clinical studies have reported the size of anatomies such as the atrium (which contains the choroid plexus) and the cerebellum are informative bio-markers, and have related these to fetal health [41, 42]. However, as these brain structures can be difficult to identify, measurement of them is not included in sonographic examination standards. This is particularly true in late gestation when skull bone begins to calcify and limits the visibility of brain structures. An automatic segmentation algorithm dedicated to ultrasound images is needed to address the above issues.

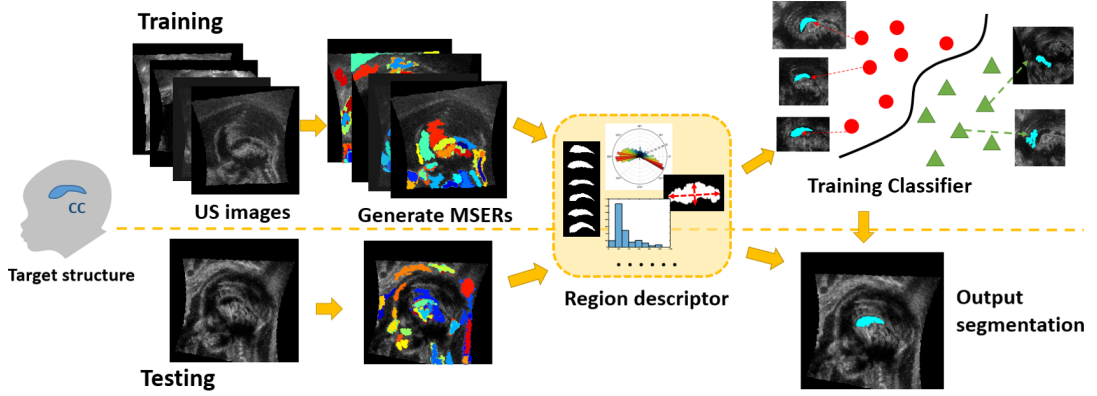


Figure 4.1: Schematic flow chart of automated segmentation framework.

4.1.1 Proposed framework

This chapter offers a new approach to automate ultrasound segmentation of fetal brain structures. The originality of this work lies in transforming the complex structure delineation problem into a region classification task which utilizes a novel region-based descriptor. Validation experiments are carried out on segmenting two structures that are important in monitoring the developing brain: the corpus callosum and the choroid plexus, to prove both the usefulness of the approach and its generality.

Figure 4.1 presents an overview of the automated segmentation framework (details of the method are given in Section. 4.2). The proposed approach consists of three steps: (1) candidate region selection; (2) region characterization, using the proposed region descriptor combined with a boosting classifier trained to identify the desired candidate region; and 3) region classification. In testing, an unseen image is passed through the same pipeline to generate an automated segmentation.

The outline of this chapter is as follows. Section. 4.2 describes the method. It first discusses the appearance of the target brain structures in Section.4.2.1. This motivates the selection of candidate regions described in Section.4.2.2. It then defines a region description model in Section.4.2.3. Section 4.2.4 describes the machine learning-based framework for region-classification (segmentation), with implementation details in Section. 4.3. Section. 4.4 presents the datasets used in the experiments. Validation experiments and results are reported in Section.4.5,

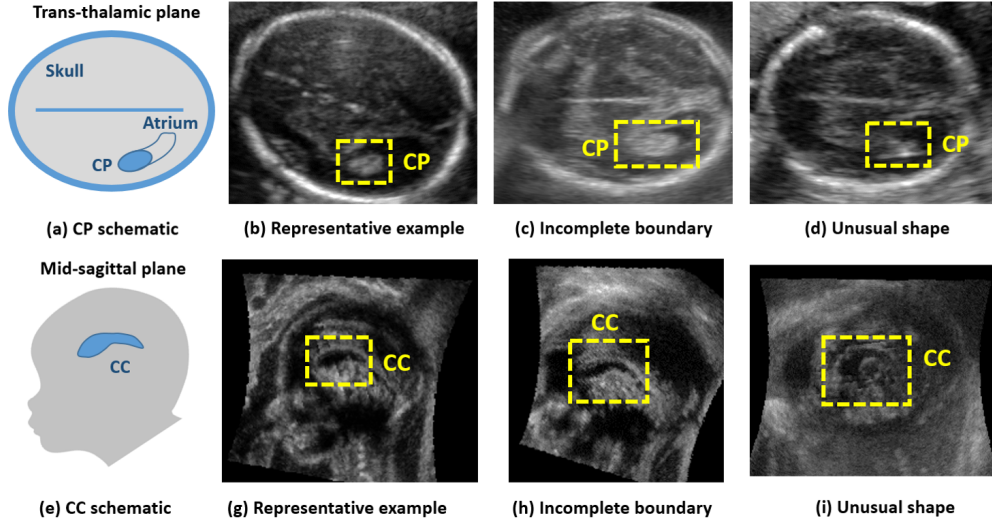


Figure 4.2: Trans-thalamic plane (first row) and mid-sagittal plane (second row) of the fetal brain. Schematics for CP and CC are shown in (a) and (e) respectively. Other sub-figures in each row are corresponding ultrasound image examples. The CP or CC is bounded by a yellow dashed box in each image.

and Section. 4.6 respectively. It concludes, in Section. 4.7, with a discussion of the contributions of the work and highlight areas of possible future work.

4.2 Method

4.2.1 Target Fetal Brain Structures

The interaction of the ultrasound beam with a fetal brain structure results in an acoustic pattern response that either yields a bright anatomical mapping of the structure or, alternatively, a dark acoustic signature. The brightness of the signature depending on its echogeneity that an experienced sonographer learns to interpret proficiently. In order to automate structure detection and segmentation, it is interesting to find how to detect, segment and characterize ‘blob-like’ acoustic responses that correspond to structures of interest (SOIs). Specifically, this work focuses on two brain structures: the choroid plexus (CP) and the corpus callosum (CC). As illustrated in Fig.4.2, the CP and CC differ in terms of intensity, shape and spatial configuration (different anatomical context) and hence provide two excellent examples on which to test the automatic detection and characterization framework.

4.2.2 Candidate Region Selection

It is assumed that, in a given ultrasound image, a fetal brain structure appears as a bright or dark region depending on the echogenicity of its tissue. Tissue with low echogenicity presents as a dark blob, and vice versa. Both CC and CP are examples of structures which are approximately homogeneous which leads to a first assumption of the proposed framework, namely that a target of interest is a local extremal region regardless of its echogenicity (whether it is a local maximum or minimum). Furthermore, it is intuitive to assume that the ideal segmentation of such a fetal brain structure should be contiguous and stable over thresholding. Based on the second assumption, we model candidate regions as Maximally Stable Extremal Regions (MSERs) [109]. Specifically, we define image I to be a mapping from the coordinate domain D to the intensity domain. Region R is a contiguous subset of D , and ∂R is the region boundary. A pixel p belongs to R : $p \in R$, while q is a pixel that belongs to ∂R . An extremal region is either a maximal extremal region or a minimal extremal region. We define R as a maximal extremal region if $I(p) > I(q)$ for all $p \in R$ and $q \in \partial R$ (or $I(p) < I(q)$ for minimal extremal region). An important property of extremal regions is their nestedness, that is, if a set of extremal regions is sorted in a monotonic order of their intensity values: $R_1, \dots, R_i, R_{i+1}, \dots$, then it follows that: $R_i \subset R_{i+1}$ for all R_i . Then, we choose the maximally stable region from the whole set of nested extremal ones. We define a function of i :

$$f(i) = (|R_{i+\delta}| - |R_{i-\delta}|) / |R_i| \quad (4.1)$$

which represents the area variation between two nested regions, where δ denotes their intensity interval and $|\cdot|$ denotes area. It is intuitive that the region is more stable if its intensity changes sharply over its boundary (large gradient), which leads to a smaller value of $f(i)$. Conversely, a larger $f(i)$ indicates instability as the area of nested extremal regions varies. Finally, an extremal region R_j is defined to be maximally stable (a MSER) if $f(j)$ is a local minimum of $f(i)$ and smaller than γ : a threshold that controls the stability of the selected region.

In practice, the aforementioned parameters for generating MSERs (controlling threshold γ and intensity interval δ) are chosen empirically and different problems usually have different optimal parameter sets. For example, the CC may be best selected by parameter set $S_a = \{\gamma_a, \delta_a\}$, while the CP is best selected using a different parameter set $S_b = \{\gamma_b, \delta_b\}$. Tuning parameters for each specific task limits the applicability of the algorithm on different tasks, and is time-consuming. To improve the adaptability as well as reduce human intervention, it generates multiple sets of MSERs using different values of γ and δ sampled from a broad range of values. The machine learning algorithm selects the appropriate region from these sets of MSERs as described in Sect. 4.2.4. As a result, no human intervention is needed to select different parameters to segment different structures. Thus, the proposed framework readily adapts to a new segmentation task. In addition, note that there may be one or more MSERs corresponding to the desired region (i.e. true-positive examples) for each image as the MSERs were generated using different parameters.

Detected MSERs (as illustrated in Fig.4.2.2) serve as candidate regions of the SOI and their features are extracted to form a region descriptor. As shown in Fig.4.3(c) many MSER candidate regions are identified but within the pool of candidates lies a segmentation of the correct structure (shown in red in Fig.4.3.(c)). To identify the correct structure, a rich region-based feature descriptor is constructed. This is subsequently used within a machine-learning algorithm to identify the correct structure. Note that the majority of MSERs belong to negative class which affects the selection of the classification algorithm in the framework design.

4.2.3 Region Descriptor

The shape of an anatomical brain structure is one of its most distinguishing features. It can be challenging to characterize fetal brain anatomy, where shape is usually irregular and has large intra-class variation. This work can characterize the shape of the SOI without an explicit model. All candidate regions are first pre-processed to remove irrelevant information (Fig.4.4). To be more precise, the regions are binarized, resized, and aligned thus eliminating the effect of intensity, size and

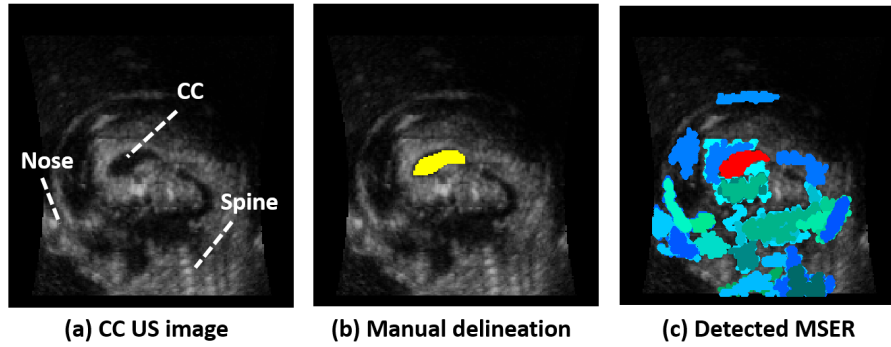


Figure 4.3: Candidate region detection. (a) shows the original fetal ultrasound image of mid-sagittal plane. The CC, the spine and the nose of the fetus are pointed out. (b) displays the manual delineation of the CC in yellow. (c) presents the detected MSERs, where the positive example is plotted in red and the negative examples are randomly assigned with blue or green for comparison. It can be seen that the majority of MSERs belong to negative class.

orientation. The region width is scaled to the same size while the aspect ratio is maintained to prevent distortion. Region alignment is achieved by rotation along the major axis to align their orientations with a common image axis.

The shape of each candidate region is then described in two ways:

Shape Descriptor 1: a set of shape templates is defined as a randomly selected subset of manual delineations. The delineations are pre-processed as described earlier.

The normalized cross correlation (NCC) between a pre-processed candidate region and each template is computed, with the maximum NCC value selected as a measure of shape similarity. Figure 4.4 shows the template-matching method applied to two separate regions. Region A (green) is the MSER generated from the desired SOI (i.e. CC, in this example), Region B (blue) is a randomly chosen MSER belonging to the background (referring to Fig.4.4). The column of templates shows six CC shape templates (randomly chosen for demonstration). The rightmost histogram compares the values of shape features extracted from Region A and B respectively. The features of Region A (green bars) have greater values than those of Region B (blue bars) as Region A's shape is more similar to the templates.

Shape Descriptor 2: A polar-histogram is constructed to describe the region area distribution centred on the region centroid. The polar region is divided evenly

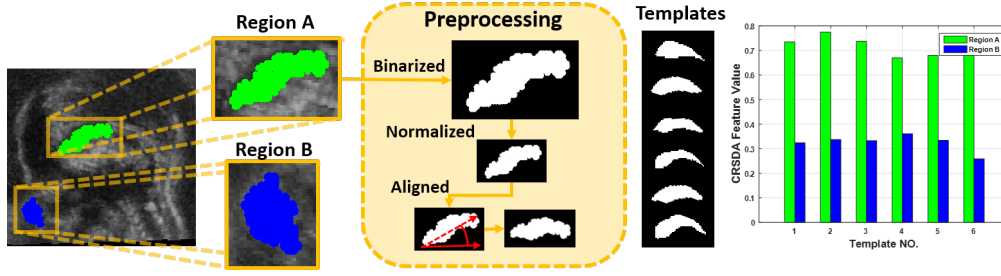


Figure 4.4: Shape Descriptor 1: pre-processing procedure and feature extraction. Shape Descriptor 1 is derived from NCC values, and its value is proportional to the shape similarity between candidate region and templates. The column of templates shows six examples of CC templates used in the experiments.

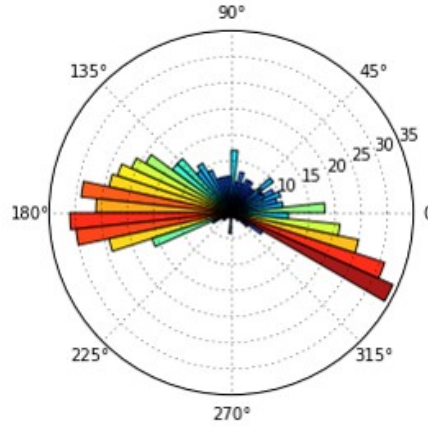


Figure 4.5: Shape Descriptor 2: The area distribution of the candidate region on a polar coordinate system. The colour red represents a large number of pixels lies within the fan section, while blue represents a small number.

into 60 bins. Each bin contains a count of bright pixels (pixels within the region mask) that lie within the corresponding fan section of the polar coordinate system (refer to Fig.4.5). This descriptor is motivated by [110] for cell detection, which used a method of boundary distribution instead. This descriptor provides 60 features.

Additionally, six other feature descriptors are derived from intensity, scale and location information, namely:

Mean intensity. It coarsely characterizes the intensity level of the region.

Histogram of pixel intensities in the region. The intensity distribution illustrated by this histogram characterize both the region stability (intensity variation) and texture (repeated intensity pattern). It also complements the region brightness information apart from the mean intensity as it is biased if large intensity

variations exist.

Length of the minor and major axes of the region. These approximate the region dimensions while being rotation-invariant. The minor and major axes are derived from the smallest ellipse that enclosed the region.

Eccentricity of the region. It is the ratio of the length of the minor axis to the major axis. A smaller eccentricity indicates a more circular region, while a larger value indicates an elongated shape.

X and Y coordinates of the region centroid. These denote the location of the candidate within the image.

Histograms of the normalized X and Y components of the region pixel co-ordinates. Specifically pixel coordinates within the region (x, y-coordinates separately) divided by the size of image. These two histograms denote the location distribution of the region which is more robust to the change of the viewpoint compared to the absolute centroid coordinates. This feature is scale-invariant with respect to the image size. This descriptor provides 60 features.

In total this gives a region feature descriptor with 190 terms.

4.2.4 Classification and voting mechanism

To preserve the generality of the algorithm, the image is not pre-cropped and minimal human intervention is allowed during parameter selection (explained in Sect. 4.2.2). As a result, candidate region selection generates a larger proportion of negative class responses relative to positive ones, e.g. 28 negative versus 1 positive case in Fig.4.3(c). This results an imbalanced dataset which may bias the classifier. To accommodate this we employ Random Under-Sampling (RUSBoost) - an iterative boosting algorithm designed to deal with data imbalance for classification [111]. In this method, *Random under-sampling* is applied to remove the majority class examples until the desired class ratio is achieved. Specifically, at iteration t , a temporary training dataset S_t is created from the dataset S using random under-sampling to remove majority class examples until a desirable percentage (e.g. 50%) of the minority class is achieved. For each candidate region R_i , the region descriptor

extracts its features and maps it as a feature point x_i in feature space X . Let $y_i \in Y$ be the class label of x_i and $D_t(x_i)$ be the weight of x_i in iteration t . A simple decision tree (weak-learner) $G_t(x_i) : X \rightarrow Y \in [0, 1]$ is then trained based on S_t . The loss function ϵ_t is defined as:

$$\epsilon_t = \sum_{i=1}^n D_t(x_i) \mathbb{1}(G_t(x_i) \neq y_i), \quad (4.2)$$

with $\mathbb{1}(\cdot)$ is an indicator function.

In the next step, the weight $D_t(x_i)$ is updated as:

$$D_{t+1}(x_i) = \frac{D_t(x_i)}{Z_t} \alpha_t (y_i G_t(x_i)), \quad (4.3)$$

where Z_t is the normalizing factor and α_t is given by:

$$\alpha_t = \frac{1}{2} \log \frac{1 - \epsilon_t}{\epsilon_t}. \quad (4.4)$$

This process increases the weights of misclassified examples and decreases those of correctly classified ones. After training is complete, all weak learners are combined to yield the final classifier.

On rare occasions, ultrasound artifacts may possess similar features (e.g. intensity or shape) to the SOI which can lead to a confusing prediction (i.e. false-positive examples). However, our experiments have shown that those artifacts are much less stable across different values of the parameters.

To combine coexisting positive examples and eliminate artifacts, a simple voting mechanism is applied after region classification. Specifically, every candidate that is classified as positive casts a vote on the corresponding region. The region with maximal votes is selected as the SOI after accumulating all the votes.

4.3 Implementation details

The stability threshold and intensity interval were set as: $\gamma = 0.1$ to 0.4 (step size = 0.05), $\delta = 0$ to 5 (step size = 0.5). The number of RUSBoost iterations were set as $t = 200$ and a learning rate of 0.25 were chosen empirically. The average runtime for the segmentation for one CP image is $12.1s$ while for a CC image is $12.6s$ on a 3.5GHZ , 16GB RAM computer.

4.4 Datasets

Experiments were conducted on ultrasound datasets of two fetal brain structures of clinical importance, CP and CC respectively. The CP was selected as it plays a primary role in the formation of cerebrospinal fluid and is associated with several congenital diseases [112, 113]. The fetal CC is considered as a sensitive biomarker of normal brain development and maturation [114]. Indeed, malformation and agenesis of the CC have been frequently related to several genetic diseases [115, 116]. The image datasets which contained each of these structures are introduced next.

Choroid Plexus Dataset: The choroid plexus appears as a bright circular region in the trans-thalamic (TT) ultrasound image plane of the fetal brain, with varying shape and size across gestation (Fig.4.2). The dataset we have used consists of 120 2D ultrasound images of the standard TT plane of the fetal brain, which were randomly selected from a large clinical study database of healthy subjects [99]. The images were of healthy fetuses between 20 to 22 gestational weeks corresponding to the period of clinical anomaly scanning for Trisomy 18 [117]. The images were acquired using a 2D linear probe (Phillips HD9) at 2-5MHz wave frequency and at a pixel resolution of 2 mm x 2 mm. The data were divided into a training group of 68 images and a testing group of 52 images.

Corpus Callosum Dataset: The corpus callosum is observed as a single comma-shaped echogenic structure on the mid-sagittal plane of fetal ultrasound brain image (Fig.4.2). Our dataset consisted of 219 2D ultrasound images of the mid-sagittal plane of the fetal brain. The subjects were randomly selected from 21 to 30 gestational weeks fetuses from the same database as described above [99]. The typical size of each image was 210x170 pixels, with each isotropic pixel measuring 0.6 mm in size. Images were divided into a training group (137 images) and a testing group (82 images).

Detain delineations of the structures are carried out by one expert, while the linear measurements (e.g. the length of the CC, the width of the CP) were performed by two experts for comparison.

4.5 Experiments

4.5.1 Classification and feature comparison

This work evaluated the performance of three different classifiers for CC and CP segmentation respectively. Specifically, it compared RUSBoost with SVM and AdaBoost. The systematic error of different classifiers is reported in terms of accuracy, sensitivity, specificity and precision (Table 4.1). Given the number of true positive (TP, i.e. the positive candidate regions that have been classified as such), true negative (TN), false positive (FP) and false negative (FN), the sensitivity is calculated as:

$$Sens = TP / (TP + FN) \quad (4.5)$$

The specificity is calculated as:

$$Spec = TN / (TN + FP) \quad (4.6)$$

The precision is calculated as:

$$Prec = TP / (TP + FP) \quad (4.7)$$

The accuracy is calculated as:

$$Acc = (TP + TN) / (TP + FP + FN + TN) \quad (4.8)$$

To demonstrate the effectiveness of the region descriptor, RUSBoost classification experiments were conducted using different feature sets: shape features only, the appearance and location features (referred as ‘other features’ in Fig.4.8) and the combination of both.

Size and shape of a brain structure change during brain development. The factor of gestational age (GA) is investigated to find whether it affects the algorithm performance. The region classification accuracy was binned and compared based on the gestational age of the subjects.

4.5.2 Segmentation accuracy

The section describes both qualitative and quantitative evaluation experiments on segmentation accuracy. Fig.4.9 and Fig.4.10 show typical CP and CC segmentation results.

Table 4.2 reports automatic segmentation accuracy using two statistics commonly used to compare the similarity of objects: Dice similarity coefficient (DSC) [118] and Hausdorff distance (d_H) [119].

Automatically derived linear measurements were compared with manual measurements by two human observers. The average of the two human experts' results is set to be the ground truth. Following ISUOG guidelines, the length of the CC and the diameter of CP were both manually measured by two different observers (to estimate inter-observer variability) and one observer repeated manual measurement to estimate intra-observer variability. The third column of Fig.4.10 shows five examples of CC length measurement. Figure 4.11 and Table 4.2 compare the inter/intra-observer variability and the deviation of automated results from the ground truth.

4.6 Results

4.6.1 Classification performance

Table 4.1 shows that the algorithm achieves good accuracy for both CC ($Acc = 99.4\%$, $Prec = 94.5\%$) and CP ($Acc = 98.0\%$, $Prec = 97.6\%$) in region identification. All the models had high specificity (= true negative rate). The value of sensitivity (= true positive rate) is interesting as a high sensitivity indicates the algorithm is not over-fitted to the dominant class and can generalize well on unseen data. The RUSBoost model achieved the best sensitivity compared to AdaBoost and SVM (CC: 97.2%, CP: 94.7%).

The Receiver Operating Characteristic (ROC) curves for each classifier is plotted in Fig.4.6. It shows that the RUSBoost classifier achieved the highest Area Under the Curve (AUC), while AdaBoost performed slightly better than SVM.

Classifiers	Accuracy%	Sensitivity%	Specificity%	Precision%
SVM-CC	98.6	82.4	99.3	90.4
AdaBoost-CC	98.7	89.6	99.6	91.1
RUSBoost-CC	99.4	97.2	99.6	94.5
SVM-CP	96.6	86.6	99.6	92.4
AdaBoost-CP	97.9	89.8	99.5	95.2
RUSBoost-CP	98.0	94.7	98.6	97.6

Table 4.1: Table of the performances assessed on the different classifiers. All indices are reported in percentage. The RUSBoost achieved the best results for both CC and CP classification.

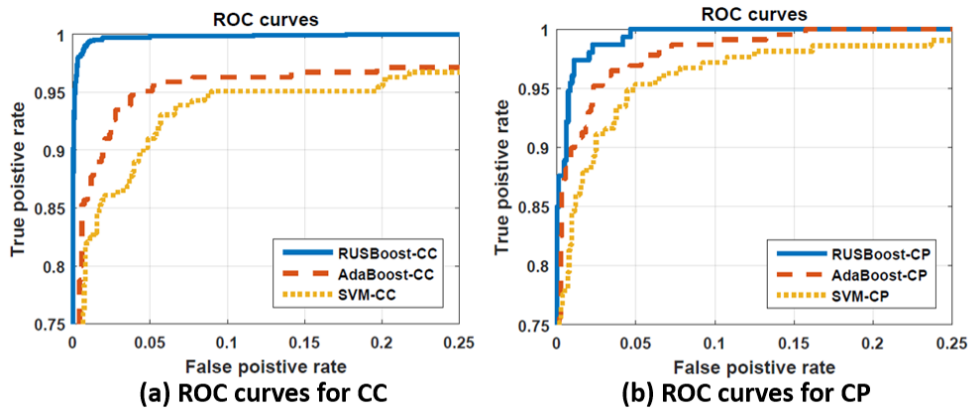


Figure 4.6: Receiver Operating Characteristic (ROC) curves for different classifiers, namely SVM, Adaboost and RUSBoost assigned with different SOI: CC (left) and CP (right) segmentation tasks. The blue solid line represents RUSBoost classifier, while the curve for AdaBoost is plotted using orange dashed line and yellow dotted line for SVM.

As the anatomical structures grow rapidly in the developing brain, their appearance and image contrast can vary greatly across gestation. This section also investigated the influence of GA on the performance of classifiers using the CC dataset as it had larger variation in terms of GA (from 21 to 33 weeks). Figure 4.7 shows classification accuracy and precision with respect to weeks of gestation for CC classification. It can be seen the variation of classifier performance is smaller than 2% in accuracy and smaller than 6% in precision. This shows our framework is robust to typical structure appearance and image contrast variation due to gestational age.

4.6.2 Feature set comparison

Figure 4.8 reports classification error using different feature sets. It shows classification based on shape features alone achieved a lower error rate, $e = 2.2\%$ at

4. Segmentation of Key Clinical Anatomical Structures in 2D Ultrasound Images

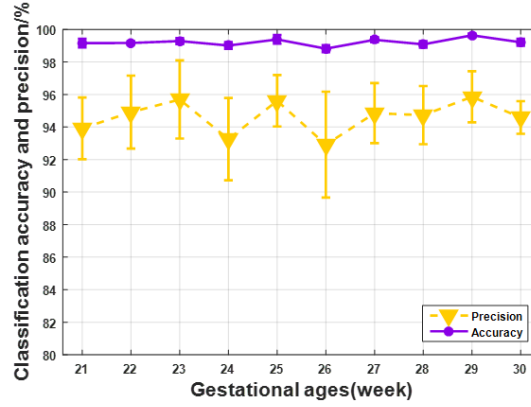


Figure 4.7: The region classification accuracy (purple) and precision (yellow) are plotted with respect to fetal gestational age (GA). The number of volumes ranged from 15 to 25 per GA week.

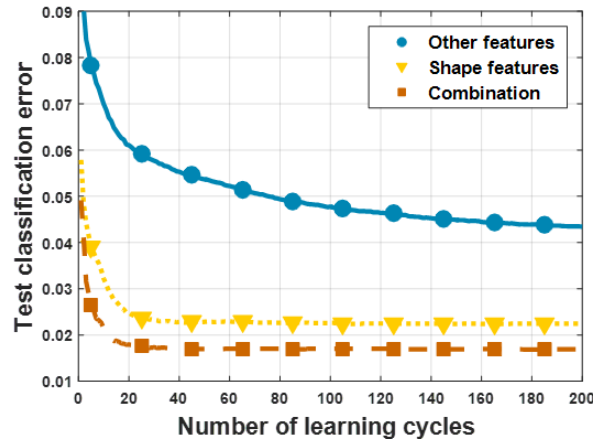


Figure 4.8: The classification error using different feature sets are plotted against learning iterations (shape features: yellow, other feature: blue and the combination: orange).

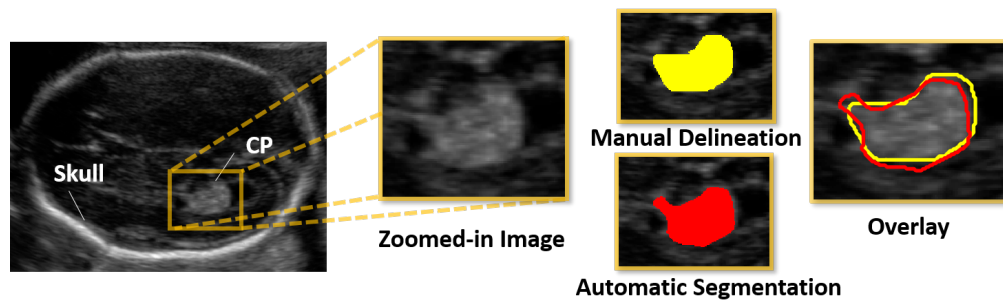


Figure 4.9: Manual and automatic segmentation of CP. Original ultrasound image is shown in the far-left. Sub-image of Choroid Plexus (CP) is enlarged to visualize the details. Manual delineation (yellow) and automatic generated segmentation (red) of CP are also displayed. The two delineations are plotted in the right-most image for comparison.

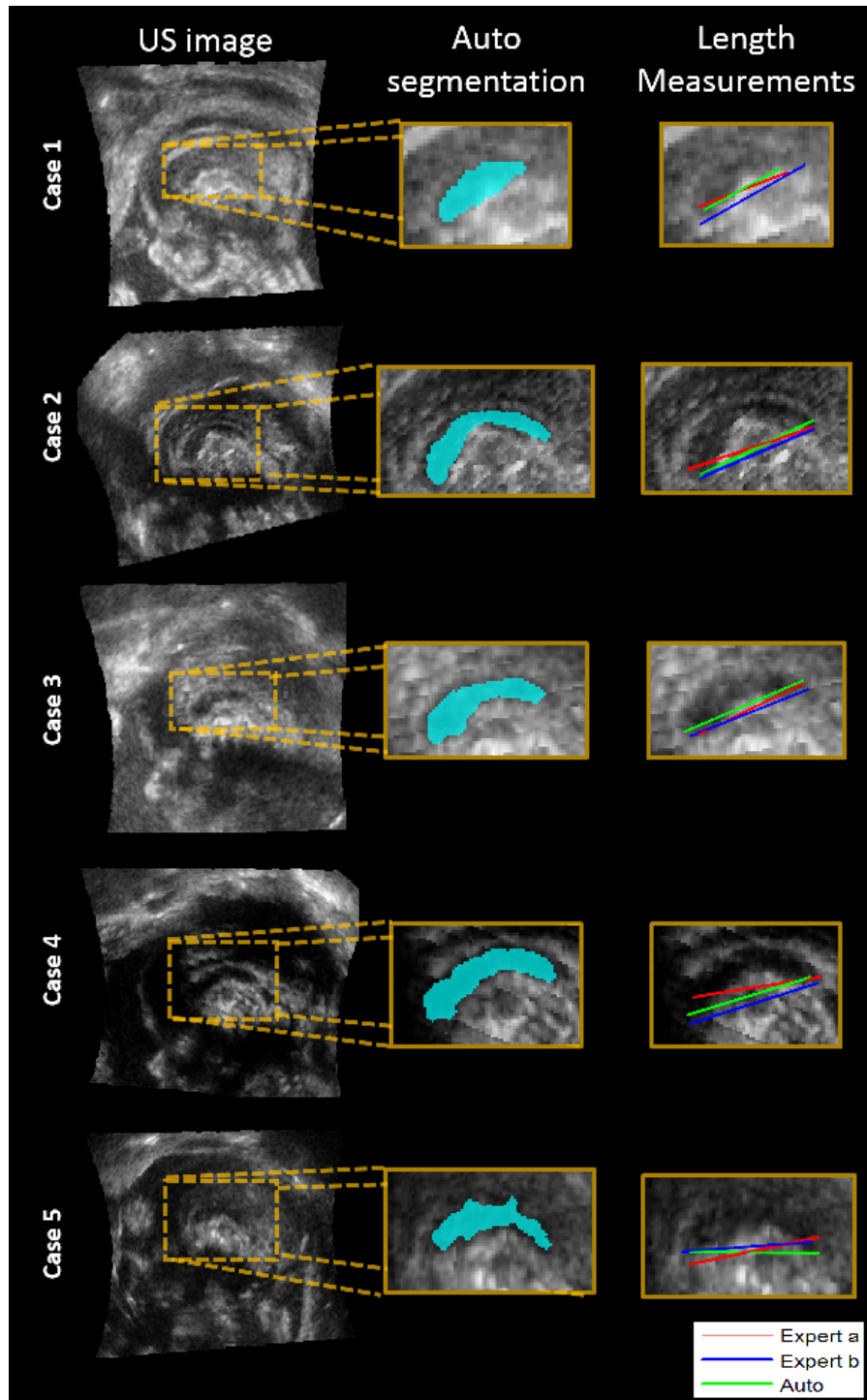


Figure 4.10: CC automatic segmentation results for 5 different cases. Each row belong to one patient. The first column are the original ultrasound images, the corresponding automatic segmentations are presented in the second columns in light blue. The third column shows the CC length measurements carried out by two different human experts and those derived from automatic segmentation (expert a: red line, expert b: blue line, automatic result: green line).

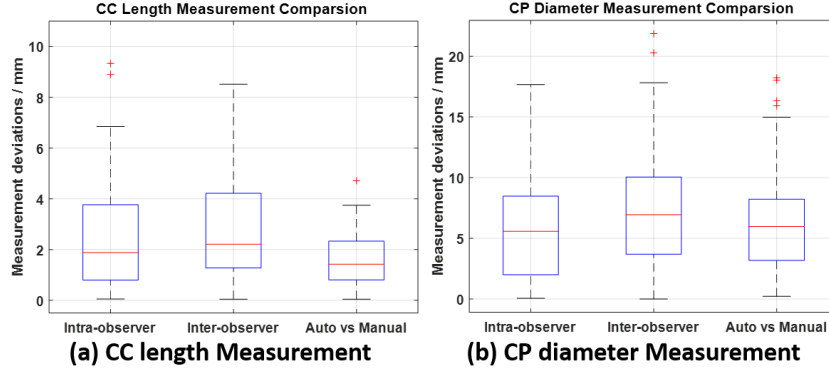


Figure 4.11: Biometric measurements evaluation. (a) The boxplot of measurements deviation of CC length between observers and automated results. (b) Counterpart boxplot for CP diameter measuring.

Brain Structure	CC		CP	
	Median	Mean $\pm$ SD	Median	Mean $\pm$ SD
Intra/mm	1.88	2.54 ± 2.23	5.6	5.9 ± 4.4
Inter/mm	2.62	2.97 ± 2.36	6.9	7.2 ± 5.2
Manual VS Auto/mm	1.59	1.81 ± 1.40	5.9	6.7 ± 4.5
d_H /mm	2.39	2.80 ± 1.91	7.3	7.7 ± 3.4
DSC	0.84	0.81 ± 0.06	0.79	0.76 ± 0.08

Table 4.2: Automatic segmentation accuracy compared with human experts' results. Biometric measurements for CC(length) and CP (diameter) are reported in the first three rows. The measurement for CC has more significant digits as the original image resolution is higher than that of CP(CC: $0.6mm$, CP: $2mm$). Intra represents intra-observer error, while Inter stands for inter-observer error. Auto is the comparison between the automated result and ground truth. The hausdorff distance and DSC between two segments (automated versus manual) are displayed in the fourth and the fifth row respectively.

the $t = 200$ iteration than that of the appearance and location features, $e = 4.3\%$. The combination achieved the lowest error rate, $e = 1.9\%$. Additionally, it can be seen (Fig.4.8) that the algorithm converged quickly with shape features alone (around the 30th iteration), whereas the algorithm had not fully converged at iteration 200 using only the appearance and location features. The fact that the classification accuracy relies heavily on the shape feature, supports our hypothesis in Sect.4.2.3 that the shape is one of the most distinguishing characteristics in fetal brain (structures) segmentation.

4.6.3 Segmentation accuracy

Previous sections shows the algorithm was able to segment SOIs with different imaging conditions. In addition, automatic delineation corresponded well to the manual delineation. Figure 4.9 shows a typical CP segmentation result (automatic: red, manual: yellow) with CP overlaid on the TT ultrasound image. Note that the automatic delineation segments the CP well from the background. Note further that in the upper-left part of the CP, where the two delineations disagree, the human expert made a subjective judgement about boundary completeness.

Figure 4.10 presents CC segmentation results for 5 different cases; the first column shows the original ultrasound images, and the second column shows the corresponding automatic segmentations in light blue. Case 1 and Case 2 in Fig. 4.10 illustrate how the approach is able to deal with large intra-class variation of the shape of the SOI. The framework is also able to accommodate varying contrast (Case 3 and Case 4), acoustic shadows (the head of CC in Case 4) and a low signal-to-noise ratio (Case 5).

The third column of Fig.4.10 compares biometry measurement carried out by two different human experts and those derived from automatic segmentation. The three lines (Expert A: red line, Expert B: blue line, Automatic Result: green line) coincide well with each other for all cases. The largest disagreement occurred in Case 4, where the head of the CC is occluded by acoustic shadows. In this case, the human observer made subjective assumptions about the CC boundary. By contrast, our algorithm may be able to capture small intensity variations which may be imperceptible by eye and offers a more stable measurement.

Figure 4.11 shows boxplots of inter/intra-observer variability and the deviation of automated results from the ground truth. It shows that the automated versus manual errors, for both CC and CP, are comparable to those of the intra/inter-observer error margins. Table 4.2 summarizes the results numerically. It can be seen that the automated-vs-manual error for CC length measurements is $1.81 \pm 1.40mm$, which is slightly smaller than the intra-observer deviation: $2.54 \pm 2.23mm$. The average error value for automated CP diameter measurement lies within intra- and inter- observer

deviation. This further shows that the designed framework can accommodate large intra-class variations in intensity, size and location of one type of brain structure and is able to generate measurements that are similar to manually derived ones.

The segmentation accuracy is also evaluated quantitatively as reported in Table.4.2. DSC values of CC segmentation: 0.81 ± 0.06 is a slightly higher than that of the CP: 0.76 ± 0.08 . This may be explained by the highly varying spatial appearance of the CP (Fig.4.2 (b,c,d)). The value of d_H (2.80 ± 1.91 mm CC, 7.7 ± 3.4 mm CP) are competitive with respect to inter-observer error of linear measurements (2.97 ± 2.36 mm CC, 7.2 ± 5.2 mm CP). It proves that the automatically generated segmentations correlate well with the manual delineations.

Furthermore, recall that the parameters used in the framework are not manually tuned separately for segmenting the CP or CC (the two experiments used the same parameter sets as given in Sect.4.3), yet it performed well for both tasks. This suggests that the method can adapt to differences in appearance, shape and spatial configuration of a SOI.

4.7 Discussion

Last section has reported experiments and their results to evaluate the proposed framework. This section analyses the rationale behind our experiment results and discuss two interesting failure cases.

Section 4.6.1 reported the performance of different classifiers in CC and CP task and concluded that the RUSBoost achieved the best performance in both tasks. It has been discovered that the difference in performance between the three classifiers is greater for CC classification than that for CP (Fig.4.6). My interpretation of this is that CC classification is a more imbalanced classification problem: only 7.6% positive examples among the whole dataset, whereas the CP has 15.44% positive examples. It results from a more complex anatomical pattern in the mid-sagittal plane than that of the trans-thalamic plane (i.e. more MSERs belong to the negative class). This shows that the RUSBoost classifier can handle the class imbalance problem superiorly and is preferable than other common classifiers for our problem.

The shape descriptor demonstrated its importance in assisting region classification (Section 4.6.2). Nonetheless, intensity and location features also provide additional information to assist classification. The first example in Fig.4.12 shows a challenging case of CP segmentation. Although the CP in that image has an unusual shape, the candidate region is still classified as positive as its intensity and location features have high prediction confidence. The resulting automatic segmentation matches well with the manual result. Case 2 in Fig.4.10 is similar as its CC has an unusual elongated shape, yet the algorithm segments the SOI and auto biometric measurements coincides well with that of the expert. It might be interesting to explore other features, suggested by these two examples, as the imaging condition and structure appearance can change greatly in fetal neurosonography.

The process of designing the region descriptors requires understanding and prior knowledge of ultrasound image and fetal brain anatomy. Recent success of neural networks (NNs) for large-scale computer vision tasks [77, 120, 121], has lead to a shift away from hand-crafted features. While NNs report excellent performance on an increasing number of medical images analysis tasks [122, 123], they are best-suited to problems with large annotated datasets, while our work is constrained by data availability. The proposed region descriptors is better suited and a simpler way for solving the presented task.

4.7.1 Challenging cases

To demonstrate the algorithm performance on challenging cases, this section further discusses the cases with the largest Hausdorff distance and the smallest DSC within the entire sets (refer to Case 2 and 3 in Fig.4.12). As the CP is connected with a thin bright structure in the original ultrasound image in Case 2, the automatic segmentation looks almost like a shooting star. As its tail section only appears in a narrow region in the upper-right direction, only one or two feature values would be different from that produced by the manual delineation using the described region descriptor. The result recorded high Hausdorff distance while still reported reasonable DSC since its head correlated well with the manual delineation.

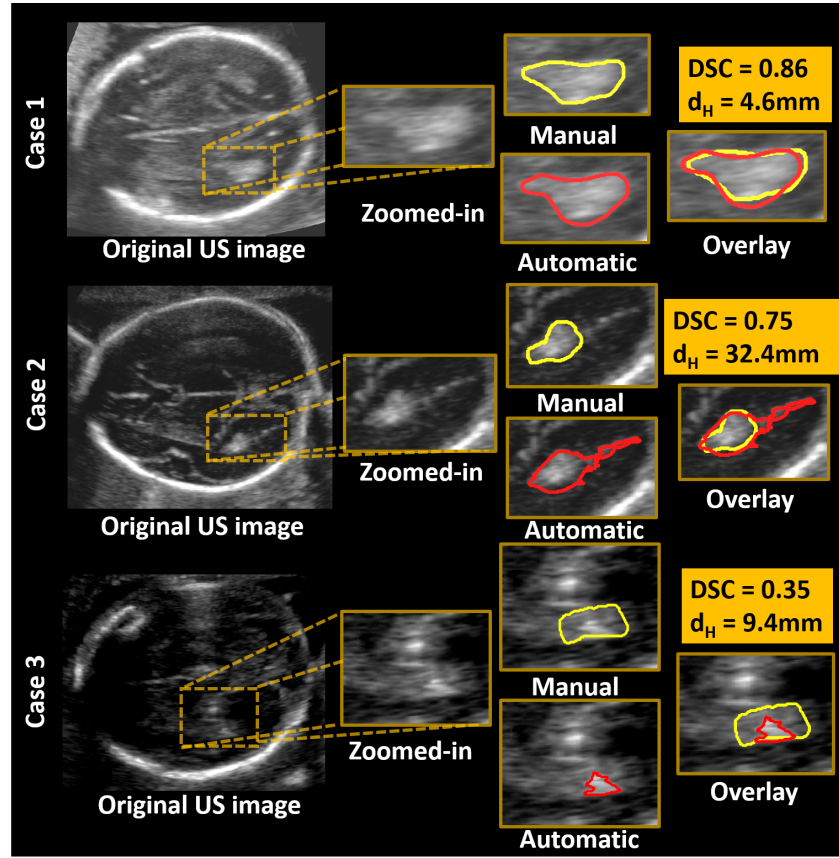


Figure 4.12: CP automatic segmentation results. Each row displays a single subject. The CP region and the segmentation results are enlarged for easier examination. The manual segmentation is plotted in yellow while the automatic result is plotted in red. The two outlines (for each case) are then overlay on the same image for comparison. The value of the DSC and d_h is also displayed in the upper-right corner.

It should be noted that the human observer ignored the ‘tail’. Segmentation methods that use pre-defined models may avoid this situation, but their adaptability towards varying anatomy shape is weak. The proposed method’s robustness towards connected small structures might be improved by adding a post-processing module in the future.

Case 3 in Fig.4.12 has the smallest DSC value as the algorithm picks up a small bright region within the manual segmentation. This behaviour is rational as the CP nearly merges with the surrounding bright tissue in the original image, while the small triangular region in the bottom-right of the image satisfies the intensity and location criteria of a typical CP. In addition, this result demonstrates that the proposed method does not have strict pre-selection criteria in candidate

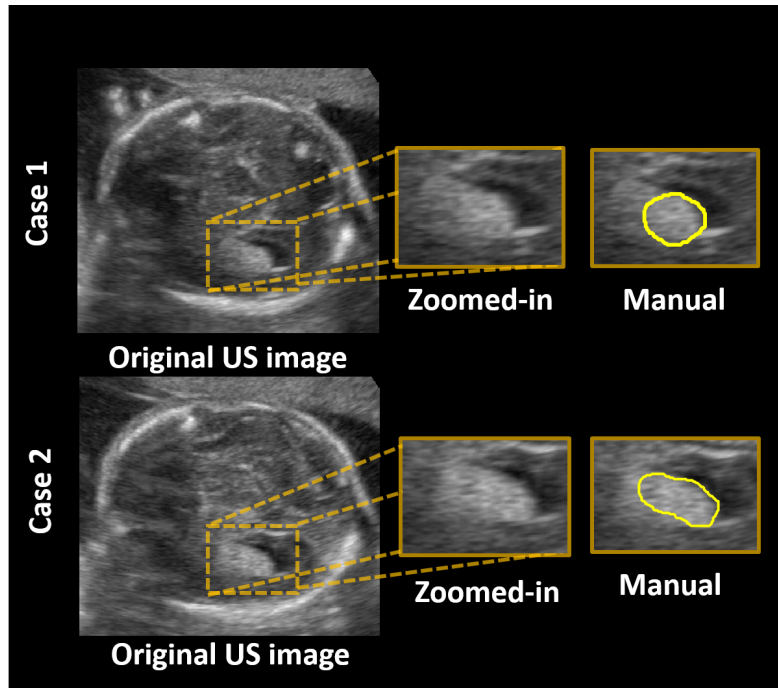


Figure 4.13: The two failure case of CP segmentation task within the test dataset. The CP image is enlarged for closer examination. The manual delineation is plotted in yellow.

region selection.

4.7.2 Failure cases analysis

This section reports two interesting cases of CP segmentation in Fig 4.13 where the proposed framework failed. It can be seen that the CP and surrounding tissues form a homogeneous bright region in both cases. Thus, the algorithm is not able to detect a suitable candidate region. We note that the enlarged CP area (according to the manual delineations) of the two cases have a substantially similar appearance, yet the same human observer delineates them quite differently (a circular outline versus a elliptical outline). The reason for the disagreement between the two manual delineations can be explained by the relatively poor contrast of the SOI with respect to the background and nearly half of the boundary delineation relied on subjective assumptions. This observation, however, echoes the previous view that manual segmentation for this task is both hard and subjectivity.

4.8 Summary

This chapter has presented a general method for automatic brain structure segmentation for 2D fetal ultrasound images. In particular, the novel shape descriptor was designed to be invariant to intensity, scale and rotation and to also capture shape characteristics. The resulting segmentations were found to match manual delineations with high accuracy. Experiments on CC and CP segmentation targeted at different image settings (mid-sagittal versus trans-thalamic plane) of the developing brain. The results demonstrated that the method is generalizable to different brain structures without the need to develop a tool for each of them individually. The method performed well on clinical data over a broad gestational age range, and is fully automatic. The experiments compared three classifiers and their performance on imbalanced datasets. The difference between the prediction accuracy of RUSBoost and the rest were even greater with the more imbalanced dataset. To further validate the method, the work also compared biometric measurements of the target structures derived by automated analysis and manual measurement. The results showed that the automated vs. manual difference was similar to inter-observer variation. This chapter also included an interesting discussion of several challenging cases which revealed some characteristics of structures in 2D ultrasound images, which may limit the practicability of boundary delineation. The next chapter will explore the localization of multiple brain structures in 3D fetal neurosonography.

5

Localizing Multiple Brain Structures in 3D Ultrasound Images

Contents

5.1	Introduction	74
5.2	View-based Projection Network (VP-Net)	78
5.2.1	Preprocessing	78
5.2.2	Network architecture design	78
5.2.3	Detection as Segmentation	83
5.2.4	Visualizing the 3D Saliency Volume	85
5.2.5	Implementation details	85
5.3	Experiments	86
5.4	Results and Discussion	90
5.4.1	Qualitative Evaluation	90
5.4.2	Quantitative evaluations	91
5.4.3	Saliency Volume	98
5.4.4	Challenging cases	98
5.5	Conclusions	101

The recent five years have witnessed the rising of deep learning technology. The advantage of a convolution neural networks (CNNs) model is it can learn end-to-end from the data, avoiding the need of hand-crafting features. Nevertheless, it faces some practical limitations, for example, the memory consumption of training a sophisticate 3D CNN model is unaffordable even on modern GPUs for high-resolution 3D images. This chapter proposes a new network architecture: View-

based Projection Networks (VP-Nets) that is specially designed for localization of structures in 3D images. Inspired by how sonographers view and interact with volumes during real-world clinical scanning, the model uses geometric projection to reduce the dimensionality of the task. As a result, the VP-Nets allows full-resolution predictions with economical memory cost. The model is tested on localizing five key brain structures in 3D fetal neurosonography. Comparison experiments with other established methods, (Random Forest and 3D U-Nets) are reported. This chapter also presents a comparison of loss functions, the effect of model depth and feature channels. The work presented here has been submitted to:

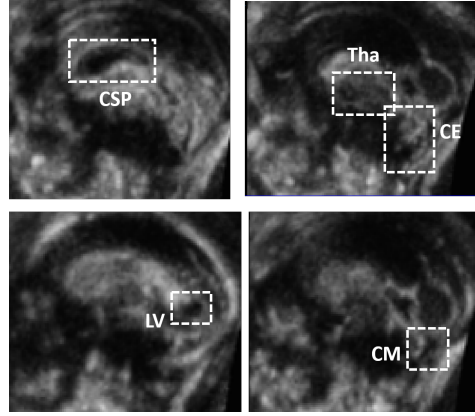
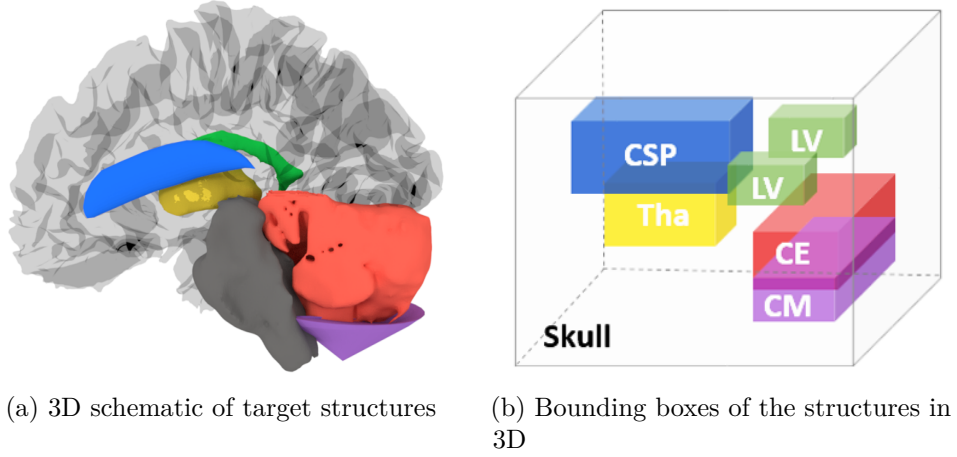
Ruobing Huang, Weidi Xie, J Alison Noble. "VP-Nets : Efficient Automatic Localization of Key Brain Structures in 3D Fetal Neurosonography". Medical Image Analysis (under second review).

In a routine ultrasound examination, multiple key brain structures should be identified in a fetal brain. This chapter aims to automate this process by building tools to localize these structures in a fetal neurosonography.

5.1 Introduction

Recall from Chapter 1, a standard fetal neurosonography examination involves identification and evaluation of the following key brain anatomies: the lateral ventricles (LV), cavum septi pellucidi (CSP), thalami (Tha), cerebellum (CE), and cisterna magna (CM) (see Fig.5.1(a) and 5.1(c)) [3]. Chapter 2 explained why localizing these structures in 3D ultrasound is challenging due to variable image quality, varying fetal brain positions and complex anatomical relationships. An automatic method to localize brain structures across a large range of gestational age (GA) is desirable to shorten the time to interpret 3D scans, e.g. assist the diagnosis of neurological conditions [124–126], assessment of fetal growth [5, 12], and gestational age estimation [8, 127].

Object localization is essentially finding a mapping from a whole 3D volume (e.g. $100 \times 100 \times 100 = 10^6$ voxels) to several digits (which denote the bounding box of the target). It is therefore a high-level abstraction task. Researchers have



(c) Structures shown in the US images

Figure 5.1: (a) Schematic of the detected structures in the right-hemisphere of the brain in 3D. Each structure is plotted as: cavum septi pellucidi (CSP) in blue, thalami (Tha) in yellow, lateral ventricles (LV) in green, cerebellum (CE) in red and cisterna magna (CM) in purple. (b) 3D spatial configuration of the bounding boxes of the target brain structures. Each box is shown as a 3D cube whose color is in accord with that in sub-figure a. Notice that there is geometric overlap between CE and CM, as well as CSP and Tha. (c) Targeted structures in sagittal planes of a US volume. The structures are bounded by white dashed boxes.

designed specialized features and complex processing systems [49, 64, 69]. Some works learn the mapping from feature representations to raw coordinates of the targeted bounding boxes [49, 64, 69]. Other works have proposed to solve structure localization in individual 2D slices [65–68], but may lose contextual information from the third dimension (for a detailed discussion refer to Chapter 2).

Volumetric fully convolutional networks (FCNs) have been proposed for **segmentation** problems in order to capture contextual information in 3D space and

have achieved good performance on MRI volumes [128, 129]. However, these models can have large memory footprints which makes it unfeasible to train a deconvolution decoder for high-resolution 3D outputs [130]. In practice, it is found that resolutions higher than 64^3 voxels bring memory issues even on a current state-of-the-art Titan X GPU (12 GB).

Aiming to incorporate merits from previous methods, an original framework **View-based Projection Networks (VP-Nets)** is proposed here to detect multiple brain structures simultaneously in 3D fetal neurosonography. Figure 5.2 presents a schematic pipeline of the proposed solution. It uses a multi-stream Convolutional Neural Networks (CNNs) to incorporate information from different anatomical views. The main contributions are:

- An end-to-end approach to detect and localize key structures in 3D fetal neurosonography. The networks are trained with bounding box masks. To the best of my knowledge, this is the first work to achieve this goal with Fully Convolutional Networks (FCNs).
- Economical time and memory consumption. Inference from a 3D input often imposes time and memory issues on CNN models. This approach, however, uses projection to reduce the dimensionality and complexity of the task, thus enable 3D outputs with higher resolution than existing 3D CNN models.
- Multi-class prediction of challenging 3D volumes. In 3D fetal neurosonography, the target structures often have different local appearance (Figure 5.1(c)) and exhibit complex spatial configurations (Figure 5.1(b)). Some structures are closely adjacent, with potentially overlapping bounding boxes (e.g. CE and CM in Figure 5.1). This further complicates the process of identifying the correct label of each pixel. VP-Nets predicts different structures separately in the last layer but sharing features in all previous layers. In this way, the prediction of different structures are not competing with each other, while interactions between them are still kept at the same time.

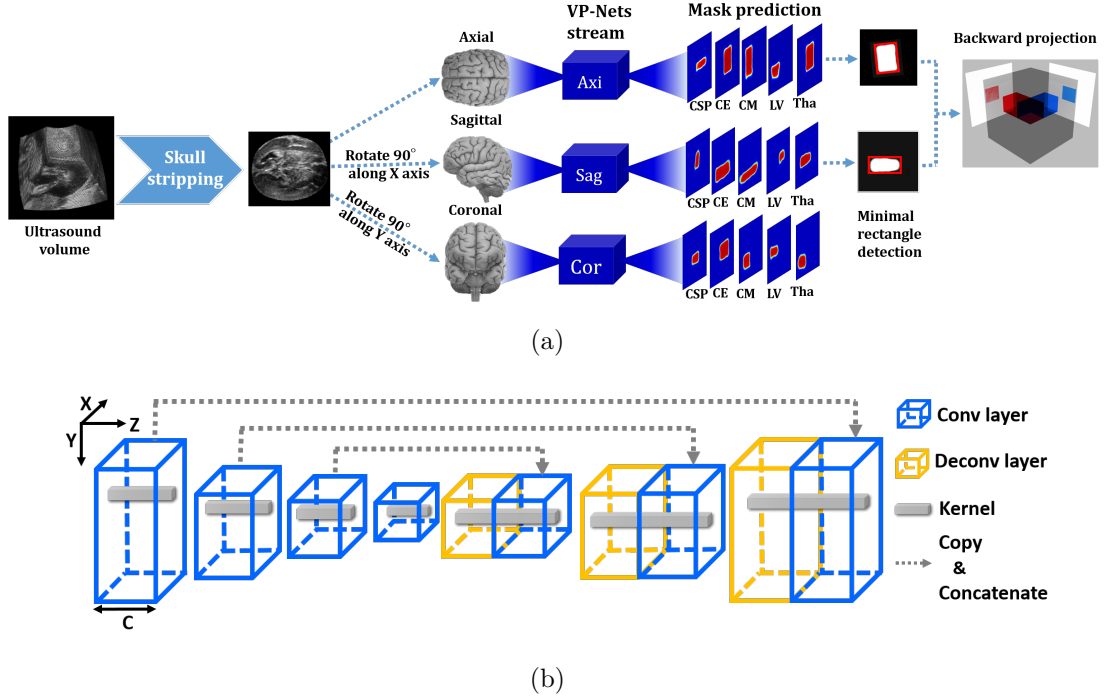


Figure 5.2: Networks design details: (a) An overview of the CNN-based image analysis pipeline. A raw ultrasound volume is preprocessed using the skull-stripping tool (refer to Section 5.2.1). The masked brain image and two of its transformed (90° rotation) volume are passed, in parallel, to three independent VP-Nets. Each VP-Net outputs the projection mask of five structures (15 projected masks for 5 structures). The minimal rectangle that encloses the prediction silhouette is detected for each structure in each view, to generate the bounding box parameters. The 3D bounding box for a structure can be reconstructed with two obtained rectangle masks with backward projection. **A graphical demo for the VP-Net model is available at: <https://youtu.be/KVxkbqWYWxc>.** (b) Each individual VP-Net shares the same spirit as 2D U-Net [122] for full-resolution prediction. In VP-Net, convolution filters in all layers only scan along the 2D X-Y plane, while always penetrating along the Z-axis to capture contextual information along that dimension.

Section 5.2 explains the core design of the multi-stream CNN model. Section 5.2.1 introduces a skull stripping tool for preprocessing, and describes the detailed design of the individual stream of VP-Nets is given. Section 5.2.3 reinterprets the localization task as a dense segmentation problem and explains in detail how the final 3D bounding box is derived from the predicted masks. Section 5.2.4 describes a method to visualize a 3D saliency volume to show that the model has learned the 3D locations of structures from 2D projections. Comparison experiments with Random Forest (RF), 3D U-Nets, and variants of VP-Nets are presented in Section

5.3. Section 5.3 investigates the effects of different loss functions and data partition. Results are reported and discussed in Section 5.4.

5.2 View-based Projection Network (VP-Net)

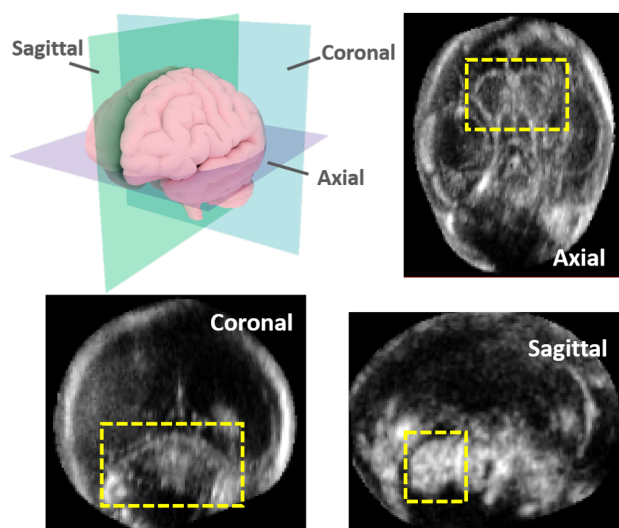
The task this chapter considers is to automatically localize five key structures in a 3D fetal US brain volumes (Figure 5.1(a)). The volumes were collected following a standard US clinical acquisition protocol. As a preprocessing step, the fetal brain is extracted using an automatic skull stripping tool described in Section 5.2.1. Then, the whole VP-Nets are trained with the target structures labelled with 3D bounding boxes by human experts (Section 5.2.2,5.2.3). During testing, the whole pipeline does not require human intervention and it can process input of different sizes as the model is fully-convolutional. Design details of the architecture are explained in subsequent sections.

5.2.1 Preprocessing

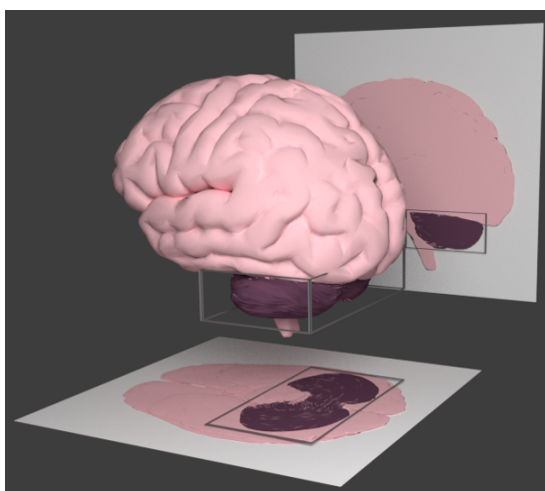
To remove the extra-cranial tissues, an automatic skull-stripping tool using a 3D U-Net model is built. Many algorithms dedicated to skull segmentation in ultrasound images have been proposed in the literature [18, 131, 132]. Here a 3D U-Net model is used, which showed good performance on similar tasks [128] (The network architecture is described in the Appendix). The model consists of contracting networks (convolution layers interwoven with pooling layers) and expansive networks (convolution layers coupled with up-sampling layers). As a result, the model accepts raw fetal brain ultrasound images as input and predicts the skull segmentation masks of the same resolution.

5.2.2 Network architecture design

The design of VP-Nets is inspired by two observations. Firstly, in clinical practice, detecting a structure in a fetal 3D US image is typically achieved by scrolling through the slices to find the largest cross-section of the structure in one of the three standard views: axial, coronal and sagittal (see Figure 5.3(a)). Then the smallest



(a)



(b)

Figure 5.3: View-based Projections.

(a) Cerebellum (CE) in different views. Yellow boxes refer to the cross section of 3D human expert annotation.

(b) Schematic of orthogonal projection. The cerebellum is highlighted in purple and bounded by a gray cube. Its orthogonal projections on axial plane (horizontal) and sagittal plane (vertical) are shown respectively. The combination of the two rectangles defines the 3D bounding box of the object.

rectangle that encloses the contour of the structure in that view is manually labelled. The same procedure is repeated for another orthogonal view. The two 2D rectangles define a cuboid which envelopes the object in 3D space with a 3D bounding box. Secondly, the orthogonal projection of a 3D object in a given direction is identical to the union set of all cross-sections along that direction (Figure 5.3(b)). In this chapter, as the considered anatomies do not possess lobes, a projection of a 3D structure is equal to its largest cross-section in the projection direction. With these two principles, the detection is solved as finding the orthogonal projection of the object.

Following these observations, the proposed framework accepts input volumes of standard anatomical views (axial, sagittal and coronal). The three volumes are passed in parallel to three identical modules. The 3D bounding box is therefore reconstructed from 2D mask predictions in the different views.

It should be noted that the proposed framework is different from other 2D or 3D CNN architectures [122, 128, 129, 133]. 2D models treat an input volume as 2D image slices: $I \in \mathbb{R}^{M \times N \times C}$, where M , N refer to spatial image dimensions (X-Y), and C denotes the number of channels (RGB or grayscale). A common way of using a 2D architecture to analyse a 3D input is to take slices from a 3D volume and process them separately. The long-range dependencies between these 2D slices is modelled by post-processing. A 3D CNN takes input volumes of $I \in \mathbb{R}^{M \times N \times D \times C}$, with the additional notation: D represents the third Z dimension of 3D data. In this case, each convolution kernel needs to scan over the whole 3D space in every layer. In contrast, each individual VP-Net (Figure 5.2(b)) shares a similar configuration with the 2D models, whereas the C are set to be the same size as the Z dimension, e.g. $C = 128$. Therefore, its deep convolution filters penetrate the whole volume along the Z-axis and only scan along the 2D X-Y plane. In this way, the model can capture 3D spatial dependencies without adding the Z dimension as used in 3D CNN. Therefore, it avoids cubically increasing memory requirements. Note that due to the penetration behaviour of the convolutional kernels, the deep kernels learn useful spatial features along the third dimension but are not expected to determine

the exact location of the object on this axis. This information is later recovered by incorporating information from the other two streams (as shown in Figure 5.2).

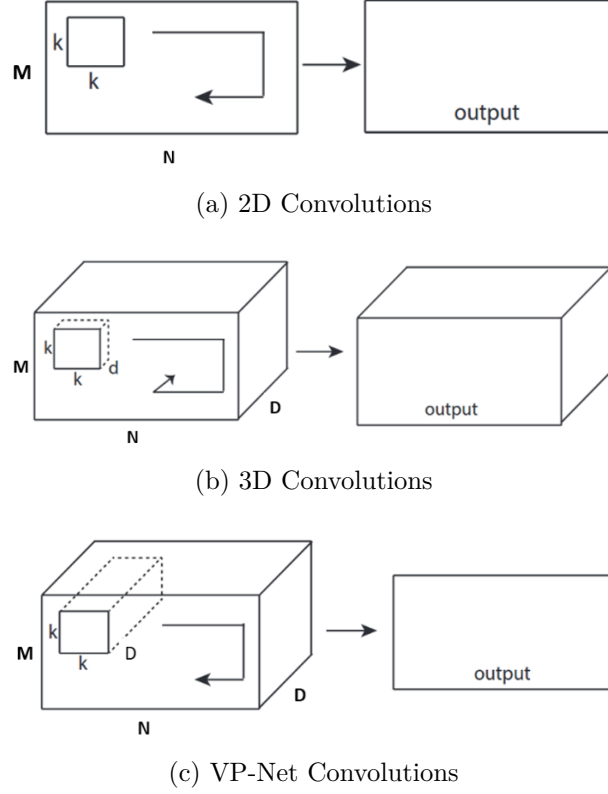


Figure 5.4: Schematics of different convolutions. M, N, D represents height, width and depth respectively. k represents the kernel size in one dimension. (a) 2D convolutions. The process is calculated in 2 direction (x,y). The output is a 2D matrix. (b) 3D convolutions. The convolution is calculated in 3 direction (x,y,z). Kernel depth d satisfies $d < D$, and is usually set to the same value as k . The output is a 3D volume. (c) VP-Net convolutions. The kernel depth is equal to the input depth D . Note that the output is also a 2D matrix.

Figure. 5.4(a),5.4(b),5.4(c) demonstrate the detail process of different types of convolution (conv). The direction of convolution computation is shown in black arrow, input depth is D , and k represents the kernel size in one dimension. The arrows showed VP-Net conv is calculated in 2 direction (x,y), while 3D conv scans 3 directions (x,y,z). Although a single VP-Net conv operation involves more values than a single 3D conv operation, the former is more computational efficient as it only involves matrix multiplication. Therefore, 3D conv takes more computation time overall (see Tab.5.3 for the run-time of 3D U-Net and VP-Nets). Given $k = 3$, the number of independent values in 2D conv is $3 \times 3 = 9$, that of the 3D conv

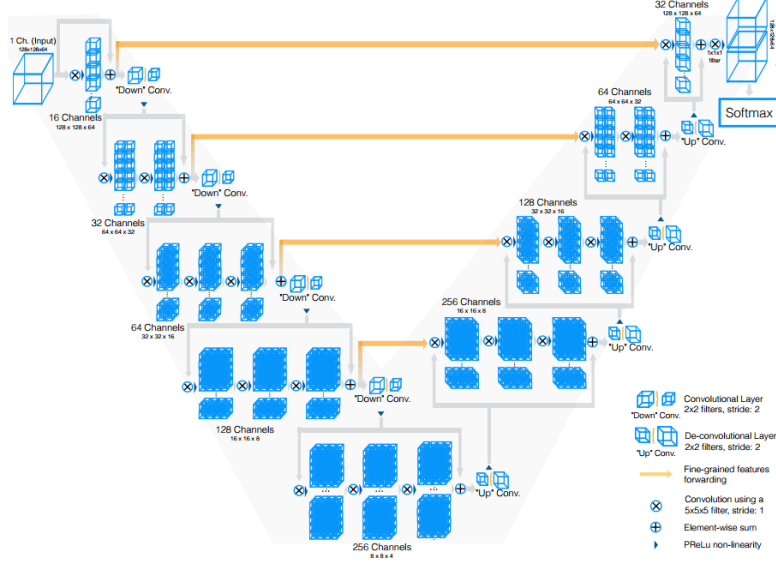


Figure 5.5: An example of 3D CNN architecture. Image source: Milletari et al. [129].

is $3 \times 3 \times 3 = 27$, and that of the VP-Net conv is $3 \times 3 \times D = 9D$. Therefore, a VP-Net model can have more parameters than a 3D CNN model, depends on the value of D . However, this number does not fully indicate the memory footprint costs during training. Note that the output of 2D conv and VP conv are both 2D matrix, while the output of 3D conv is 3D volume. Define the number of used kernel in one layer as $\hat{C}$, the output feature map of VP-Net conv is $M \times N \times \hat{C}$, while that of the 3D conv is $M \times N \times D \times \hat{C}$. Therefore, a single 3D conv layer can cost as large as 1.1GB memory with input size of $128 \times 128 \times 128$ voxel (see Appendix 2). Figure.5.5 gives an example of a 3D CNN model described in [129], it shows many intermediate feature activation **volumes** need to be stored during training. This is the main reason why 3D CNN has larger memory consumption despite it might have less model parameter. It is also interesting to point out that the widely-used VGG16 model has 138 million parameters [78], which is 3.6 times larger than that of the most sophisticate VP-Net model (38 million, see Tab. 5.3). This suggests fitting a model with large number of parameters does not necessarily lead to poor testing performance, as long as proper training paradigm and enough training data is used.

5.2.3 Detection as Segmentation

Instead of regressing raw coordinates [49, 64, 69], each VP-Net projects the desired structures to its corresponding anatomical views and **segments** the bounding-box masks of structures.

The loss function is therefore defined as :

$$E(\theta_{vp_1}, \theta_{vp_2}, \theta_{vp_3}) = \sum_{i=1}^3 \sum_{s=1}^S \sum_{MN} \mathcal{L}_{MN}^s(\theta_{vp_i}) \quad (5.1)$$

, where θ_{vp_i} refers to the parameters in the i th VP-Net ($i \in 1 \rightarrow 3$), S corresponds to the number of target structures (in our case $S = 5$) and M, N refer to volume dimensions: X and Y respectively. Sigmoid units are used for each pixel on the mask, indicating the probability of the target structure. Different loss functions are investigated in Section 5.3, e.g. traditional binary cross-entropy, dice coefficient loss.

This detection-as-segmentation design avoids fitting a highly abstract mapping function (feature representations to bounding box coordinates), and increases the foreground/background ratio (i.e. alleviates label sparsity) without further annotations. In reality, the approach increases the foreground/background ratio from 0.6% to 5.0% (3D to 2D). Since the receptive fields of different pixels correspond to different image regions, each of them captures a different variation on the appearance of the input. The dense, pixel-wise mask supervision can therefore be regarded as a sliding-window scheme, which serves as a natural way of data augmentation (same as random cropping of the input).

Finally, the mapping from the predicted silhouettes (in different views) to the final 3D bounding boxes takes two steps. First, the rectangle R that encloses the silhouette mask (denote as T) is calculated as follows:

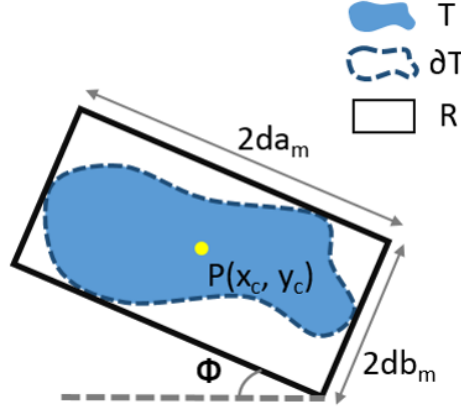


Figure 5.6: Rectangle detection denotation. The prediction mask T is plotted in solid blue, and its contour is plotted in dashed blue line. The centroid $p(x_c, y_c)$, the orientation ϕ , the width $2da_m$ and the height $2db_m$ are labelled respectively. The black box denotes the detected minimal rectangle.

1. The centroid of R : $p(x_c, y_c)$ is obtained by finding the mean of pixel position x_i, y_i along the x and y axis for all pixel $p(x_i, y_i) \in T$.
2. The contour of T is extracted as ∂T by calculating the gradient of T .
3. The ellipse that best fits the contour ∂T is detected using the least square fitting. The orientation ϕ of the ellipse (the angle between the major axis and the image axis) is recorded.
4. For every point $p(x_j, y_j) \in \partial T$, the distance between it and the centroid $p(x_c, y_c)$ along the direction of ϕ and that in the perpendicular direction are recorded as da_j and db_j respectively.
5. The maximal value of $da_j \in da$ and $db_j \in db$ are selected as da_m, db_m .
6. The rectangle R is defined by $p(x_c, y_c)$, ϕ , $2da_m$, and $2db_m$ which correspond to the centroid, orientation, the width and height of the R respectively (as shown in Figure 5.6).

A unique rectangle mask can be created with the obtained parameters and the process is repeated for three anatomical views. A 3D bounding box can be found thorough backward projection using the rectangle mask from any two views (see Figure 5.7). This could be simply obtained by: a) stacking each mask N_v times along the corresponding view to create a view-based projection volume P_v , where N_v is the depth of 3D volume in the view $v \in V$; b) The intersection volume In of two different P_v (e.g. P_a from axial view and P_c from coronal view) can be extracted by a simple summation followed by thresholding. In practice, these two steps only involve basic matrix manipulations thus the backward-projection can be efficiently implemented with a few lines of code.

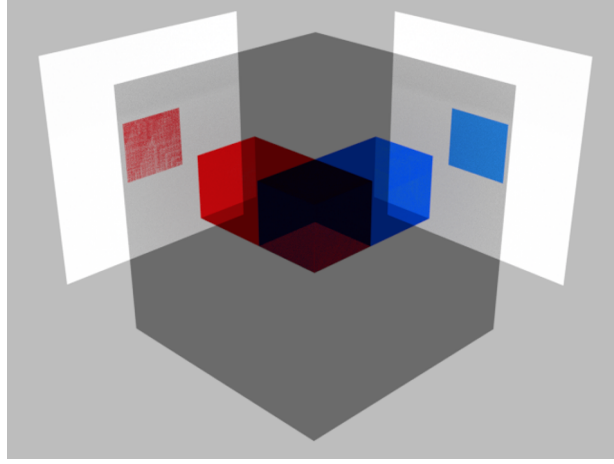


Figure 5.7: Back-ward projection. The red and blue are rectangle masks detected from two different views. The 3D bounding box can be obtained by intersection of their projection (dark purple). Note that here a rectangle mask is not stacked along the whole volume (as described in text) for a better visualisation of the intersection region.

5.2.4 Visualizing the 3D Saliency Volume

The visualization of a 3D saliency volume allows one to see what deep filters have learnt along the third dimension, with only 2D supervision (X-Y axis masks) during training. Formally, each stream of one VP-Net can be formulated as:

$$P = f(I; \theta) \quad (5.2)$$

where I is the input volume, and $P \in R^{128 \times 128 \times S}$ is the output predicted masks of the target structures, θ represents the learned weights, S is the number of target structures and f is a non-linear function mapping from input volume to prediction. Therefore, by perturbing the predicted score of a **single pixel** p_i ($p_i \in P$), $\frac{\partial p_i}{\partial I}$ is calculated as the saliency volume for p_i (adapted from [120]). Intuitively, the resulting saliency volume indicates the voxels in the input volume with the largest influence on output prediction.

5.2.5 Implementation details

The same training strategy is applied for all tested networks (except for batch size). The training was done end-to-end simultaneously via RMSProp [134] from scratch. The initial learning rate was initialized as 10^{-3} , and further decayed by a factor

of 10 every 15 epochs. The weights were initialized randomly, batch-normalization was used after each convolutional layer, followed by ReLU as non-linearity. Max-pooling was used to down-sample the feature map. Data augmentation included: (1) random cropping and rescaling with a factor of $s = 0.7$, (2) rotation of an angle randomly selected within $\theta = \pm 20^\circ$, (3) random volume-wise flip i.e. reflection of the slices with a probability of $p = 0.5$.

5.3 Experiments

Datasets The dataset includes 285 3D US fetal brain images acquired following standard clinical protocol of collecting trans-abdominal fetal neurosonograms [99]. The volumes were randomly selected from scans of different healthy fetuses age range between 20 and 29 gestational weeks. All the scans were collected using a Philips HD-9 ultrasound machine with curvilinear abdominal transducers C5-2, C6-3, V7-3. The typical size of each volume was $180 \times 200 \times 170$ voxels, with isotropic voxels measuring $(0.6 \times 0.6 \times 0.6 \text{ mm}^3)$ in size. Volumes were divided into a training (200 volumes), validation (40 volumes), and testing set (45 volumes). Within the dataset, 285 CSP, 263 LV, 285 Tha, 266 CE, and 249 CM were visible based on visual assessment of the 3D volumes. Note that some structures are not visible due to the presence of acoustic shadows or inferior image quality.

Skull segmentation accuracy The only pre-processing step the proposed model requires is skull-stripping. A 3D U-Net model was trained based on the manual delineated skull masks. Table.5.1 reports the accuracy of skull segmentation. The results show that the skull-stripping tool localizes and segments the skull accurately from the raw input image. The tool achieved a 3D IOU of $86.5 \pm 5.1\%$, which indicates good overlap between the prediction and the ground truth in all dimensions. During test time, it takes 0.23s to process one single volume. As part of the whole pipeline, this automatic tool provides a fast and accurate way to prepare the raw images for the main task: structure detection.

Table 5.1: Skull segmentation accuracy. The predicted skull mask and its corresponding ground truth are compared in: 1) the distance between their centers , 2) the difference between their scales (average across three dimensions), and 3) the 3D Intersection over Union (IOU) between the two.

Skull size(mm)	Cen Dev(mm)	Sca Dev(mm)	3D IOU (%)
51.2 ± 12.5	1.6 ± 1.0	2.1 ± 1.8	86.5 ± 5.1

Proposed VP-Nets Architectures In total, 6 VP-Nets architectures were implemented. They varied in two ways: depth $l = 7$ or 9 layers, and number of convolution kernels in every layer ($C = 64, 128$ or 256). Each stream of VP-Nets starts with several convolutional and batch-norm blocks which were coupled with a max-pooling layer to reduce feature map resolution to obtain high-level features. Then a sequence of up-sampling (de-convolution) layers were added subsequently, each of which is concatenated with the corresponds features from the contracting path to retain high-resolution information. For simplicity, the same architecture was applied for each of the three individual streams. Note that different streams do not share parameters, thus different view-based representations can be learned. The number of filters was kept unchanged in every layer throughout the whole network. Our most sophisticated model Network F has 9 ConvBlocks and 256 convolutional filters for each convolutional layer, in total it costs 1.2G memory. Details of the network architecture and memory consumption can be found in Table 6 in the Appendix. Comparing with the total memory consumption of the 3D U-Net, Network F only takes a quarter of the former.

Biometric Measurements The diameter of CE and the width of CM is of particular clinical interest to help evaluate fetal maturation and to diagnose congenital diseases [135–137]. Using the predicted masks, automatic measurement of biometry can be derived. The accuracy of these clinically-relevant measurements is investigated by comparing them to results obtained from manual measurements by a human expert.

Baseline 1: Regression Random Forest A 3D detection framework is implemented based on the regression random forest algorithm [64]. In this approach, each 3D bounding box is defined by a 6 component vector (namely, $\mathbf{b} = \{b^L, b^R, b^A, b^P, b^H, b^F\}$), where each component represents the corresponding axis-aligned bounding box. Its superscripts follow the radiological convention as: L = left, R = right, A = anterior, P = posterior, H = head and F = foot. Note that this method was not designed for different orientations. The model aimed to find the function mapping $\mathbf{b} = f(\mathbf{Q})$, where $\mathbf{Q} \in R^{D \times 1}$ is the extracted Haar-like features, D refers to its dimensionality (2048 in our experiment), $\mathbf{B} \in R^{6N \times 1}$ corresponds to the coordinates of the bounding boxes for N structures. In this experiments, 25 trees is used, with a tree depth of 8. For consistency, a similar data augmentation technique was used in this experiment.

Baseline 2: 3D U-Net A 3D U-Net is built with input $\mathbf{I} \in R^{128 \times 128 \times 128 \times 1}$, and a 3D bounding box prediction: $\mathbf{P} \in R^{64 \times 64 \times 64 \times 5}$ (the largest resolution allowed by the available GPU). Overall the 3D U-Net consists of four convolution $\wedge$ pooling layers and four up-sampling $\wedge$ concatenated layers. Details of the network architecture and memory consumption can be found in the Appendix (Table 2). Although only 64 convolutional kernels are applied for all layers, the memory cost for a single input volume during the feed-forward process amounts to 2.1 G (without considering the memory consumption for parameters). Theoretically, during training, a naive back-propagation process (without special optimization) should take the same amount of memory consumption as feed-forward propagation [138], i.e. 4.2G in total for one 3D volume. Thus, even to predict at half the resolution of the original input, a state-of-the-art GPU (Titan X, 12G) can have problems in training networks with a batch size larger than 2.

Evaluation Metrics For all experiments, structure detection was evaluated using Intersection Over Union (Jaccard index) to evaluate the overlap of bounding boxes

between ground truth and prediction:

$$J_b(\mathbb{C}_g, \mathbb{C}_p) = \frac{\|\mathbb{C}_g \cap \mathbb{C}_p\|}{\|\mathbb{C}_g \cup \mathbb{C}_p\|} \quad (5.3)$$

, where $\mathbb{C}_g$ and $\mathbb{C}_p$ refer to the ground-truth and prediction respectively, and $\|\mathbb{C}_g \cup \mathbb{C}_p\| = \|\mathbb{C}_g\| + \|\mathbb{C}_p\| - \|\mathbb{C}_g \cap \mathbb{C}_p\|$.

Cross Validation To further estimate how the proposed model will generalize to the dataset, 5-fold cross-validation was applied to the VP-Net F. The dataset is partitioned equally into 5 sub-samples (each contains 57 subjects). Each cross-validation model is trained using the same hyper-parameter and data augmentation techniques as described in Section 5.2.5.

Loss function Apart from the traditional binary-cross entropy as a baseline experiment, the effect of different loss functions is also evaluated. In medical image processing, it is common that the anatomy of interest only occupies a small portion of the whole image. As a result, the loss is trapped in local minimal and the networks tend to yield negative predictions. The VP-Nets model has already increased the foreground/ background ratio 8 times through projection (approximately from 0.6% to 5%). However, the value of this ratio (5%) is still relatively low. To tackle the class imbalance problem, several works used customized loss functions based on a dice coefficient or IOU metric [129, 139]. Here we explore whether the detection accuracy can be further improved with a dice loss function, defined as:

$$\mathcal{L} = 1 - \frac{2 \sum_{i=1}^N \mathbb{C}_{gi} \mathbb{C}_{pi} + \epsilon}{\sum_{i=1}^N \mathbb{C}_{gi}^2 + \sum_{i=1}^N \mathbb{C}_{pi}^2 + \epsilon} \quad (5.4)$$

, where N is the number of samples, $\mathbb{C}_{gi} \in \mathbb{C}_g$ denotes the i th ground truth sample, $\mathbb{C}_{pi} \in \mathbb{C}_p$ denotes the i th prediction sample and ϵ is a small number to avoid division of zero. The gradient of the loss can be computed w.r.t. $\mathbb{C}_{pj}$:

$$\frac{\partial \mathcal{L}}{\partial \mathbb{C}_{pj}} = 2 \frac{2 \mathbb{C}_{pj} \sum_{i=1}^N \mathbb{C}_{gi} \mathbb{C}_{pi} - \mathbb{C}_{gj} (\sum_{i=1}^N \mathbb{C}_{gi}^2 + \sum_{i=1}^N \mathbb{C}_{pi}^2)}{(\sum_{i=1}^N \mathbb{C}_{gi}^2 + \sum_{i=1}^N \mathbb{C}_{pi}^2)^2}, \quad (5.5)$$

without considering the effects of ϵ . To compare the impact of this loss function, another model is trained with the same VP-Nets F architecture using the dice

coefficient loss is trained. The same dataset, optimizer and data-augmentation techniques were employed.

5.4 Results and Discussion

This section reports accuracy results on targeted structures such as center deviation error, size difference error, biometry error and 2D & 3D IOU, where 2D IOU is computed for a cross-sectional plane only, and 3D IOU is calculated over a volume.

5.4.1 Qualitative Evaluation

Qualitative results are shown in Figure 5.8, where 2D bounding boxes for individual structures are drawn on cross-sections of different views. The ground truth is shown in a yellow dashed box while the prediction is shown in a red dashed box. The figure shows examples of challenging anatomical configurations present in fetal neurosonography, which display varying contrast and orientation. During data augmentation, part of the brain is cropped. All images are resized to the same size for visualization purposes. It can be seen that the VP-Nets and 3D U-Nets outperform RFs in both finding the location and determining the scale of structures. It should be pointed out that the same data-augmentation technique was used for both RF and CNN experiments. This shows that the intrinsic limitation of features presented in [64], which are not orientation invariant, as they were designed for CT scans where subjects present in identical orientation. On the contrary, although rotation-invariant is not built-in characteristics of CNN, it can learn hierarchical representations to cope with this variation given enough training images or suitable data-augmentation. It can reduce the necessity to design bespoke rotation-invariant features for ultrasound images. Figure 5.8(d) shows that the LV comprises bright and dark blobs in the axial plane. Both VP-Nets and 3D U-Nets picked out this unique pattern and were able to identify the LV quite well. However, as its boundary is ambiguous in the sagittal view, both models correlate well with the ground truth. This phenomenon coincides with typical human observation. A human expert usually finds one structure easier to identify in one or two views. This phenomenon

is affected by the way in which the ultrasound beam interacts with different tissues within the fetal brain. It suggests that the imaging protocol may need to be adapted if a specific structure is of particular interest. That is, the image may be required from the sagittal view to diagnose congenital disease related to CSP.

Although these examples are drawn from different test subjects, one can still identify the overlap between CE, CM and Tha (Figure 5.8(a), 5.8(b) and 5.8(e)). This means some voxels cannot be assigned with a single label which is an additional challenge posed by our task. This is similar to attribute learning in computer vision, where multiple visual nouns and adjectives are learned to describe one object [140]. By predicting different structures separately, the proposed network avoids the ambiguity of the label of each pixel/voxel. Meanwhile, this strategy ensures that a large number of features can be shared within one anatomical view which simplifies the whole system as well.

5.4.2 Quantitative evaluations

Detection accuracy of VP-Net-F The accuracy result of our best VP-Net model: VP-Net-F, is first reported in Table 5.2. From the reported average length of each structure, it can be seen that different structures vary in size and each structure has a large intra-class variation. As shown in Table 5.2, our model copes well with this challenge, as the size prediction error is only 2.2 ± 1.8 mm (average across different structures). The table also reports 3D IOU between prediction and annotation bounding boxes. The model achieved a 3D IOU larger than 62% in average. The CE detection task achieved the highest accuracy. One conjecture is it resulted from its distinct shape and relatively large size. In contrast, the LV scored the lowest accuracy among the five structures. It may be interpreted as the size is considerably smaller compared with other structures, which leads to a more imbalanced class.

As a side-product, the VP-net provides automatic fetal brain structure biometry. On average, the automatic measurements correlated well with the manual ones. The reported deviation is less than 2.0 mm.

Table 5.2: Results of **VP-Nets F**. The largest diameter of each structures is given in the second column for comparison. The prediction and ground truth (GT) are compared in terms of the center deviation (Cen Dev), scale deviation (Sca Dev) and 3D IOU. The automatically generated measurements of CE and CM is also reported.

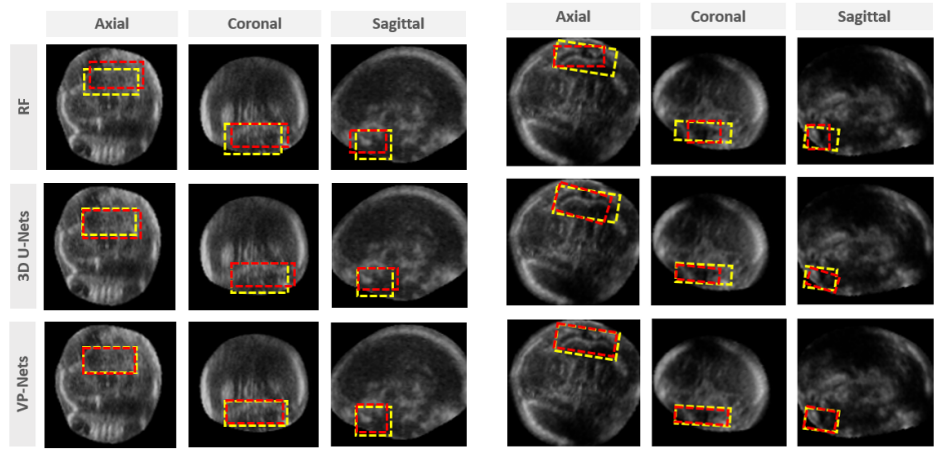
Structure	GT (mm)	Cen Dev(mm)	Sca Dev(mm)	3D IoU (%)	Biometry Err(mm)
CSP	31.4 ± 5.8	1.8 ± 1.5	2.2 ± 1.8	63.2 ± 15.2	–
LV	9.0 ± 2.5	2.4 ± 2.0	1.5 ± 1.4	48.7 ± 19.1	–
Tha	25.8 ± 3.8	1.9 ± 1.7	2.1 ± 1.9	68.1 ± 14.7	–
CE	26.9 ± 14.0	2.0 ± 1.9	2.4 ± 2.2	66.5 ± 14.6	1.5 ± 0.7
CM	16.2 ± 5.9	2.1 ± 1.9	2.3 ± 2.1	65.9 ± 15.3	2.0 ± 0.9

Comparison of different models Table 5.3 reports quantitative results for 3D U-Net and the 6 VP-Nets (VP-Net-A to VP-Net-F). The results show that 3D U-Net and VP-Nets all outperform RF and are capable of dealing with multiple structures with varying shapes and scales. The results show the benefit of the end-to-end trainable model over hand-crafted features. Interestingly, the RF model is accurate in predicting the scale of the bounding boxes while it scores poorly in the 2D and 3D IOU metrics as the localization predictions are slightly deviated from the ground truth.

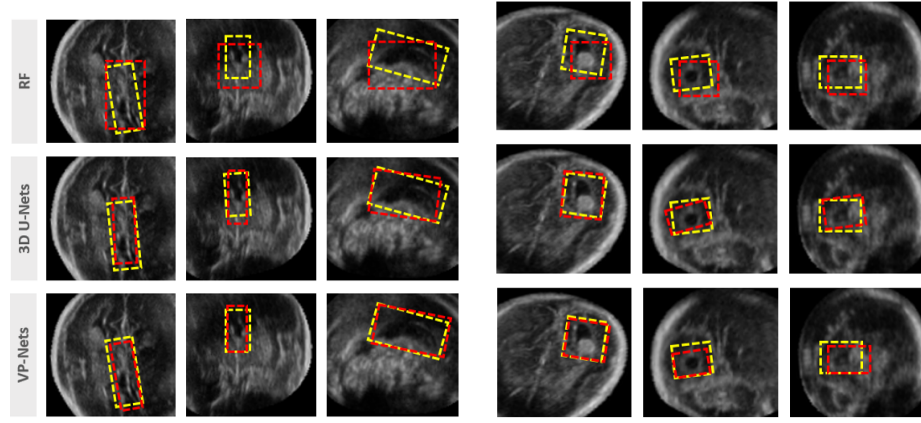
VP-Net-F achieved the highest accuracy among all calculated metrics. It may benefit from its large model capacity, i.e. the largest number of trainable parameters compare to its counterparts. Its performance may be further enhanced by increasing the model size given enough training data.

VP-Nets scored higher accuracy given a model with similar number of parameters. For example, VP-Net-E showed higher accuracy than 3D U-Nets in both center and scale detection (Table 5.3). It may suggest that by decreasing label sparsity using projection, the VP-Net model deals better with the over-fitting.

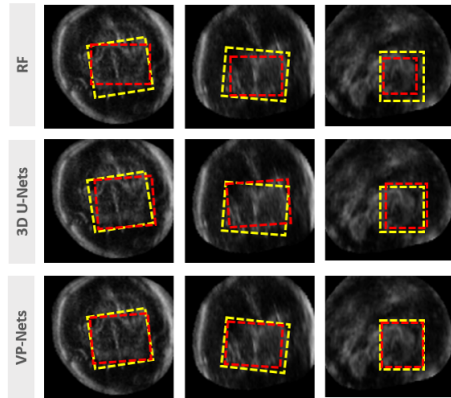
All detection methods achieve the best performance on the axial view. This observation echoes with the clinical observation that structures are easier to identify on axial views.



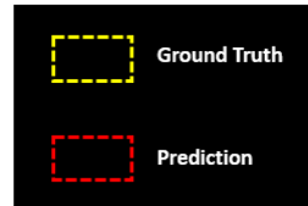
(a) Qualitative result of **CE** detection (b) Qualitative result of **CM** detection



(c) Qualitative result of **CC** detection (d) Qualitative result of **LV** detection



(e) Qualitative result of **Tha** detection



(f) Legend

Figure 5.8: Example of predicted bounding boxes of targeted structure in different anatomical views. Yellow box show the ground truth, and Red boxes are predictions.

First to last row: Results from Random Forest, 3D U-Net, VP-Nets.

First to last column: Results from Axial, Coronal, Sagittal views.

Note that there are overlap between CE and CM in all the views (as shown in a and b). Tha coincides with CE and CM in coronal view as well. The results show the prediction of VP-Nets coincided better with ground truth than that of 3D U-nets and random forest.

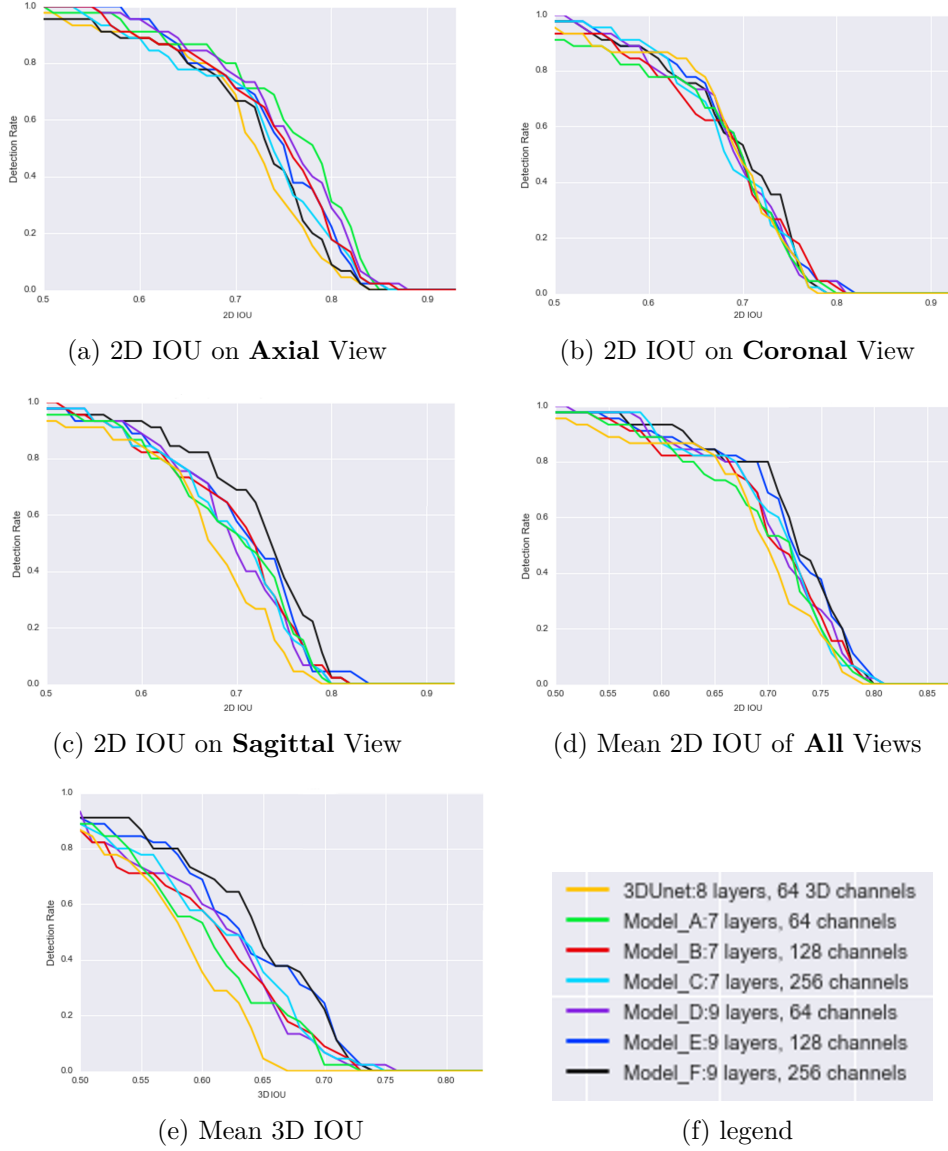


Figure 5.9: Mean 2D & 3D IOU curves of 3D U-Nets (yellow) and VP-Nets (other colors). (b) shows that the 3D U-Nets performed similarly with other VP-Nets on coronal view in 2D IOU. All VP-Nets achieved higher accuracy when combine the three views together (e). Comparing different variants of VP-Nets, Model F(black) achieved the highest accuracy. It shows that larger feature channel and deeper layers can increase the network performance.

Table 5.3: Quantitative results of different models. Column 2,3,4 list the channels, depths and the number of trainable parameters (Par) of each model. The prediction and ground truth (GT) are compared in terms of the center deviation (Cen Dev), scale deviation (Sca Dev), 2D IOU and 3D IOU. RF has the lowest 3D IOU accuracy as the predicted center locations have the largest error margin. The VP-Nets F achieves the highest accuracy in all metrics as the model has the largest model capacity (trainable parameters numbers). 3D U-Nets shows similar accuracy in mean 2D IOU deviation, but lags behind in 3D IOU. Refer to the text to see detail analysis. The last column report the runtime of different CNN models. The VP-Nets models use considerably less time than the 3D U-Nets.

Model	Channels	Depths	Par (10^6)	Cen Dev (mm)	Sca Dev (mm)	2D IOU (%)	3D IOU (%)	Run time
RF	—	—	—	6.6 ± 4.8	3.9 ± 3.6	55.6 ± 16.7	15.1 ± 14.5	
3D U-Nets	64	8	9	2.9 ± 2.7	3.1 ± 2.3	68.5 ± 9.2	56.5 ± 19.6	0.36s
VP-Nets A	64	7	2	3.0 ± 2.8	3.2 ± 2.2	67.8 ± 9.1	57.1 ± 17.1	0.04s
VP-Nets B	128	7	3	2.8 ± 2.2	3.3 ± 2.0	69.1 ± 7.9	58.6 ± 18.1	0.07s
VP-Nets C	256	7	29	2.6 ± 2.3	3.1 ± 1.8	70.0 ± 8.3	57.2 ± 17.0	0.07s
VP-Nets D	64	9	7	2.7 ± 1.9	2.9 ± 2.0	69.2 ± 8.6	59.3 ± 16.9	0.04s
VP-Nets E	128	9	9	2.5 ± 1.7	2.5 ± 1.8	70.2 ± 7.5	60.8 ± 17.5	0.08s
VP-Nets F	256	9	38	2.0 ± 1.9	2.2 ± 1.8	71.7 ± 8.1	62.0 ± 15.8	0.08s

When comparing variants of VP-Nets, deeper networks usually lead to better performance, e.g. VP-Net-D *versus* VP-Net-A. Moreover, the number of channels plays an important role as shown by VP-Nets A, B and C. However, 3D U-Nets cannot afford a large channel size with current standard GPU capacity, this limits the model performance in practice. Interestingly, 3D U-Nets perform similarly to VP-Nets C in the mean 2D IOU metrics (Figure 5.9(d)), while its performance is lower than almost all VP-Nets in terms of 3D IOU measurements (Figure 5.9(e)). This indicates that 3D IOU is a more discriminating metric than the mean of 2D IOU metric as the former is the product of accuracy of the prediction along each orthogonal axis. In other words, a high 3D IOU score requires that the prediction is consistent with the ground truth in all three axis, while a high 2D IOU may not satisfy this condition.

This point is further investigated by reporting the 2D IOU on different views separately (as shown in Figure 5.9(a), 5.9(b), 5.9(c)). Figure 5.9(b) shows 3D U-Nets achieve comparable performance in the coronal view, while it performed inferiorly in sagittal view (Figure 5.9(c)). On the contrary, all the VP-Nets performed similarly across different views. This may be explained by the fact that the design presented by this chapter simplifies the task for VP-Nets as each stream only needs to deal with one single anatomical view, thus leading to overall superior performance. Furthermore, as the memory burden is reduced from cubic to quadratic, VP-Nets support a larger batch size and prediction with higher resolution masks (equal to the input image).

The last column of Table 5.3 lists the running time of different methods (per volume). All the VP-Nets models need substantially less time than the 3D U-Nets, which proves the previous analysis that the proposed VP-Nets model can cut down the time consumption. The fastest, VP-Nets A, only need 0.04s to process an input volume. The VP-Nets F uses 0.08s, a quarter of that of the 3D U-Nets, and achieves the best localization accuracy. This indicates the applicability of the proposed method on large datasets in the future.

Cross-validation As the dataset is randomly sampled from a large clinical database, it is interesting to see whether the partition of training and testing data will affect the model performance. A 5-fold cross-validation was carried out using the same training paradigm of the previous experiment. Table 5.4 reports the results of each fold and the average based on VP-Net-F, showing that the partition of dataset has minor effect on the model performance and the proposed method generalize well on the dataset.

Table 5.4: Cross validation results. The prediction and ground truth (GT) are compared in terms of the center deviation (Cen Dev), scale deviation (Sca Dev), 2D IOU and 3D IOU. The model performed similarly across different folds.

Fold No.	Cen Dev (mm)	Sca Dev (mm)	2D IOU (%)	3D IOU (%)
Fold 1	1.9 ± 1.7	2.1 ± 1.7	72.6 ± 7.2	62.6 ± 16.0
Fold 2	1.9 ± 1.8	2.3 ± 1.9	71.0 ± 7.6	61.3 ± 15.9
Fold 3	2.2 ± 1.9	2.5 ± 1.9	70.9 ± 8.3	60.6 ± 16.8
Fold 4	1.9 ± 1.7	2.2 ± 1.9	71.4 ± 7.9	62.3 ± 16.0
Fold 5	2.1 ± 2.0	2.4 ± 1.8	71.2 ± 8.6	61.7 ± 16.5
Average	2.0 ± 1.8	2.3 ± 1.9	71.3 ± 7.7	60.9 ± 16.1

Dice coefficient Loss Table.5.5 compares the results of models trained by different loss function, while other training hyper parameters were set to be identical. The results show that the model trained with dice coefficient loss outperformed the previous model in finding the locations and the scales of the targets. It shows that the dice loss handles the class imbalance problem relatively better, therefore it can boost the prediction accuracy by a small margin.

Table 5.5: Comparison of models trained with binary cross-entropy and dice coefficient loss. The model performance is reported as: the center deviation (Cen Dev), scale deviation (Sca Dev), 2D IOU and 3D IOU. The model trained with dice coefficient loss outperformed the model trained with binary cross-entropy.

Loss	Cen Dev (mm)	Sca Dev (mm)	2D IOU (%)	3D IOU (%)
Cross-entropy	2.0 ± 1.9	2.2 ± 1.8	71.7 ± 8.1	62.0 ± 15.8
Dice	1.8 ± 1.4	1.9 ± 1.5	74.6 ± 7.1	63.2 ± 14.7

5.4.3 Saliency Volume

An example of a 3D saliency volume derived from the CE prediction (from the test set) in the axial stream (X-Y dimension) is shown in Figure 5.10. Salient points scattered far away from the CE indicate the incorporated context information, while those clusters around the structure itself represents regional characteristics captured by the model. For example, the bow-shape edge of the CE is of particular interest (Figure 5.10a): it shows these salient points are consistently correlated with salient features of the target structure. This visualization indicates that the trained model is able to capture the most discriminant features of the target structure. *More interestingly*, the salient points (red spots) closely correlate with the target structure along the Z axis (see Figure 5.10b). It suggests that the model can pinpoint the location of the object in the third dimension Z, despite supervision given only in the X-Y plane,

5.4.4 Challenging cases

The model is investigated further by discussing three particular challenging test cases. In the first case, the original US image quality is relatively poor (Figure 5.11). Figure 5.11(a) is a schematic of the CE in the axial plane (a trans-cerebellum plane). The CE is the white eyeglass-shape structure, while CM is the dark ‘M’-shape structure¹. However, as the US image is corrupted by acoustic shadows, half of the CE is invisible (Figure 5.11(b)). The human expert had to label the structure based on subjective assumptions (overlaid as yellow box in Figure 5.11(c)). Figure 5.11(d) shows the output confidence map of the VP-Nets model in red, where a higher intensity indicates higher confidence. The prediction is more confident around the region where the CE is visible.

The second case is the one where prediction of the CSP bounding box has the largest scale deviation from the ground truth in the test set. Similar to the above example, the schematic of CSP in the mid-sagittal plane is displayed in

¹The reader can also refer to the upper-right figure in Figure 5.3(a) for a clearer visualization of the CE in the trans-cerebellum plane.

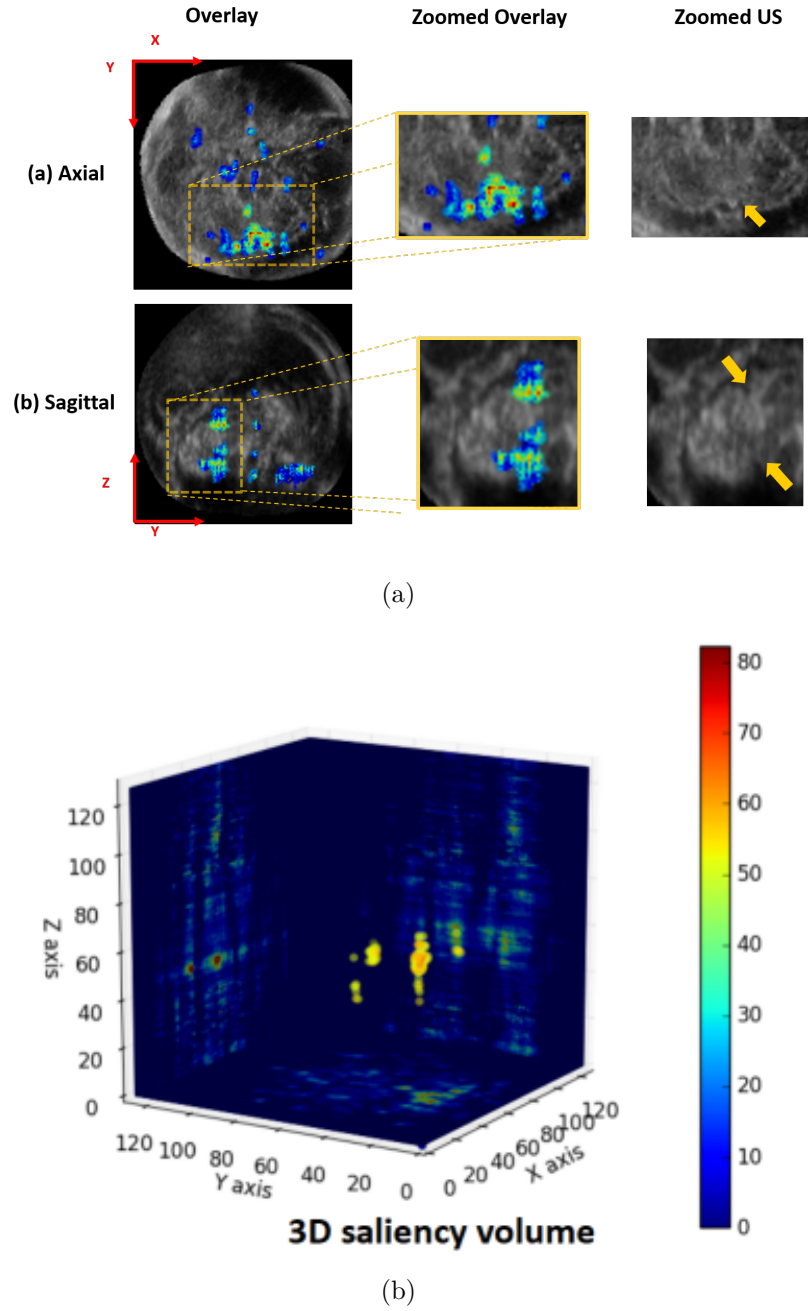


Figure 5.10: Visualization of a 3D saliency volume derived from CE detection in the axial view. The right-hand figure shows the points with top 2% gradient values in 3D. The color red represents higher gradient values, indicating a higher influence on the prediction outcome. Sub-figure (a)(b) shows a axial and a sagittal US slice with 3D saliency points projected on each view respectively. The images and the original US are zoomed in for detail visualization.

Figure 5.12(a). Compare the ground truth (Figure 5.12(c)) and the prediction (Figure 5.12(d)). They disagree mostly at the rear region of the CSP. Refer to the

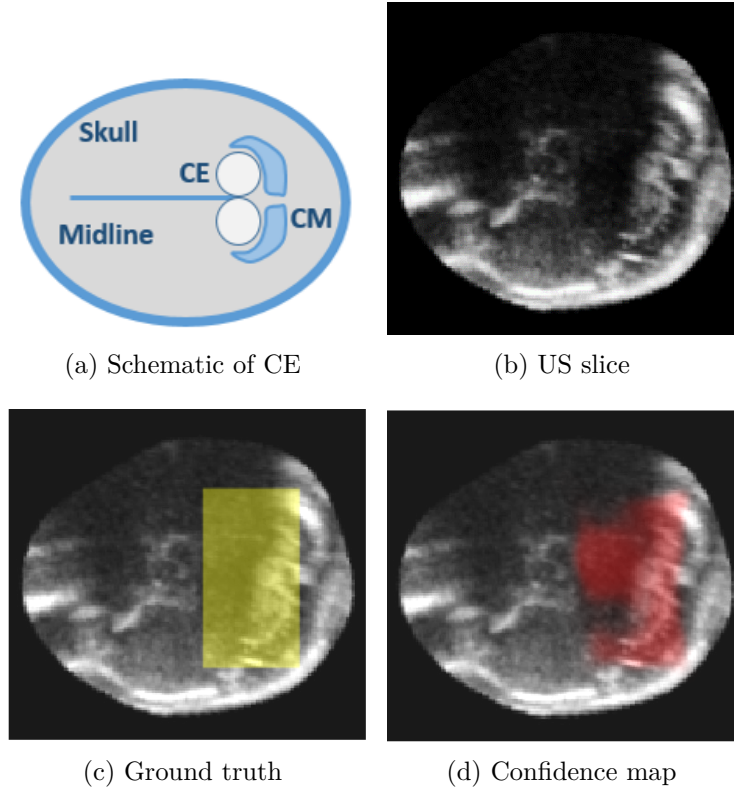


Figure 5.11: Challenging case of CE in test sets. (a) shows the schematic of anatomical configuration of CE in axial plane. The white eyeglass-shaped structure is the CE. (b) is the US slice of the discussed case. Half of the CE is blocked by acoustic shadows. (c) shows the ground truth bounding box of CE (overlaid in yellow). The confidence map predicted by VP-Nets F is shown as red in (d). In this case, IoU of the raw prediction map and the ground truth is 68%.

original US slice (Figure 5.12(b)), this may be caused by the poor visibility of the tail of the CSP, which is connected with a small dark structure.

The third case is where the human expert indicated that the structure is not present while the proposed model detected a structure (Figure 5.13(b)). Referring to the location pointed out by Figure 5.13(c), one can argue that part of the LV is visible. It is possible that the VP-Nets has captured some unique ultrasonic patterns that may be hard to interpret by human eye.

The fourth case shows the detection results where 2D IOU with the ground truth is low (45.6%, Figure 5.13(f)). Comparing Figure 5.13(e) and Figure 5.13(f), one can see the results have a high correlation. Nevertheless, as the area of the LV is relatively small, a small deviation in the scale or center prediction causes

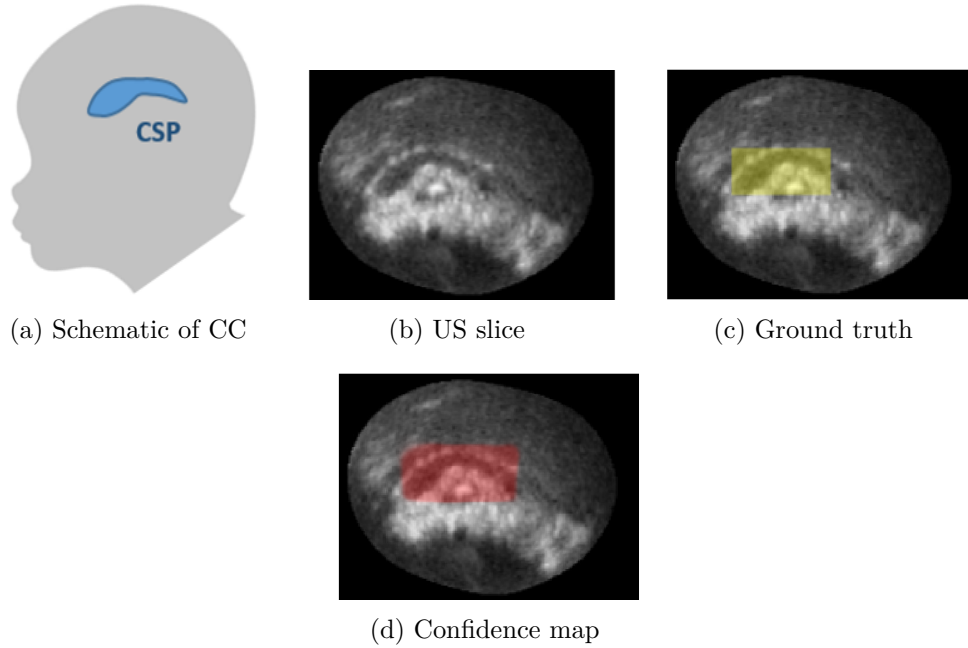


Figure 5.12: Challenging case of CC in test sets. (a) is the schematic of CC in mid-sagittal plane. The CC is shown as the blue comma-shape structure. The corresponding mid-sagittal slice of the US volume is given in (b). The ground truth is plotted as transparent yellow rectangle in (c). Similarly, the predicted confidence map is shown as red rectangle in (d).

a significant decrease in IOU estimation.

5.5 Conclusions

This chapter has presented a new efficient approach to 3D detection of multiple brain structures in fetal neurosonography through a FCN-based model: VP-Nets. The model does not require a large amount of training data and learns from coarsely annotated volumetric images (bounding-box masks). Models with different depths and feature channels were compared and the prediction accuracy was reported. Cross-validation experiments demonstrated that the proposed model generalize well on the dataset. Different loss functions were also tested, the results showed the dice-coefficient loss has beneficial effects on the model performance compared to classical binary cross-entropy. Visualization of the 3D saliency volume demonstrated that the network is capable of localizing key structure in 3D space without any supervision in the third dimension. The chapter ended with discussion of several

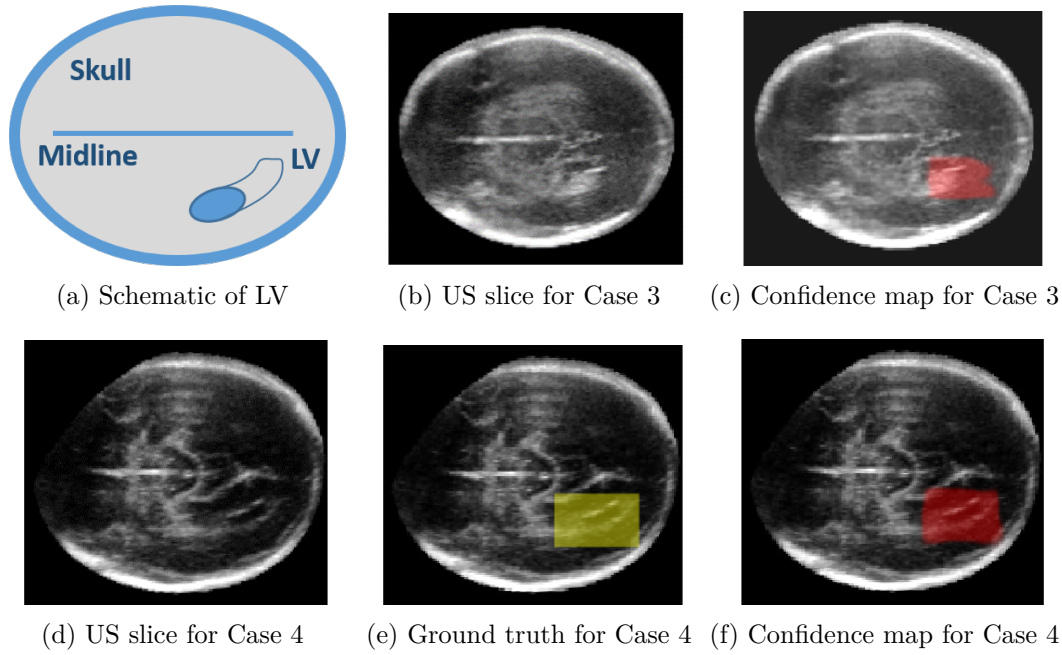


Figure 5.13: Challenging case of LV in test sets. (a) is the schematic of LV in axial plane. (b), (d) show the corresponding axial image of the third and the fourth case respectively. No ground truth is displayed for (b) as the human expert reckoned LV is not visible in this case. (e) show the ground truth label of the LV for the fourth case. The predicted confidence map is superimposed on the corresponding ultrasound image in (c) and (f).

challenging cases which revealed how the model performed when the structure visibility is limited. The developed tool can accelerate the process of routine examinations and serves as a fundamental step to process and understand large fetal neurosonography datasets in the future. The applicability of the proposed model is not limited to ultrasound images, but rather to any other 3D medical imaging modalities where detection is desirable.

6

Automatic Similarity Registration of Fetal Brain : A Pilot Study

Contents

6.1	Introduction	104
6.2	Fetal Brain Talairach (FBT) coordinate system	105
6.2.1	Fetal Corpus Callosum	107
6.2.2	Establishing the unified coordinate system	108
6.3	Fetal brain similarity registration with 3D CNNs . . .	110
6.3.1	Learning to align brains in the axial view	111
6.3.2	Learning rotations in three anatomical views	116
6.3.3	A full similarity 3D registration pipeline	123
6.4	Conclusion and future study	126

Previous chapters have considered visualization, segmentation and detection algorithms for fetal neurosonography. Regardless of their objectives, they all face one common problem: the varying positions, orientations and scales of the fetal brains within the image volume. These 9 degrees of freedom pose unique challenges in the analysis of fetal neurosonography, in contrast to MRI studies of human brains which are images in similar orientations and positions. Previous chapters (Chapter 3,4,5) have focused on crafting features or designing algorithms that are robust to these variations. Instead, this chapter describes how to reduce this effect fetal head positions. Specifically, an similarity fetal brain registration algorithm

is developed for spatial alignment of ultrasound images. The application of the developed method can map subjects into the same coordinate system therefore simplifying subsequent analysis. Section 6.2 introduces a brain coordinate system that is widely used in the MRI community: the Talairach system. Based on that, a Fetal Brain Talairach (FBT) coordinate system is proposed that is bespoke for 3D fetal neurosonography (see Section.6.2.2).

Section 6.3 tests and discusses approaches of brain registration based on deep learning. We start by describing a method to align images in one orientation and proceed to generalize the idea to a full 3D similarity registration. Experiments are designed and tested with increasing levels of difficulty. The results demonstrate the capability of 3D CNNs to perform 3D fetal ultrasound image registration. This pilot study also illustrates how neural networks behave in solving a problem that does not have a unique solution, and the importance of balanced supervision.

6.1 Introduction

Chapter 2.5 reviewed previous work on registration. Image registration algorithms can be classified into two broad categories: linear and non-rigid (or 'elastic') based on the transformation models they use. Rigid body registration and affine registration are two common forms of linear image registration. The difference between the two lies in the degrees of freedom each method allows. In 3D, the former accepts 6 degrees of freedom covering translation and rotation in 3D space. While the latter allows for up to 12 degrees of freedom, with the additional scaling and sheering factors allow changing in the size and the shape of an object. In this study, the objective of the registration method is bespoke for the task, where rotation, translation and scaling are incorporated as the fetal brain appears in different orientations, positions and sizes in clinical fetal neurosonography. Figure 6.1 shows a montage of 30 mid-axial slices of randomly selected 3D ultrasound images. Comparing the sub-image in the bottom-left of the figure with its neighbours, it can be seen the brain has large orientation variation in the axial view. The third row of the figure shows position deviation witnessed in raw images. Meanwhile, the size of brain in the middle of

the fourth row differs dramatically with its neighbours. The figure also indicates additional challenges in the ultrasound images such as intensity variations, acoustic shadows, presence of speckles as discussed in the previous chapters. However, the sheering components are excluded in this study as sheering distorts the shape of brain structures which is not a natural variation caused by a scanning procedure. It should be noted that pathological cases may include dramatic shape distortion. However, this study targets at healthy subjects which are the main concerns in fetal neurosonography examinations.

As a result, the registration method presented here aims to estimate the orientation, translation, and scaling of the fetal brain in raw 3D ultrasound images and to geometrically map them into a common coordinate system. It accepts 9 degrees of freedom (3 translation, 3 orientation and 3 scale parameters), which is a special case of an affine registration, and is called similarity registration.

6.2 Fetal Brain Talairach (FBT) coordinate system

The basic foundation of modern adult brain atlas studies is the use of a standardized 3D coordinate frame of reference. It serves as a powerful medium that can combine brain images from hundreds of subjects. This framework allows the exploration of inter-/intra-individual variance for disease diagnosis and population studies. Some coordinate systems have been established for the adult brain, such as the Talairach space [141], MNI atlases [142, 143], FreeSurfer atlases [144], etc. The Talairach coordinate system, which is also known as the stereotaxic coordinate system, is one of the most widely used systems. It was first proposed by Jean Talairach and Gabor Szijia in 1967, and is defined by two anchors: the anterior commissure (AC) and posterior commissure (PC) (see Fig.6.2). The AC and PC are both nerve fibres that connect the two hemispheres across the midline. The AC-PC line that these two anchors define is required to be horizontal, and the mid-sagittal plane is required to be vertical. The AC is set to be the origin of the system, and the y-axis points from the PC to AC, the x-axis is in the left

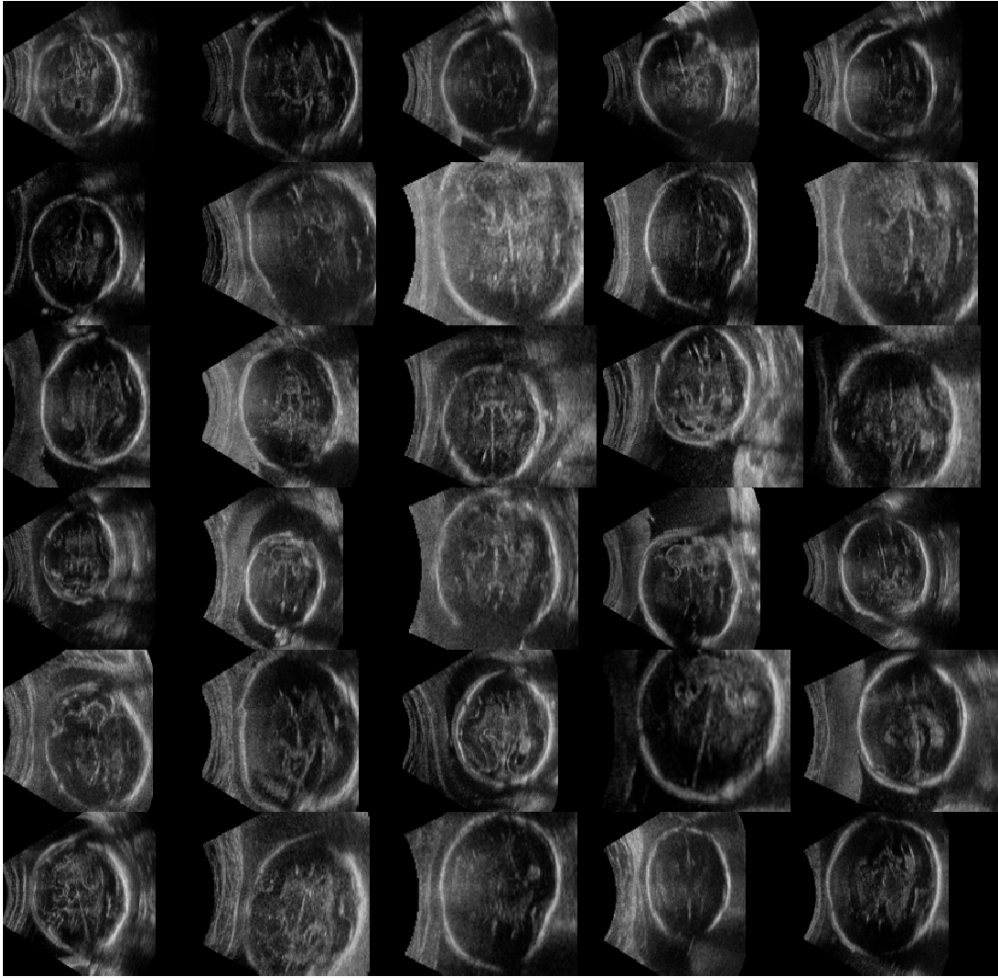


Figure 6.1: Montage of mid-axial slices from randomly selected subjects. All images is cropped in the image centre (a 150×150 pixel window) for better visualization.

and the right direction, and the z-axis is in the superior and inferior direction (up and down). Therefore, the two landmarks can define a rigid transformation that maps an individual brain image to the standard Talairach system. It can then be used to compare multiple samples to study brain development or to integrate information from different modalities (e.g. MRI and PET).

In fetal neurosonography, the Talairach transformation cannot be directly applied as the AC and PC are not visible. Instead, we propose a new coordinate system, which we call the Fetal Brain Talairach (FBT). FBT is adapted from the Talairach system but is specialized for 3D ultrasound images of the fetal brain. To the best of our knowledge, it is the first time that such system has been established for fetal neurosonography.

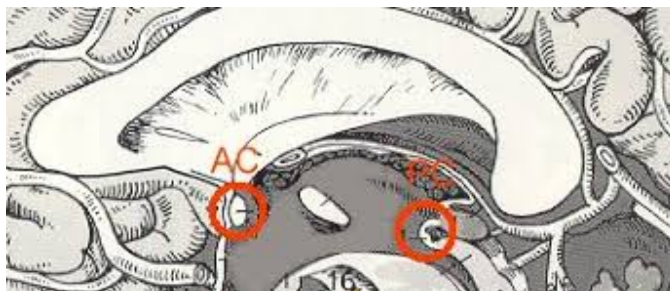


Figure 6.2: Illustration of the anterior commissure (AC) and posterior commissure (PC) in the brain. The AC and PC are marked with orange circles. Image credited to Lilla Zöllei, slides for WhyNHow seminar 12/04/08.

6.2.1 Fetal Corpus Callosum

The corpus callosum (CC) is located at the mid-sagittal plane, and is the main connector between the two cerebral hemispheres, as shown in Figure 6.3(a). In the human embryo, the earliest callosal axons appear at 74 days, the genu and the splenium are recognized at 84 days, and adult morphology is achieved by 115 days [145] (shown in Figure 6.3(b)).

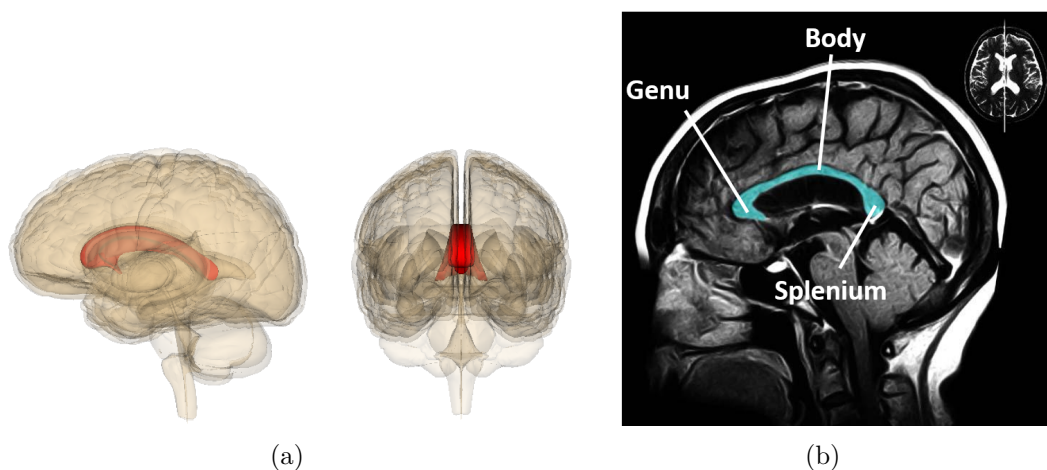


Figure 6.3: The human corpus callosum. (a) The corpus callosum locates at mid-sagittal section of the brain. Image source: Life Science Databases from Anatomography via Wikipedia. (b) The human CC consists of three parts: the genu, the body and the splenium. Image resource: Pre AP Biology Study Guide with Mrsedward at Mountain View College.

The fetal CC is considered as a sensitive biomarker of normal brain development and maturation [114]. In sonography studies, the corpus callosum is defined as the anechoic structure delimited superiorly by the sulcus of the CC and the

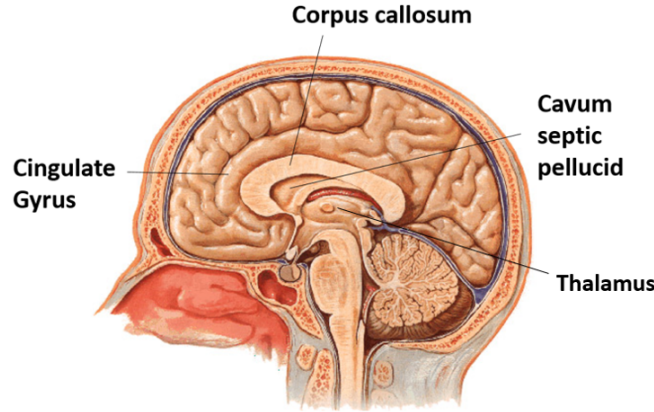


Figure 6.4: Illustration of the spatial location of the corpus callosum (CC) and the cavum septic pellucidi (CSP) in the brain. The two are closely adjacent and superior to the cingulate gyrus. The thalamus located beneath them is also labelled.

cingulate gyrus and inferiorly by the cavum septic pellucidi (CSP) [146] (their spatial relationship is shown in Fig. 6.4). In a mid-sagittal scan that is obtained directly using 2D ultrasonography, the CC is seen as a thin sonolucent¹ strip with echogenic contours overlying the CSP. However, in 3D fetal ultrasound, the CC is not clearly differentiated from the inferior CSP [147][148][149]. Instead, a thin sonolucent CC, a single comma-shaped echogenic structure including both elements (the CC and the CSP) is observed, as shown in Figure 6.5(b). Measurements of the CC-CSP complex in 2D and 3D images demonstrates a high agreement with those of the CC, and indicates that the CC can be accurately inferred from the sonolucent CC-CSP complex [147][148].

6.2.2 Establishing the unified coordinate system

The anatomical position of the CC-CSP complex is spatially similar to the AC and PC. Hence, we propose that the endpoints of this complex are chosen to be the reference landmarks for the 3D fetal brain ultrasound images. Then the line between these two endpoints defines the y-axis of the FBT (see Fig.6.6), to restrict the orientation of the brain in the axial and sagittal view. Different from the Talairach system, where the anterior landmark is set to be the origin, it might be more intuitive

¹In ultrasonography, sonolucent means containing few or no echoes.

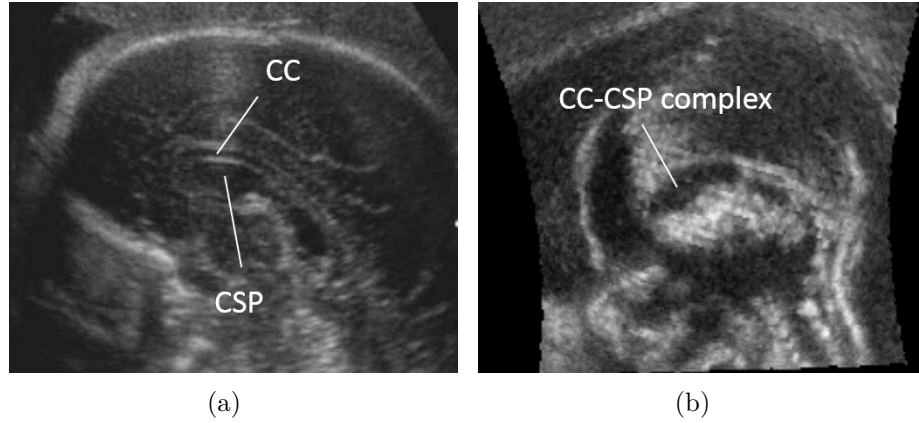


Figure 6.5: The fetal corpus callosum in ultrasound images. (a) 2D ultrasound images of fetal CC, observed as a dark thin strip. (b) 3D ultrasound image demonstrates a single comma-shaped echogenic structure representing the CC-CSP complex.

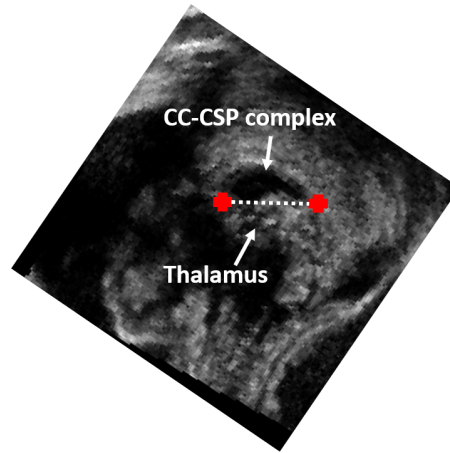


Figure 6.6: The mid-sagittal plane of a 3D fetal brain ultrasound image. The reference landmarks for the Fetal Brain Talairach (FBT) coordinate system are marked in red cross. The line between the two landmarks (showed as white dashed line) defines the y-axis. The CC-CSP complex and the thalamus are highlighted by white arrows.

to set the centre of the brain mass to be the origin. Chapter 3 described in detail how the mid-sagittal plane can be obtained in a fetal ultrasound volume. This plane is set to be vertical to the image axis, which adds a constraint on the orientation of the brain in the coronal view. Moreover, as it is not uncommon for the fetal brain to have an inverted position, an additional requirement should be included in the FBT system. The CC-CSP complex is required to be superior to the thalamus, which is a large mass of gray matter also located in the midline (see Fig.6.4).

Now, the FBT coordinates system is fully established and an individual fetal

neurosonography scan can be registered to the same space. The translation between two transformed brains is thus the distance between their respective origins. Furthermore, the scale difference between two warped subjects is defined as the difference between the size of their skulls. Figure 6.7 shows an example of a transformed fetal brain viewed from four different angles. The horizontal slice is the trans-thalamic plane, while the vertical plane is the mid-sagittal plane. The skull boundary is represented in green frames. Reference landmarks used to define the y-axis are highlighted in red. The FBT axes (x,y,z) are shown in white dashed lines. Equipped with the FBT system, the next section investigates algorithms that can automate affine transformation estimation.

6.3 Fetal brain similarity registration with 3D CNNs

One of the most straight forward ways to register a 3D fetal brain ultrasound image to the FBT system would be to combine the plane detection method of Chapter 3 and the CC segmentation algorithm of Chapter 4. This sequential method would be limited in two ways. Firstly, by error propagation - if the algorithm failed to detect a good mid-sagittal plane in the first step, the next part will inevitably result in failure. Secondly, losing context information - the algorithm proposed for CC segmentation in the 2D mid-sagittal plane has proved to be fast and accurate (see Chapter 4 for details). However, using this approach to detect the reference landmarks of a 3D image discards valuable anatomical information. For example, a full FBT system requires knowledge of the spatial configuration between the CC-CSP complex and the thalamus. The latter is not easily discernible in the mid-sagittal plane in 3D ultrasound (see Fig.6.8(a)). Nevertheless, the thalamus has a distinct sonographic pattern and appears in a butterfly shape in the axial slice (see Fig.6.8(b)). Identifying the thalamus using this information is thus simpler and more reliable. To address these problems, the following study endeavours to define a unified approach to 3D similarity brain registration for ultrasound images that is able to utilize 3D contextual information. Due to the complexity of the problem,

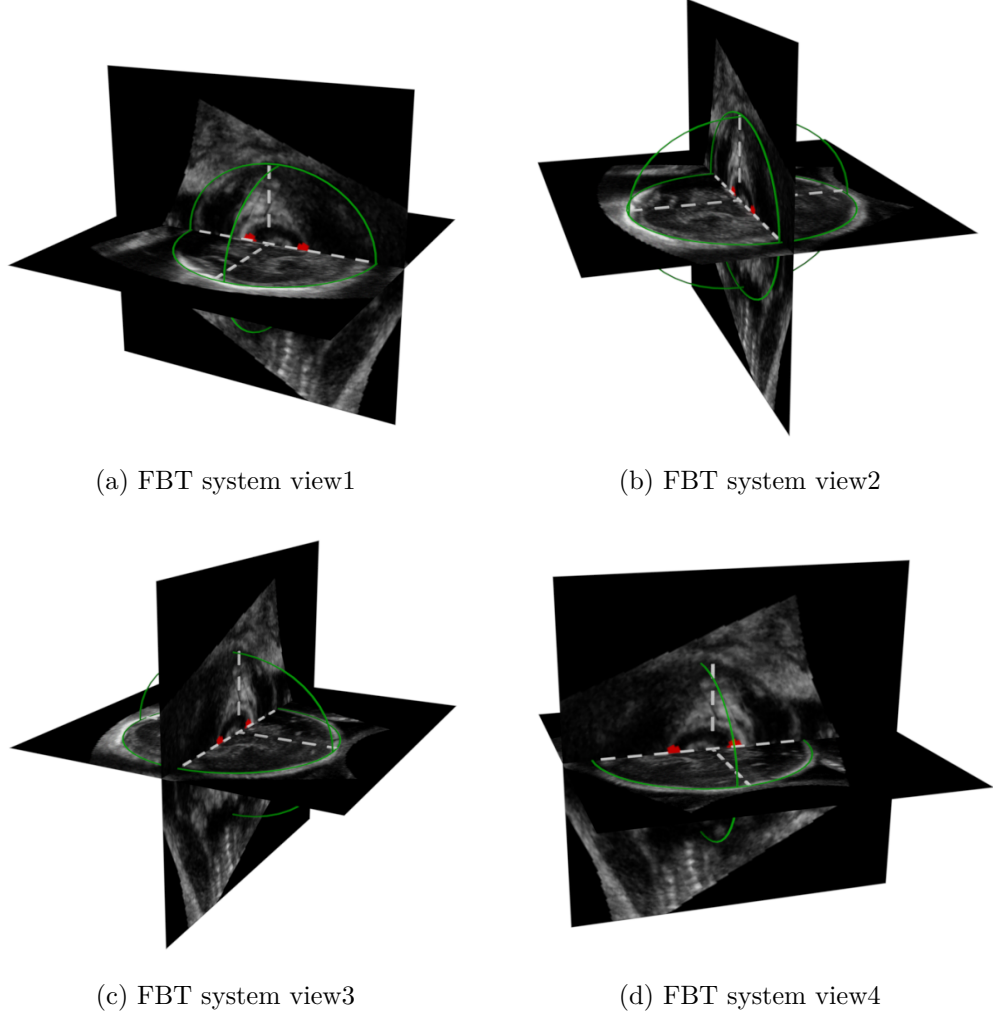


Figure 6.7: The established FBT system shown from four different angles. The vertical plane in each image is the mid-sagittal slice while the horizontal plane is the trans-thalamic plane. The skull boundary is represented in green ellipse frame. The reference landmarks (endpoints of the CC-CSP complex) are marked in red. The x,y,z axis of the system is plotted in white dashed line.

we broke down the task into a simpler problem, e.g. one direction alignment, and gradually increased the complexity of the models' objective.

6.3.1 Learning to align brains in the axial view

Deep learning algorithms have demonstrated incredible performance on classifying slices and detecting objects of interest in ultrasound images [68, 150]. However, to the best of our knowledge, no similar works has been done to estimate brain orientation in fetal neurosonography based on CNNs. In the vast body of literature

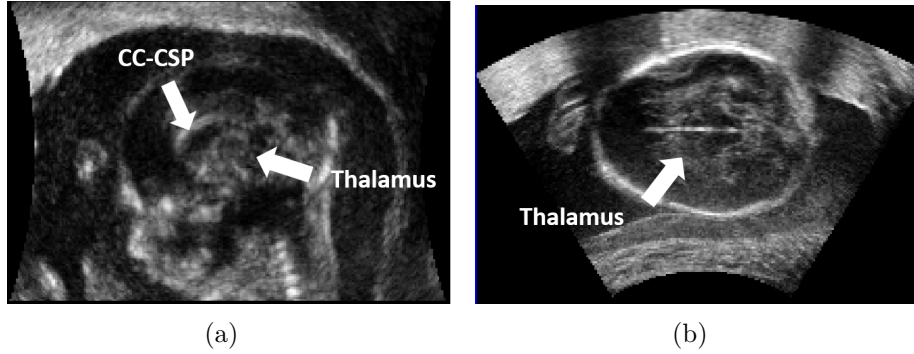


Figure 6.8: An example of fetal thalamus observed in ultrasound slices. (a) In the mid-sagittal slice, the thalamus is the structure beneath the CC-CSP complex (both pointed out by a white arrow). The thalamus does not possess a clear boundary in the mid-plane and appears as an obscure blob. (b) In the axial view, the thalamus consist of two symmetrical dark spot that combined in a butterfly shape. It is easier to identify the thalamus in this view for this unique pattern.

of computer vision, less attention has been paid to estimation of large rotation as its effect on their main subjects: natural images, are comparatively insignificant. A common solution in this scenario is to apply a small random rotation to increase the model robustness towards different orientation. By exposing the CNNs to enough simulated samples, the models can learn features to capture this variation. Some works have proposed methods to build rotation-invariant filters or networks to overcome this issue for remote sensing analysis or texture images [151, 152], whose nature is the opposite of this task. The ability of CNNs to learn the orientation of objects explicitly in ultrasound images has not been reported before.

To look at this problem in more detail, we first consider how to detect fetal brain orientation in one single view: the axial view. The key idea here is to learn the orientation angle in an unsupervised manner, via orientation simulation. All brain images were pre-aligned manually in the investigated view and a large random rotation transformation is applied and its value is used as supervision during training. An advantage of this method is it helps alleviate the need of a huge amount of data to train an over-parameterized CNN model. Since the rotation is simulated, in theory, the network can see an infinite amount of transformed data during training. Another contribution of this work is the addition of a 3D Spatial Transformer Network (STN). STNs were first introduced in [153] as insertable blocks that learn

to resample feature maps until the best alignment is achieved for a 2D classification task. In the literature of computer vision, this is also known as attention learning, as in theory, the STN should eliminate background information and focus on, i.e. zoom in, the region of interest [154]. The STN consists of two parts: (1) a localization network which predict the transformation parameters and (2) a grid generator that resamples the feature map. Here, instead of transforming the intermediate feature map, the grid generator transforms the input image based on the parameters that the localization network predicts. The transformed image is then compared with the original image (before rotation simulation). The difference between the two is a minimum if the prediction is optimal (close to the ground truth value).

To establish a fully 3D approach, it is intuitive to adapt the 3D U-Net baseline used in Chapter 5, which accepts a 3D fetal brain image as input and outputs bounding box masks for 5 key brain structures respectively. Figure 6.9 presents a schematic of the proposed model. The input is fetal brain images extracted from fetal neurosonography, shown as the black block. The dark blue represents conv block (convolution layer $\wedge$ batch normalization layer $\wedge$ ReLu activation layer) and light blue symbolizes the concatenated feature map, the STN module is shown in pink. In total, the model has six tasks: predicting the orientation of brain in the axial view, segmenting 3D bounding box of five key brain structures (output shown as grey blocks). Note that the STN also warps the input based on the predicted orientation and the result is used as a regularizer during training (shown in grey). The final loss function can be written as:

$$E(\theta) = \left(\sum_{s=1}^S \mathcal{L}_{loc}^s(\theta) \right) + \mathcal{L}_{ang}(\theta) + \alpha \mathcal{L}_{reg}(\theta) \quad (6.1)$$

, where θ is the network parameters, S corresponds to the number of structures (here $S = 5$), $\mathcal{L}_{loc}$ denotes the structure localization loss, and $\mathcal{L}_{ang}$ denotes the angle prediction loss. $\mathcal{L}_{reg}$ represents the regularizer loss, while α is the regularization strength which is set to a small number: $\alpha = 0.001$.

The STN is inserted in the middle of the CNN model where the feature map has the largest perceptive field and the smallest spatial size. In this way, the STN module

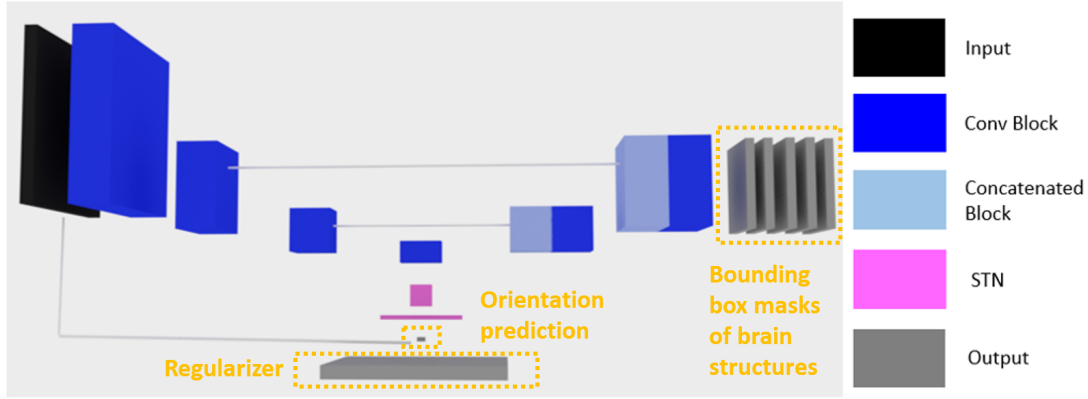


Figure 6.9: Sketch of the proposed 3D model. The basics architecture is similar to a 3D U-Nets, and takes an ultrasound brain image as input (black). The output consists of three parts: the rotation deviation of the brain in the axial view, an image transformed by the STN, bounding box masks of 5 key brain structures. Different neural network blocks are plotted in different colours (shown in the right of the figure).

does not need to down-sample the feature map aggressively before predicting the one orientation parameter. It avoids the possibility of blocking the gradients during back propagation (a bottleneck scenario). However, it is possible to insert the module in other positions of the network. Exploring this might be considered in future work.

Dataset To assess this CNN architecture we used a dataset consisting of 285 3D ultrasound fetal brain images of different healthy subjects. The fetuses age range was from 20 to 28 weeks. The typical size of each volume was $180 \times 200 \times 170$ voxels, with isotropic voxels measuring $(0.6 \times 0.6 \times 0.6 \text{ mm}^3)$ in size.

Pre-processing and Orientation Simulation The brains were extracted using the skull-stripping tool introduced in Chapter 5. All the images were aligned manually in the axial view. The dataset was partitioned into a subset of 240 samples for training and a subset of 45 samples for testing. For simplicity, the orientation simulation was applied as: rotation of an angle randomly selected within $\theta = \pm 60^\circ$ in the axial view. This rotation range was selected to be larger than natural images orientations, which typically have deviation less than $\pm 17^\circ$. The values of the simulated angles were recorded and used as supervision for STN prediction. Note that this orientation simulation can be alternately framed as a

type of data-augmentation and perform during the same stage with other type of augmentation. However, apart from the common function of data augmentation: to increase the dataset, the orientation simulation is also part of supervision during training. The nature of the model is essentially learning to register objects (aligning one orientation in this case) from synthetic data. It is therefore different from simple data augmentation where random perturbation is applied on an input, but the exact value of the perturbation is discarded.

Implementation Details The training was done end-to-end from scratch via an Adam optimizer [155], with an initial learning rate set as 10^{-3} . The weights were generated using the orthogonal initializer. The feature map was down-sampled three times use max-pooling to increase the receptive field and up-sampled twice to accommodate GPU memory limitation.

Results and Discussion The experiments were executed twice using the same model and training paradigm, with one alteration on the rotation supervision. The first attempt (Trial 1) of learning the angle deviation reported an error margin larger than the ground truth deviation (see Table. 6.1). After carefully examination, it was found that feeding the raw angles as ground truth (GT) is not appropriate as the same rotation can be represented by different numerical values due to angle wrapping, e.g. $-330^\circ, 30^\circ, 390^\circ$ indicate the same rotation transformation. If we frame the CNN model as a non-linear function as: $y = f(x; W, b)$, where x is the input image, W, b are learned weights and bias, and y is the output angle. In this case, the function $f(\cdot)$ does not have a unique solution for a given input x . An ideal prediction (close to the GT) could jump in steps of 360° based on the same feature activations. Depending on the networks to learn the desirable rotation range ($\pm 180^\circ$) without any regularization might be impractical and hard to achieve. Trial 2 thus modified the supervision given to the orientation prediction task (i.e. angle wrapping). To be more precise, a GT angle θ_d is first transformed from degrees to radians: $\theta_r = \frac{\pi}{180} \times \theta_d$, and divided by a factor of π : $\theta = \frac{\theta_r}{\pi}$ to normalize the value. Then a sigmoid function is applied: $S(\theta) = \frac{1}{1+e^\theta}$, to limit the value to the range of

Trial	Angle dev°	Cen Dev (mm)	Sca Dev (mm)	3D IOU(%)
1	69.7 $\pm$ 129.8	3.5 $\pm$ 2.9	3.9 $\pm$ 3.2	50.2 $\pm$ 20.4
2	5.3 $\pm$ 3.9	3.4 $\pm$ 3.0	3.9 $\pm$ 3.1	51.0 $\pm$ 21.3

Table 6.1: Quantitative results of two trials. Trial 1 is the model trained with unwrapped angles. Trial 2 is the model trained with corrected angles. Column 2 reports the angle deviation between the prediction and the ground truth in degree. It can be seen that unwrapped angle confuses the networks and lead to inferior predictions. In addition, column 3,4,5 reports the accuracy of detection tasks in terms of the centre deviation (Cen Dev), scale deviation (Sca Dev), and 3D Intersection over Union (IOU). The results show that the two trials achieved similar accuracy in detection.

0 $\rightarrow$ 1. The proposed model was then trained again using the same dataset and hyper-parameters. The second row of Table.6.1 reports the results. It can be seen that the angle prediction error reduced substantially after this simple modification.

Table. 6.1 also reports the model performance on the detection tasks obtained from other streams of the model (the five structure localization tasks, see Fig. 6.9). Compare the two rows, it shows the second trial outperformed the first one by a small margin while the overall results are similar. It is interesting to see that the failure of one stream within a multi-stream model did not destroy the whole model. One explanation of this might be that the model was trained to minimize the loss from six tasks simultaneously, therefore one single task has limited effects on the overall results. Note that both the results obtained from the two trials are inferior to those reported in Chapter 5. The conjecture is that the data were simulated with aggressive angle deviations that are not observed in real-world data. It is important to note the core purpose of these two experiments was to test the ability of the 3D CNN architecture to estimate the targeted orientation, while the structure detection tasks were included as additional supervision (they highlight the position of key brain structure in an image) to assist the rotation prediction. Overall, the results demonstrated that the CNN model can learn to predict the object orientation explicitly, while a suitable constraint should be applied on angle supervision during training.

6.3.2 Learning rotations in three anatomical views

This section builds on the result of Section 6.3.1 to address the following questions:

1. Can the model learn in an unsupervised manner or with less supervision?
2. Can the model be extended to predict rotation angles of three anatomical views simultaneously?
3. Is gestational age a factor of model performance?

A series of experiments have been designed accordingly. To test if extensive structural supervision was necessary (question 1), a one stream model (**Half-Net A**) was implemented that takes a brain image as input and outputs the rotation deviation of the axial view (i.e. simulated angle supervision only). To answer question 2, this model was extended in the final prediction layer to predict the rotation of the brain in all three anatomical views at the same time (denoted as **Half-Net B**). Validation experiment is designed to test whether age affects the model performance (question 3). Furthermore, as discussed earlier, the reference landmarks are key in deciding the orientation of the brain during the manual alignment process. It is rational to use these landmarks as additional supervision. Another model: **Full-Net**, was implemented based on **Half-Net B**, with an expansion stream that receives the CC-CSP landmarks supervision (as a further investigation of question 1). To make the full use of the spatial information represented by the landmarks, 3D sphere masks were created centred at the landmarks coordinates. The radius of the sphere was set to 5 pixels to increase the foreground-background ratio. Figure 6.10 shows an example of the created mask in the sagittal view.

Dataset The dataset consisted of 389 3D ultrasound fetal brain images of different healthy subjects, with age range from 18 to 30 weeks. The typical size of each volume is $200 \times 220 \times 200$ voxels, with isotropic voxels measuring $(0.6 \times 0.6 \times 0.6 \text{ mm}^3)$ in size. The dataset are partitioned into a subset of 312 samples for training and a subset of 77 samples for testing.

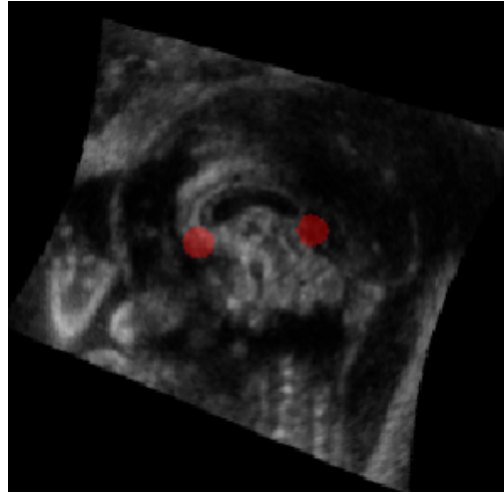


Figure 6.10: Schematic of the 3D sphere mask created for reference landmarks in the mid-sagittal plan. The landmarks are plotted in transparent red circles respectively.

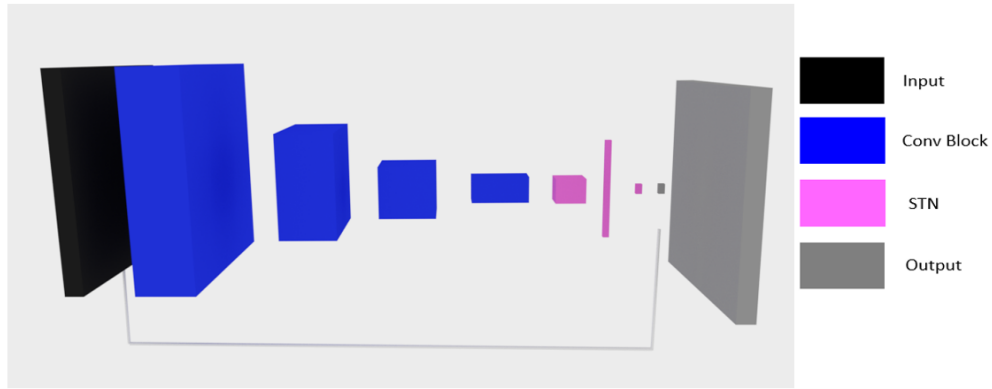
Pre-processing and Orientation Simulation The brains were extracted using the skull-stripping tool introduced in Chapter 5. All the images were manually aligned in three anatomical views. Note that the orientation simulation is performed in all three anatomical views for this experiment. Each aligned image was rotated by an angle randomly selected within $\theta = \pm 30^\circ$ for each anatomical view. It should be pointed out that the angle deviation range is smaller than that in the previous experiments to simulate realistic transformations. The value of simulated angles were recorded to trained the proposed CNN model. The CC-CSP landmarks were manually labelled for all the images. The landmark mask was rotated accordingly with its corresponding input image on the fly.

Implementation Details The training was done end-to-end from scratch via an Adam optimizer [155], with an initial learning rate set as 10^{-3} . The weights were generated using the orthogonal initializer. The feature map was down-sampled three times using max-pooling to increase the receptive field. The STN then flattened the 3D feature map and connected it to a fully-connected layer, yielding a 128 prediction layer for Half-Net A and a 128 layer for Half-Net B. Figure 6.11(a), 6.11(b) shows a schematic of the Half-Net A model. Full-Net is extended from Half-Net B, with two immediate up-sampling layers that concatenate with the previous feature

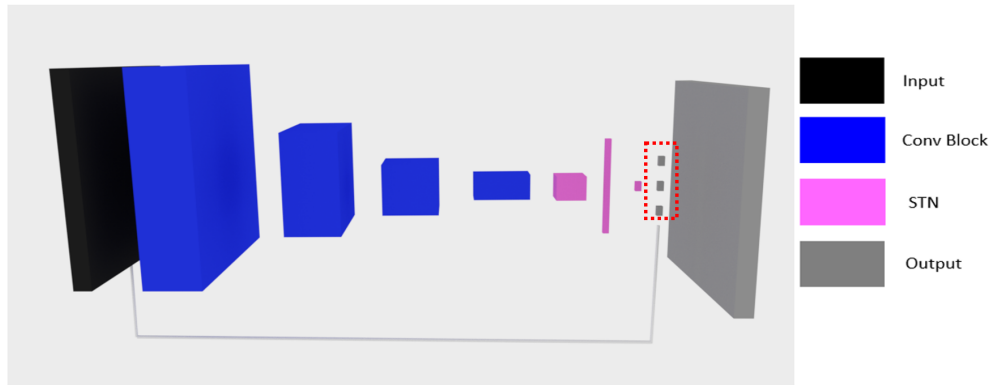
map of the same resolution. The expansive stream receives landmark supervision as 3D sphere masks (see Fig.6.11(c)). The network architecture is similar to that shown in Fig.6.9, except the expansive stream only has one output: landmark mask segmentation. Mean-square-error loss was used for angle prediction, while binary cross-entropy loss was used for landmarks segmentation. Inputs were resized to $96 \times 96 \times 96$ voxels before feeding into the networks.

Results and discussion Table. 6.2 reports the prediction accuracy of the three models in terms of angle deviation (from the GT) and landmark localization error (for Full-Net). The **Half-Net A** achieved good accuracy without any structural supervision. The angle prediction error is lower than that reported in Table. 6.1. This may due to the smaller rotation range ($\pm 30^\circ$ instead of $\pm 60^\circ$) applied on the input image. **Half-Net B** achieved a mean angle prediction error less than or equal to 5.2 degrees for all views. It shows that the 3D model is able to estimate the 3D orientation of the brain in the 3D space. The Full-Net scored the lowest error rate among the three. This demonstrates well that the landmarks masks provide good supervision for improving network precision. Note that the Full-Net also outperformed the model with extensive structural supervision (see Trial 2 in Table. 6.1). This may be because the five structures segmentation tasks distracted the model. The first conclusion drawn from this is that a suitable level of supervision should be imposed during training to better assist the target learning task. Furthermore, the problem boils down to finding ways to weight different tasks for a multi-stream model. Future studies will explore more flexible weighting mechanisms which can be adjusted based on real-time training loss. The localization error is also examined by calculating the distance between the centre of the predicted landmark mask and that of the GT. Results show that the network was able to localize the landmarks with an error margin lower than $1.6mm$ in average.

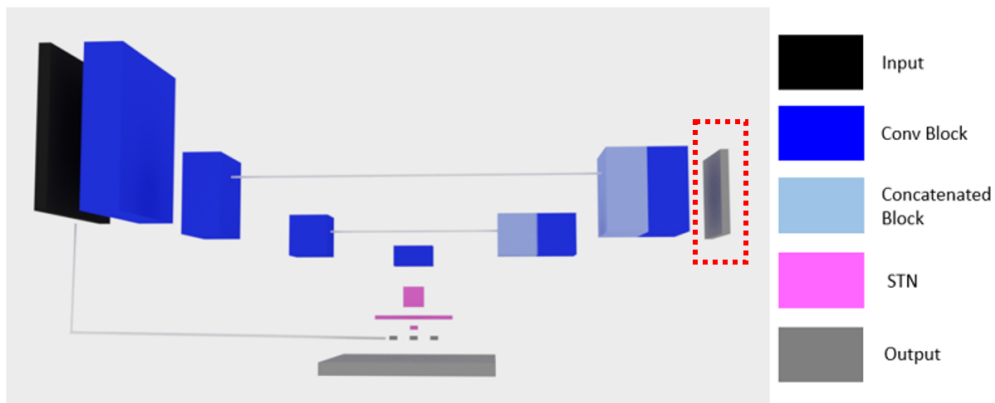
Figure 6.12 plots the estimated angles versus the ground truth for different views separately. In the left-hand side sub-figures, the x-axis is the GT value of the orientation in the corresponding view, while the y-axis is the prediction results.



(a)

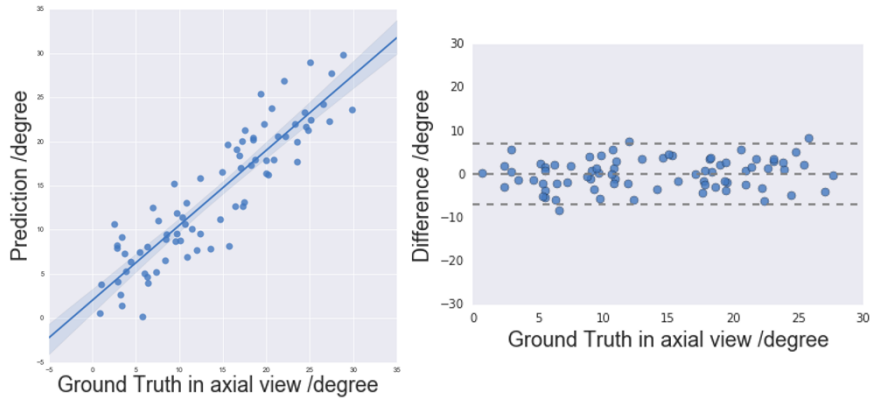


(b)

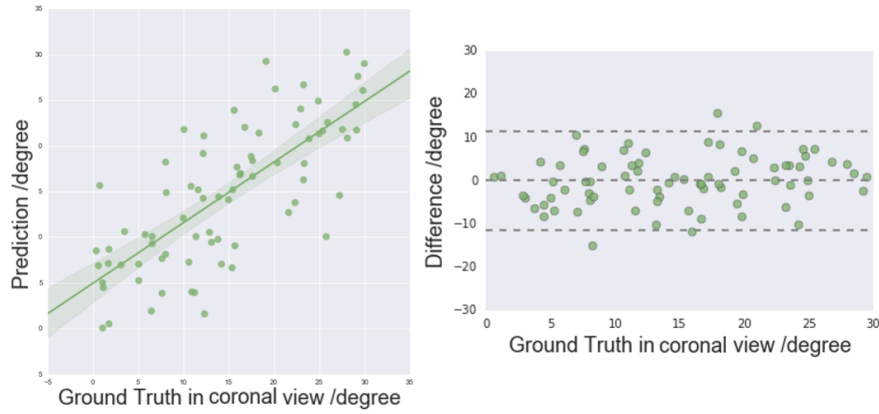


(c)

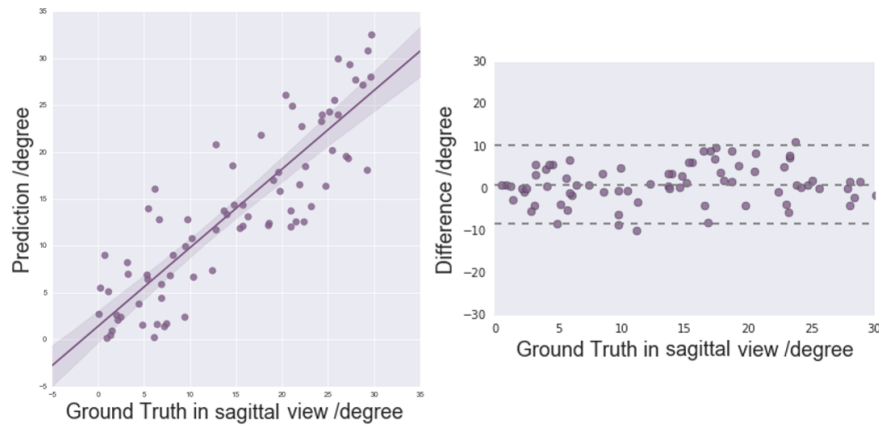
Figure 6.11: Schematic of the proposed models. (a) is the Half-Net A. It consists of several convolutional block (convolution $\wedge$ batch normalization $\wedge$ ReLu activation $\wedge$ Max-pooling layer), a STN module (pink) and prediction layer. The model takes an ultrasound brain image as input (black), and outputs the rotation angles of the image and an image transformed by the STN. (b) is the Half-Net B model. Note that the only difference is that the STN outputs three orientation estimations (highlighted by a red dashed box) instead of one. (c) is the Full-Net model. It is similar to the model shown in Fig.6.9, while the expansive stream only have one output: the CSP landmarks masks (the right most block in (c)).



(a) Axial view



(b) Coronal view



(c) Sagittal view

Figure 6.12: Results of the Full-Net model. Each row shows results of each anatomical view: axial - blue, coronal - green, and sagittal - purple. The sub-figure in the left in each row shows the plot of angle deviation between prediction and ground truth. The x-axis is the ground truth deviation in degrees, while the y axis is the predicted angles in degrees. Points located close to $y = x$ represent a good prediction. A straight line that best fits the scatters is calculated using logistic regression. The sub-figures in the right column are Bland-Altman plots, where the y-axis is the difference between the ground truth and the prediction.

Model	Axi ang dev°	Cor ang dev°	Sag axi dev°	Loc err(mm)
Half-Net A	5.0 ± 4.1	-	-	-
Half-Net B	4.7 ± 3.9	5.2 ± 4.2	4.5 ± 3.7	-
Full-Net	3.9 ± 3.1	4.9 ± 3.8	4.1 ± 3.2	1.6 ± 1.2

Table 6.2: Prediction error rate of the three tested model: Half-Net A, Half-Net B, and Full-Net. Column 2,3,4 reports the angle prediction error in Axial (Axi ang dev), Coronal (Cor ang dev), and Sagittal (Sag ang dev) view respectively. The Full-Net scored the lowest error rate in all three views.

Therefore, ideal predictions should approximate the $y = x$ line which has the same distance from the two axis. The right-hand side sub-figures are Bland–Altman plots. Figure 6.12(a) shows that the prediction corresponds well with the GT in the axial view, and the error does not increase with a larger GT value (where the alignment considered to be more difficult). Fig. 6.12(b) shows the model is less accurate in the coronal view. This phenomenon agrees with examinations of fetal neurosonography in practice where the coronal view is considered as the most challenging one in which to identify structures. Meanwhile, the reference landmarks give direct information on the brain orientation in the sagittal view and axial view which might help improve prediction in these two views.

Due to skull calcification and brain growth during pregnancy, the appearance of ultrasound scans change across gestation (see Fig. 6.14). It is thus interesting to examine if the model performs differently on subjects from different age groups. Figure. 6.13 plots the age of subjects versus the mean angle prediction error (mean of three views). The model achieves similar accuracy across a broad age range (gestational 18 to 30 weeks). This shows that the model is robust with respect to imaging variation caused by age.

In summary, this section shows that a CNN model can learn 3D orientations of brains with simulated angles (unsupervised), and that suitable landmarks supervision can help improve performance. The model we proposed was found to be robust towards image appearance variation caused by brain maturation.

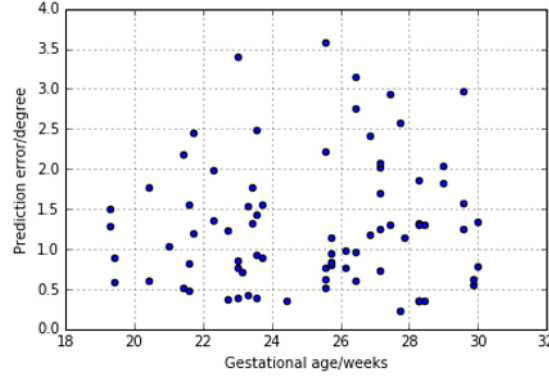
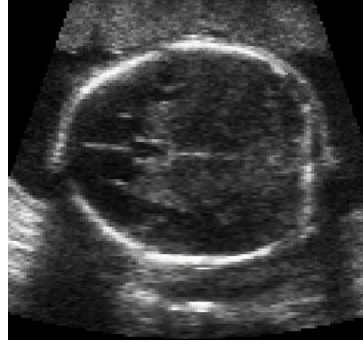


Figure 6.13: Plots of the age of each subject (x axis) versus the absolute value of its mean angle deviation (y axis). The figure shows that the model performed similarly across the gestational age despite the contrast and structural variance observed among different age groups. Note that the y-axis limits is set as $0 \rightarrow 4^\circ$ to zoom in on the difference between subjects.



(a) Subject of 18th GA weeks



(b) Subject of 30th GA weeks

Figure 6.14: Comparison of typical example of ultrasound scans collected at different gestational age (GA). The trans-thalamic plane of fetal head are selected for anatomical consistency. (a) is a scan collected at the 18th gestational week while (b) is a scan collected at the 30th gestational week. It can be seen that the image of the younger fetus has higher contrast and the skull size is smaller. The image of older fetus has lower contrast and the anatomy has more complex structures due to brain maturation. This is a common pattern observed in fetal neurosonography.

6.3.3 A full similarity 3D registration pipeline

This section takes a further step to incorporate translation and scaling factors into the model of Section 6.3.2 to build a full 3D similarity registration pipeline. It should be noted that estimation of translation and scale are closely related to skull detection in a raw image while previous experiments were tested on skull-stripped images. An additional skull segmentation task is thus added to the Full-Net (see Section.6.3.2).

In the new model, the STN module predicts the rotation parameters while the skull mask generates the translation and scaling parameters for the final transformation. As a result, an integrated pipeline is completed that takes an original 3D ultrasound image as input and outputs a full 3D similarity transformation matrix that maps the 3D image into the FTB coordinate system (see Fig 6.15).

Dataset The dataset consisted of 389 3D ultrasound fetal brain images of different healthy subjects, with age range from 18 to 30 weeks. The typical size of each volume is $200 \times 220 \times 200$ voxels, with isotropic voxels measuring $(0.6 \times 0.6 \times 0.6 \text{ mm}^3)$ in size.

Pre-processing and Orientation Simulation Note that different from Section 6.3.1 and 6.3.2 where the extra-cranial tissues were removed, this experiment used the uncropped ultrasound images. All training images were pre-aligned manually to the standard orientation (defined by the FBT system). Orientation simulation was carried out in the same way introduced in Sect. 6.3.2. The test were done on un-aligned images.

Implementation Details The same network architecture and training paradigm was used as for Section 6.3.2. Note that due to the computer memory required for an 3D CNN model, the input image size was limited to $96 \times 96 \times 96$ to afford a model with appropriate depth. Down-sampling the input image directly loses useful image information that is not recoverable by a network. Instead, an input image is randomly cropped with a factor of s_x, s_y, s_z from 0.5 to 0.7 in x, y, z dimension respectively during training. This scheme can also be regarded as an aggressive data augmentation which would help the model to learn translation-invariant features. Note that during testing, images were cropped in the centre of the volume for consistency (crop size: $128 \times 128 \times 128$ voxels).

Results and Discussion Quantitative results of the proposed pipeline are reported in Table. 6.3. The first row lists the angle deviation error in each anatomical view and the mean of them in each subject. The coronal view was found to be the

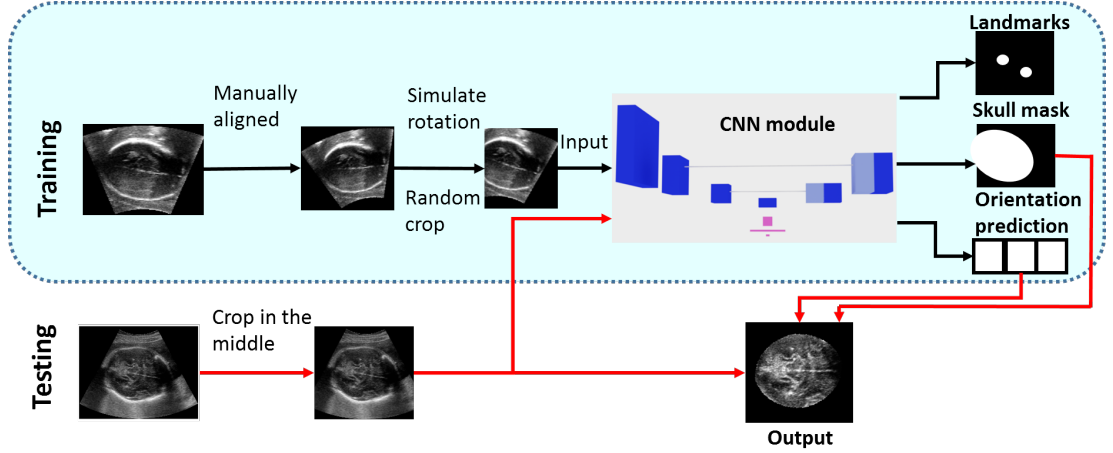


Figure 6.15: Schematic of proposed pipeline. During training, an image is pre-aligned and processed with suitable orientation simulation and data-augmentation (e.g. rotation and cropping). The CNN module is trained to predict the orientation of the brain, segment the skull and detect the landmarks. During testing, an unseen image is cropped in the middle and passed into the trained CNN model. The image is subsequently rotated based on the predicted orientation angles and masked by the predicted skull mask. The output is a rigid aligned brain image based on the FBT standard.

most challenging while the model is the most confident in the axial view. Note that the error margin is larger than that of the **Full-Net** model (reported in Table. 6.2), which is as expected given that the input is now (cropped) raw ultrasound images versus the skull-stripped images used in the previous experiment. As a result, the task is much more challenging but more realistic due to the presence of maternal tissues. The second and the third rows report the translation and scaling error in each dimension. The results show that the model is able to accurately detect and localize the brain in three dimensions. Interestingly, the coronal view has the highest translation and scaling prediction accuracy. One conjecture is that the error margin might be correlated with the size of the brain and the brain is relatively smaller in the coronal view.

Three interesting cases are now discussed as qualitative examples. Here, the discussion focuses on the accuracy of orientation estimation as this is the most challenging task of the pipeline. Figure 6.16(a) shows the registration results of the subject with the largest orientation angle ($GT = 17^\circ$) in the axial view within the test set. It proves that the model can handle large orientation deviation. Meanwhile,

Component	Axial	Coronal	Sagittal	Mean
Angle dev /°	4.8 ± 3.7	5.5 ± 4.4	5.3 ± 4.1	5.2 ± 4.0
Translation /pixel	0.5 ± 0.4	0.4 ± 0.2	0.7 ± 0.6	1.0 ± 0.6
Scale /pixel	3.0 ± 2.8	2.2 ± 1.4	3.1 ± 2.3	2.8 ± 2.4

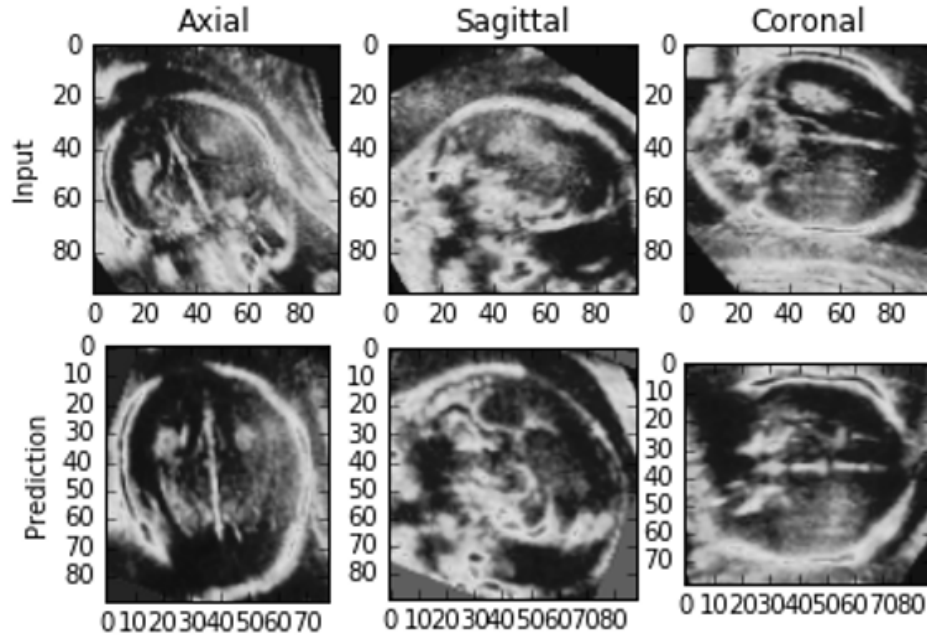
Table 6.3: Similarity registration error of the final pipeline. The rows reports angle deviation, translation and scale errors. The columns represent axial, coronal, sagittal view and the mean across them respectively. Note that the mean error of translation here is not the mean across different views, but the rather is the distance of the predicted centre and the GT centre in 3D space.

the pipeline also aligned the brain nicely in the sagittal and coronal view (see the 2 bottom-right images in Fig. 6.16(a)). The transformed image is also centred correctly in the middle of brain, demonstrating that the network can find the correct translation and scaling parameters. The example in Fig. 6.16(b) has the largest mean angle deviation in the test set: 12° in axial, 18° in sagittal, and 1° in coronal. The prediction result demonstrates the model handles certain challenging cases well where subjects have a large orientation in two views simultaneously. Meanwhile, although part of the brain was not visible, the model was able to detect the skull and re-centre the image correctly.

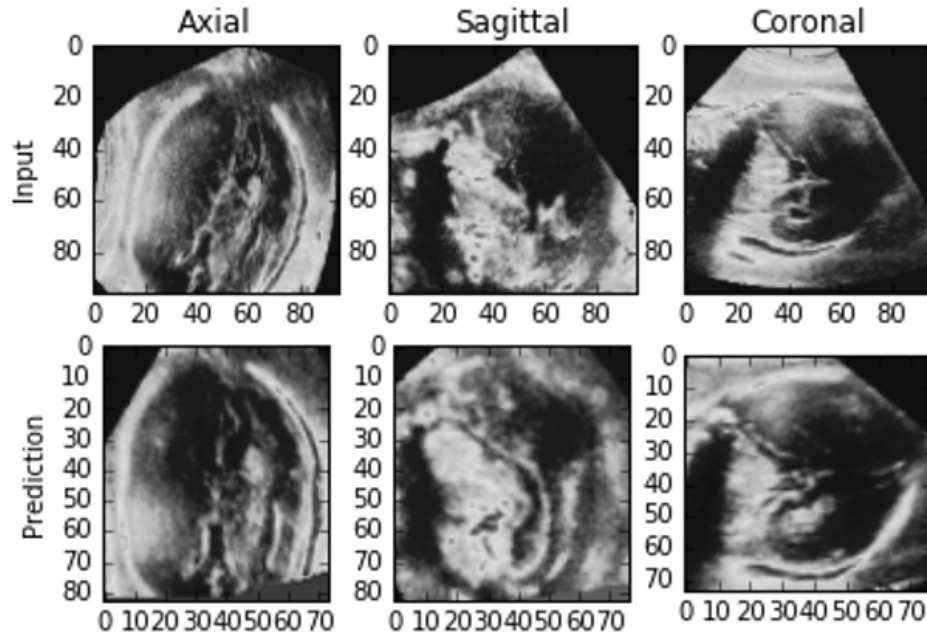
Figure 6.17 displays the case with the largest prediction rotation error in the test set: 2° in axial, 9° in sagittal, and 15° in coronal deviation from the GT. The conjecture is that the image quality is inferior for this particular case, which may have resulted from skull calcification in late pregnancy (subject age: 28 gestational week).

6.4 Conclusion and future study

This chapter first introduced a standard 3D fetal brain coordinate system (FBT system) for the fetal brain in the context of ultrasound images. The system was inspired by the Talairach space proposed by the MRI brain research community, and adapted accordingly with the unique anatomical patterns observed in 3D ultrasound images. Subsequently, the study investigated CNN approaches to rigid registering fetal brain based on the FBT system. The experiments revealed that: 1) A 3D CNN model can be trained to learn the orientation from data explicitly; 2) A balanced supervision yields the best results in predicting the rotation transformation; 3) the



(a) Example a



(b) Example b

Figure 6.16: Examples of the similarity registration results from the test set. The first row for each subject is the input image, the second row is the transformed image based on the prediction. Each column is the middle slice of the volume in each dimension. The input image of Example a has the largest axial rotation deviation ($GT = 17^\circ$) in the test set. The input of Example b has the largest mean angle deviation in the test set: 12° in axial, 18° in sagittal, and 1° in coronal. The prediction result demonstrates the model handles well on the subjects with large orientation and deviations in multiple views.

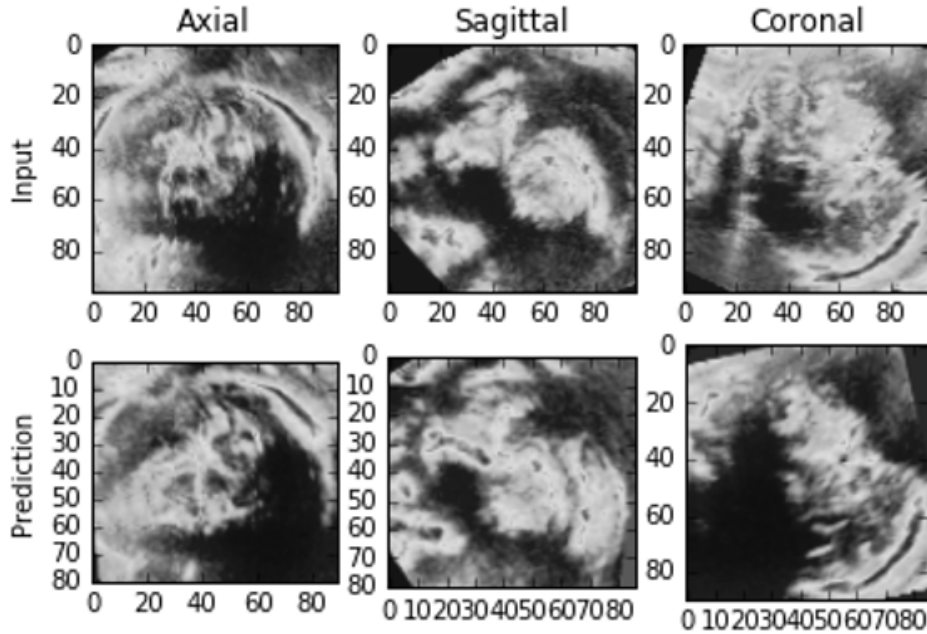


Figure 6.17: The case of the largest mean orientation prediction error in the test set. It might be explained that the image quality is not ideal thus obstructing structural visibility.

trained model is robust towards biological variation caused by brain maturation. Finally, a fully automatic similarity registration pipeline was built which directly registers a raw ultrasound image to the proposed FBT systems. The findings reported here not only can be applied to fetal neurosonography, but also to register other 3D images of different modalities. Furthermore, this pilot study also raises some interesting questions. For example, is there a better way to weight different tasks in training a multi-stream model? In addition, recall that the input image has to be cropped and resized due to computer memory limitations. Can we circumvent this problem given limited GPU capacity? Alternatively, should the model be applied on a raw volume in a sliding window manner or averaged over a multiple of locations? Can registration accuracy be further improved using other supervisions? The investigation of these questions is left as future work.

7

Conclusion

This thesis has presented a series of automatic image analysis algorithms for visualization, segmentation, localization and registration purposes in fetal neurosonography. This chapter summarizes the main contributions of this thesis and discusses the limitations of the present work and directions for future research.

7.1 Contributions

The main contributions of this thesis include:

- 1. A skull detection algorithm to remove the extra-cranial tissues and a plane detection pipeline for visualizing the mid-sagittal profile**

We presented a fully automated skull detection algorithm for fetal neurosonography. The algorithm was designed based on template matching and is easy to implement. We further explored the algorithm performance by manipulating the resolution of input images. Results showed that reduction of resolution can accelerate processing procedure and boost accuracy, while aggressive image compression has a negative effect. Though the algorithm was originally proposed for 2D ultrasound images, we showed how it can be extended to extract skull in 3D images.

Built on top of this skull detection algorithm, we designed a plane detection pipeline that extracts the mid-sagittal plane from a 3D fetal brain ultrasound

images. The pipeline first selected a stack of axial slices, then it detected the skull on these slices. The midline of brain in each slice was subsequently detected via the Hough transform. Finally, the mid-sagittal plane is defined by fitting a plane to the detected midlines. This fully automated pipeline achieved comparable plane extraction accuracy with inter-/intra-observer results. The application of the proposed pipeline can reduce the time and effort required to visualize the mid-sagittal plane, which contains important anatomical information.

2. A general framework for automatic segmentation of fetal brain structures in 2D ultrasound images

We described a general framework to automatically segment fetal brain structures in 2D ultrasound images. This framework generated maximal stable extremal regions (MSERs) as candidate regions and used an original region descriptor to characterize these candidate regions. Random Under-Sampling boost (RUSboost) is employed to select the desired region out of the pool of candidates. The proposed framework is tested to segment two clinical important brain structures - the corpus callosum (CC) and the choroid plexus (CP) - whose appearance and image context are considerably different. Validation experiments showed that the shape of a structure is a more distinguishing feature compared to its mean intensity value and its location within an image. The proposed region descriptor was able to extract reliable shape features and achieved high classification accuracy. The fact that our framework was able to delineating the CC and the CP despite their appearance difference validates the adaptability and the robustness towards imaging condition variations in ultrasound images. The framework can be extended to segment other 2D structures brain structures in fetal neurosonography given appropriate training data.

3. Efficient algorithm to localize multiple fetal brain structures in 3D ultrasound images

We proposed a novel View-based Projection Networks (VP-Nets) model for structures localization in 3D ultrasound images. The VP-Nets use geometric projection to reduce the dimensionality and complexity of the task, therefore enabling high 3D-resolution output which is unaffordable using current modern

GPUs to process a fully 3D CNN model. During testing, the trained VP-Nets model needs less than 0.1s to process a brain volume. It output the bounding boxes of different structures separately to avoid label ambiguity in overlap regions. The five brain structures considered by the model are those examined in a standard clinical examination: lateral ventricles (LV), the thalamus (Tha), the cerebellum (CB) and cisterna magna (CM), and cavum septi pellucidi (CSP). The application of the model can reduce the time to analyse a brain ultrasound scan in daily routine. Due to its fast run-time, the trained model could be used to process large datasets and analyse the fetal brain development among different populations in the future.

4. Fetal Brain Talairach (FBT) system and an automatic affine registration method for 3D fetal brain images

To overcome the challenges introduced by varying position and orientation of fetal brains in ultrasound, we described a standard coordinate system, which we call the Fetal Brain Talairach (FBT) system. The FBT system is adapted from the Talairach system used by the MRI research community, and accustomed to fetal neurosonography by choosing reliable landmarks that are visible in ultrasound whilst have similar anatomical context to that used in MRI. The FBT system allows any individual fetal 3D neurosonogram to be registered to a common space. Its application facilitates analysis and comparison of brain development among different subjects or the same subject at different ages. It also enables a fundamental step towards a full non-rigid fetal brain registration which will be considered in the future work.

With the established FBT system, we further proposed an automatic affine registration algorithm for 3D fetal neurosonography. The CNN-based algorithm uses simulated orientation during training to achieve a highly accurate estimation of brain orientation. Several experiments were conducted to test the method and the effects of different supervision. The final model takes a fetal neurosonography scan as input and outputs a 3D affine registered brain image (extra-cranial tissues are removed). Validation experiments showed that the trained model was robust towards biological variation caused by brain maturation. The application of the tool

could be used to transfer segmentation maps between subjects, analyse individual brain growth across gestation etc. It could simplify routine examination as brains are mapped into a standard coordinate system so the brain structures should locate at a similar spatial position. These applications could be explored in future work.

7.2 Limitations and Future Work

Incorporation of context information in plane extraction

Chapter 3 proposed an automatic approach to extract the MSP plane in 3D fetal brain ultrasound volumes. The pipeline took advantage of the appearance of the MSP in different anatomical directions and used a traditional image processing approach to detect the plane. In spite that the algorithm achieved good accuracy on the validation dataset, its performance and robustness might be further improved by incorporating context information. For example, in later project (during data collection for **Chapter 6**), it is found that the brain can have large orientation deviation in two or more axes from the image axis in some extreme cases. In this scenario, the MSP is distorted and its appearance violated the assumption made in **Chapter 3** thus the proposed framework might fail. Although these cases are rare and might result from violation of a clinical protocol (where a suitable position of the probe is suggested), it is possible to overcome this issue by considering context information, e.g. the location of mid-brain structures or the relative position between the plane and the skull.

Some of the research presented in this thesis has already considered this aspect. For example, the affine registration tool presented in **Chapter 6** has a large respective field though multiple down-sampling of feature maps. With its skull-stripping module and orientation prediction, the mid-sagittal plane can be found by selecting the middle slice of a registered brain. In future work, it might be interesting to compare the MSPs extracted in this way to manual selected planes.

Applicability of detail delineation in ultrasound images

Chapter 4 presented a method for automatic segmentation of fetal brain structures and reported good segmentation results. However, in the discussion

of some challenging cases, it is found that sometimes the target structures have indiscernible boundaries due to poor image quality or adjacency to other brain tissues. Human observers often make subjective assumptions to overcome this issue while the algorithm failed to handle these situations well. One solution to this problem is to re-acquire the scan and improve the image contrast. However, this might not be always applicable as it requires time and effort from both the clinicians and the patients. It then opens up the question whether detail delineation is necessary for complex fetal brain structures in ultrasound due to the challenges it faces and lack of golden standard. It is important to remember that the original purpose of segmenting a structure is to analyse the difference and changes of a structure (from the same or different individuals) which are visible to the human eyes. It is true that this information is crucial to understand structural development and malformation caused by congenital disease. However, it might be the case that some of the information contained in ultrasound images is hard to, or cannot be, interpreted by the naked eyes. Moreover, it is possible that some of the structures or image regions are affected by the disease but this is overlooked by current clinical diagnosis standard. For example, as discussed in **Chapter 2**, the fetal alcohol spectrum disorder (FASD) is one of the most common congenital diseases. Prenatal alcohol exposure can damage the fetal brain in different ways and different degrees which are not fully understood. Clinical studies report different findings as some pointed out the volume reduction of the cerebellum [37, 38], other reported the malformation of corpus callosum [39, 40], while other structures (e.g. Basal ganglia) are also mentioned. There is no unified agreements on which brain structures are affected and which are not, due to the complexity of this disorder. Other brain regions might also be impaired by alcohol exposure, and future diagnosis can benefit greatly from these additional information.

With the advancement of modern machine learning and the establishment of large medical datasets, it is interesting to see if cutting-edge image analysis algorithms can learn from the data and assist diagnosis. It would be interesting to see if the algorithms would agree with clinical findings or pinpoint regions that

have previously been overlooked by the human eye. This would be fascinating to investigate in the future if the disease datasets are available.

Possibility of building an end-to-end model

Chapter 5 presented an efficient model to localize multiple key brain structures in fetal neurosonography. The model is fast, accurate and has an affordable memory footprint. However, the model requires a raw fetal neurosonogram to be preprocessed via an automatic skull-stripping algorithm. This pipeline is designed due to the extreme low foreground-to-background ratio of a brain structure compared to a whole fetal neurosonography scan. Though the whole pipeline is fully automatic, it is a piecewise approach that allows error propagation. Future research might investigate direct localizing structure from a raw volume. Moreover, the VP-Nets model outputs bounding box masks of a structure in different anatomical directions simultaneously. To reconstruct the 3D bounding box of a target using backward-projection, it is sufficient to use two predicted masks collected from different anatomical direction. It might be interesting to investigate which two view should be selected for each individual structure for a better prediction. This point is left to explore as future work.

Future work in registering fetal brains in ultrasound

Chapter 6 described a standard coordinate system for fetal neurosonography and discussed methods to automatically align fetal brains ultrasound volumes using affine registration. The experiments revealed that the CNN model is able to capture the orientation of a target in 3D space with proper supervision and angle wrapping. However, this research also identified some questions. Currently, a 3D image is resized to 96×96 before feeding into the networks due to limited GPU capacity. Is there any way to circumvent this information loss though model modification? Is there a better way to supervise the registration networks or weight different supervision based on a real-time loss? During testing, an input volume is cropped in the middle to have consistent resolution with training. Should we instead adopt a sliding window approach during testing and will that affect the result? These are all interesting questions to consider and answer in the future.

Moreover, another direction of future research would be the extension to non-rigid 3D registration. With the unique ultrasonic patterns contained in the images, it might be inapplicable to simulate non-rigid transformation on ultrasound images to train a CNN model. If manual supervision is necessary, what structures should be labelled and how should they be labelled remain as open questions.

7.3 Summary

The research presented in this thesis focused on developing original image analysing algorithms for fetal neurosonography. The work has explored traditional image processing algorithms, signal processing techniques, and cutting-edge machine learning algorithms. We have developed a suite of algorithms with potential for clinical use which may simplify routine clinical examination. During the development of these algorithms, we discovered challenges to analyse ultrasound images and designed specific algorithms to overcome them. Future work will aim to further improve the presented approaches through investigating technical details of the established frameworks and incorporation of larger datasets. The performance of the presented models on disease/malformed cases is also interesting and might be examined in the future.

Appendices

Layer Type	Conv	BNorm	Conv	BNorm
Filter Size	$3 \times 3 \times 3 \times C$	–	$3 \times 3 \times 3 \times C$	–

Table 1: Convolution Block Details for the 3D U-Net. BNorm refers to Batch Normalization [156]. ReLU is used after each BNorm.

Layer Type	Output size	Memory Cost
Input:	$128 \times 128 \times 128 \times 1$	$\approx 0.008\text{G}$
ConvBlock 1:	$128 \times 128 \times 128 \times 64$	$\approx 1.073\text{G}$
MaxPooling 1:	$64 \times 64 \times 64 \times 64$	$\approx 0.067\text{G}$
ConvBlock 2:	$64 \times 64 \times 64 \times 64$	$\approx 0.134\text{G}$
MaxPooling 2:	$32 \times 32 \times 32 \times 64$	$\approx 0.008\text{G}$
ConvBlock 3:	$32 \times 32 \times 32 \times 64$	$\approx 0.016\text{G}$
MaxPooling 3:	$16 \times 16 \times 16 \times 64$	$\approx 0.001\text{G}$
ConvBlock 4:	$16 \times 16 \times 16 \times 64$	$\approx 0.002\text{G}$
MaxPooling 4:	$8 \times 8 \times 8 \times 64$	$\approx 0.0001\text{G}$
ConvBlock 5:	$8 \times 8 \times 8 \times 64$	$\approx 0.0002\text{G}$
Deconv 5:	$16 \times 16 \times 16 \times 64$	$\approx 0.002\text{G}$
Merge 5:	$16 \times 16 \times 16 \times 128$	$\approx 0.004\text{G}$
ConvBlock 6:	$16 \times 16 \times 16 \times 64$	$\approx 0.002\text{G}$
Deconv 6:	$32 \times 32 \times 32 \times 64$	$\approx 0.016\text{G}$
Merge 6:	$32 \times 32 \times 32 \times 128$	$\approx 0.032\text{G}$
ConvBlock 7:	$32 \times 32 \times 32 \times 64$	$\approx 0.064\text{G}$
Deconv 7:	$64 \times 64 \times 64 \times 64$	$\approx 0.134\text{G}$
Merge 7:	$64 \times 64 \times 64 \times 128$	$\approx 0.268\text{G}$
ConvBlock 8:	$64 \times 64 \times 64 \times 64$	$\approx 0.268\text{G}$
Output(CE):	$64 \times 64 \times 64 \times 1$	$\approx 0.001\text{G}$
Output(CM):	$64 \times 64 \times 64 \times 1$	$\approx 0.001\text{G}$
Output(CSP):	$64 \times 64 \times 64 \times 1$	$\approx 0.001\text{G}$
Output(LV):	$64 \times 64 \times 64 \times 1$	$\approx 0.001\text{G}$
Output(THA):	$64 \times 64 \times 64 \times 1$	$\approx 0.001\text{G}$
–	–	Sum $\approx 2.10\text{G}$

Table 2: Architecture of 3D U-Net and memory cost for single input volume during feed-forward process. To compute gradients in the backward pass, the activation of all layers should be stored. Therefore, each back-propagation step takes approximately twice the memory and the compute time than the forward propagation step [138]. For all layers, we simply applied 64 convolutional filters. Thus, while training on single input volume, the sum of memory consumption for the whole 3D-UNet should be 4.2G.

Layer Type	Conv	BNorm	Conv	BNorm
Filter Size	$3 \times 3 \times 3 \times C$	–	$3 \times 3 \times 3 \times C$	–

Table 3: ConvBlock Detail for the Skull segmentation Network. BNorm refers to Batch Normalization [156]. ReLU is used after each BNorm.

Layer Type	Output size
Input:	$96 \times 96 \times 96 \times 1$
ConvBlock 1: $96 \times 96 \times 96 \times 64$	
MaxPooling 1:	$48 \times 48 \times 48 \times 64$
ConvBlock 2:	$48 \times 48 \times 48 \times 64$
MaxPooling 2:	$24 \times 24 \times 24 \times 64$
ConvBlock 3:	$24 \times 24 \times 24 \times 64$
MaxPooling 3:	$12 \times 12 \times 12 \times 64$
ConvBlock 4:	$12 \times 12 \times 12 \times 64$
MaxPooling 4:	$6 \times 6 \times 6 \times 64$
ConvBlock 5:	$6 \times 6 \times 6 \times 64$
Deconv 5:	$12 \times 12 \times 12 \times 64$
Merge 5:	$12 \times 12 \times 12 \times 128$
ConvBlock 6:	$12 \times 12 \times 12 \times 64$
Deconv 6:	$24 \times 24 \times 24 \times 64$
Merge 6:	$24 \times 24 \times 24 \times 128$
ConvBlock 7:	$24 \times 24 \times 24 \times 64$
Deconv 7:	$48 \times 48 \times 48 \times 64$
Merge 7:	$48 \times 48 \times 48 \times 128$
ConvBlock 8:	$48 \times 48 \times 48 \times 64$
Deconv 8:	$96 \times 96 \times 96 \times 64$
Merge 8:	$96 \times 96 \times 96 \times 128$
ConvBlock 9:	$96 \times 96 \times 96 \times 64$
Output:	$96 \times 96 \times 96 \times 1$

Table 4: Architecture of the skull segmentation network. The whole structure is similar to 3D U-Net and has 9 layers. For simplicity, each layer has 64 convolutional filters. Due to the memory consumption of 3D U-Net, the resolution of input images is reduced to $96 \times 96 \times 96$. The network output skull segmentation result in the same resolution.

Layer Type	Conv	BNorm	Conv	BNorm
Filter Size	$3 \times 3 \times C$	—	$3 \times 3 \times C$	—

Table 5: Convolution Block Details for the VP-Net F. BNorm refers to Batch Normalization [156], ReLU is used after each BNorm.

Layer Type	Output size	Memory Cost
Input:	$128 \times 128 \times 128$	$\approx 0.032\text{G}$
ConvBlock 1:	$128 \times 128 \times 256$	$\approx 0.032\text{G}$
MaxPooling 1:	$64 \times 64 \times 256$	$\approx 0.004\text{G}$
ConvBlock 2:	$64 \times 64 \times 256$	$\approx 0.008\text{G}$
MaxPooling 2:	$32 \times 32 \times 256$	$\approx 0.0012\text{G}$
ConvBlock 3:	$32 \times 32 \times 256$	$\approx 0.0024\text{G}$
MaxPooling 3:	$16 \times 16 \times 256$	$\approx 2.6e-4\text{G}$
ConvBlock 4:	$16 \times 16 \times 256$	$\approx 5.2e-4\text{G}$
MaxPooling 4:	$8 \times 8 \times 256$	$\approx 6.4e-5\text{G}$
ConvBlock 5:	$8 \times 8 \times 256$	$\approx 1.3e-4\text{G}$
Deconv 5:	$16 \times 16 \times 256$	$\approx 5.2e-4\text{G}$
Merge 5:	$16 \times 16 \times 512$	$\approx 1.0e-3\text{G}$
ConvBlock 6:	$16 \times 16 \times 256$	$\approx 1.0e-3\text{G}$
Deconv 6:	$32 \times 32 \times 256$	$\approx 0.0024\text{G}$
Merge 6:	$32 \times 32 \times 512$	$\approx 0.0048\text{G}$
ConvBlock 7:	$32 \times 32 \times 256$	$\approx 0.0048\text{G}$
Deconv 7:	$64 \times 64 \times 256$	$\approx 0.004\text{G}$
Merge 7:	$64 \times 64 \times 512$	$\approx 0.008\text{G}$
ConvBlock 8:	$64 \times 64 \times 256$	$\approx 0.008\text{G}$
Deconv 8:	$128 \times 128 \times 256$	$\approx 0.016\text{G}$
Merge 8:	$128 \times 128 \times 512$	$\approx 0.032\text{G}$
ConvBlock 9:	$128 \times 128 \times 256$	$\approx 0.032\text{G}$
Output(CE):	$128 \times 128 \times 1$	$\approx 2.6e-4\text{G}$
Output(CM):	$128 \times 128 \times 1$	$\approx 2.6e-4\text{G}$
Output(CC):	$128 \times 128 \times 1$	$\approx 2.6e-4\text{G}$
Output(LV):	$128 \times 128 \times 1$	$\approx 2.6e-4\text{G}$
Output(THA):	$128 \times 128 \times 1$	$\approx 2.6e-4\text{G}$
—	—	Sum $\approx 0.2\text{G}$

Table 6: Architecture of a single stream of **VP-Net F**. It has (9 ConvBlocks) and its memory cost for single stream during feed-forward process are listed. Theoretically, back-propagation should take the same amount of memory. For all layers, we simply applied 256 convolutional filters. Thus, while training on single input volume, the sum of memory consumption for the whole VP-Nets (three streams) should be only **1.2G**.

References

- [1] Joan Stiles and Terry L Jernigan. “The basics of brain development”. In: *Neuropsychology review* 20.4 (2010), pp. 327–348.
- [2] Arzou Ahsan. *Obstetrics & Gynecology*. Lippincott Williams & Wilkins, 2007.
- [3] ISUOG et al. “Sonographic examination of the fetal central nervous system: guidelines for performing the ‘basic examination’ and the ‘fetal neurosonogram’.” In: *Ultrasound in obstetrics & gynecology* 29.1 (2007), p. 109.
- [4] LW Chan et al. “Volumetric (3D) imaging reduces inter-and intraobserver variation of fetal biometry measurements”. In: *Ultrasound in Obstetrics & Gynecology* 33.4 (2009), pp. 447–452.
- [5] TC Chang et al. “Ultrasonic fetal weight estimation: Analysis of inter-and intra-observer variability”. In: *Journal of clinical ultrasound* 21.8 (1993), pp. 515–519.
- [6] Beryl R Benacerraf. “Three-dimensional fetal sonography use and misuse”. In: *Journal of Ultrasound in Medicine* 21.10 (2002), pp. 1063–1067.
- [7] Shidong Tong et al. “Intra-and inter-observer variability and reliability of prostate volume measurement via two-dimensional and three-dimensional ultrasound imaging”. In: *Ultrasound in medicine & biology* 24.5 (1998), pp. 673–681.
- [8] FP Hadlock et al. “Fetal head circumference: relation to menstrual age”. In: *American Journal of Roentgenology* 138.4 (1982), pp. 649–653.
- [9] Frank P Hadlock et al. “Estimating fetal age: effect of head shape on BPD”. In: *American Journal of Roentgenology* 137.1 (1981), pp. 83–85.
- [10] P Schwärzler et al. “Sex-specific antenatal reference growth charts for uncomplicated singleton pregnancies at 15–40 weeks of gestation”. In: *Ultrasound in obstetrics & gynecology* 23.1 (2004), pp. 23–29.
- [11] RJM Snijders and KH Nicolaides. “Fetal biometry at 14–40 weeks’ gestation”. In: *Ultrasound in Obstetrics & Gynecology* 4.1 (1994), pp. 34–48.
- [12] Frank P Hadlock et al. “Estimation of fetal weight with the use of head, body, and femur measurements—a prospective study”. In: *American journal of obstetrics and gynecology* 151.3 (1985), pp. 333–337.
- [13] Valiollah Salari et al. “Automated measurements of fetal head from ultrasound images”. In: *Medical Imaging IV: Image Processing, Newport Beach, CA, USA* 1233 (1990), pp. 213–216.
- [14] IE Zador et al. “Ultrasound measurement of the fetal head: computer versus operator”. In: *Ultrasound in Obstetrics & Gynecology* 1.3 (1991), pp. 208–211.

- [15] George K Matsopoulos and S Marshall. "Use of morphological image processing techniques for the measurement of a fetal head from ultrasound images". In: *Pattern Recognition* 27.10 (1994), pp. 1317–1324.
- [16] Christine W Hanna and Abou Bakr M Youssef. "Automated measurements in obstetric ultrasound images". In: *Image Processing, 1997. Proceedings., International Conference on*. Vol. 3. IEEE. 1997, pp. 504–507.
- [17] Vikram Chalana et al. "Automatic fetal head measurements from sonographic images". In: *Academic radiology* 3.8 (1996), pp. 628–635.
- [18] Wei Lu, Jinglu Tan, and Randall Floyd. "Automated fetal head detection and measurement in ultrasound images by iterative randomized Hough transform". In: *Ultrasound in medicine & biology* 31.7 (2005), pp. 929–936.
- [19] Ahmet Ekin. "Feature-based brain mid-sagittal plane detection by RANSAC". In: *Signal Processing Conference, 2006 14th European*. IEEE. 2006, pp. 1–4.
- [20] Sylvain Prima, Sébastien Ourselin, and Nicholas Ayache. "Computation of the mid-sagittal plane in 3-D brain images". In: *IEEE Transactions on Medical Imaging* 21.2 (2002), pp. 122–138.
- [21] E Varvarigos et al. "Transabdominal multiplanar scan vs. B-mode in examination of fetal physiologic brain sagittal scan". In: *Ultrasound in Obstetrics and Gynecology* 20 (2002), p. 84.
- [22] I Kapouleas et al. "Registration of three-dimensional MR and PET images of the human brain without markers." In: *Radiology* 181.3 (1991), pp. 731–739.
- [23] Larry Junck et al. "Correlation methods for the centering, rotation, and alignment of functional brain images." In: *Journal of nuclear medicine: official publication, Society of Nuclear Medicine* 31.7 (1990), pp. 1220–1226.
- [24] Satoshi Minoshima et al. "An automated method for rotational correction and centering of three-dimensional functional brain images". In: *Journal of Nuclear Medicine* 33.8 (1992), pp. 1579–1585.
- [25] Marijn E Brummer. "Hough transform detection of the longitudinal fissure in tomographic head images". In: *IEEE Transactions on Medical Imaging* 10.1 (1991), pp. 74–81.
- [26] Babak A Ardekani et al. "Automatic detection of the mid-sagittal plane in 3-D brain images". In: *IEEE transactions on medical imaging* 16.6 (1997), pp. 947–952.
- [27] Alexander V Tuzikov, Olivier Colliot, and Isabelle Bloch. "Brain symmetry plane computation in MR images using inertia axes and optimization". In: *Pattern Recognition, 2002. Proceedings. 16th International Conference on*. Vol. 1. IEEE. 2002, pp. 516–519.
- [28] G Pilu et al. "Sonography of fetal agenesis of the corpus callosum: a survey of 35 cases". In: *Ultrasound in Obstetrics & Gynecology* 3.5 (1993), pp. 318–329.
- [29] Claude d'Ercole et al. "Prenatal diagnosis of fetal corpus callosum agenesis by ultrasonography and magnetic resonance imaging". In: *Prenatal diagnosis* 18.3 (1998), pp. 247–253.
- [30] Alfred B Kurtz et al. "Ultrasound criteria for in utero diagnosis of microcephaly". In: *Journal of Clinical Ultrasound* 8.1 (1980), pp. 11–16.

- [31] NS Den Hollander et al. "Congenital microcephaly detected by prenatal ultrasound: genetic aspects and clinical significance". In: *Ultrasound in Obstetrics & Gynecology* 15.4 (2000), pp. 282–287.
- [32] Robin Smith, John C Leonidas, and J Maytal. "The value of head ultrasound in infants with macrocephaly". In: *Pediatric radiology* 28.3 (1998), pp. 143–146.
- [33] Albert E Chudley et al. "Fetal alcohol spectrum disorder: Canadian guidelines for diagnosis". In: *Canadian Medical Association Journal* 172.5 suppl (2005), S1–S21.
- [34] Gideon Koren et al. "Fetal alcohol spectrum disorder". In: *Canadian Medical Association Journal* 169.11 (2003), pp. 1181–1185.
- [35] JFRL Bertrand, LL Floyd, Mary Kate Weber, et al. "Fetal alcohol syndrome. Guidelines for referral and diagnosis". In: (2005).
- [36] Robert J Sokol, Virginia Delaney-Black, and Beth Nordstrom. "Fetal alcohol spectrum disorder". In: *Jama* 290.22 (2003), pp. 2996–2999.
- [37] Sarah N Mattson, Terry L Jernigan, and Edward P Riley. "MRI and prenatal alcohol exposure: Images provide insight into FAS". In: *Alcohol Research and Health* 18.1 (1994), p. 49.
- [38] Sarah L Archibald et al. "Brain dysmorphology in individuals with severe prenatal alcohol exposure". In: *Developmental medicine and child neurology* 43.3 (2001), pp. 148–154.
- [39] Edward P Riley et al. "Abnormalities of the corpus callosum in children prenatally exposed to alcohol". In: *Alcoholism: Clinical and Experimental Research* 19.5 (1995), pp. 1198–1202.
- [40] Fred L Bookstein et al. "Midline corpus callosum is a neuroanatomical focus of fetal alcohol damage". In: *The Anatomical Record* 269.3 (2002), pp. 162–174.
- [41] PL Hilpert, BE Hall, and AB Kurtz. "The atria of the fetal lateral ventricles: a sonographic study of normal atrial size and choroid plexus volume." In: *AJR. American journal of roentgenology* 164.3 (1995), pp. 731–734.
- [42] E Albert Reece et al. "Fetal cerebellar growth unaffected by intrauterine growth retardation: a new parameter for prenatal diagnosis". In: *American journal of obstetrics and gynecology* 157.3 (1987), pp. 632–638.
- [43] Essam A Elgamal et al. "Agenesis of the corpus callosum associated with spinal open neural tube defect". In: *Saudi medical journal* 35.Suppl 1 (2014), S57.
- [44] Wesley Lee et al. "Transverse cerebellar diameter: a useful predictor of gestational age for fetuses with asymmetric growth retardation". In: *American journal of obstetrics and gynecology* 165.4 (1991), pp. 1044–1050.
- [45] Reuven Achiron and Anat Achiron. "Development of the human fetal corpus callosum: a high-resolution, cross-sectional sonographic study". In: *Ultrasound in obstetrics & gynecology* 18.4 (2001), pp. 343–347.
- [46] Selami Serhatlioglu et al. "Sonographic measurement of the fetal cerebellum, cisterna magna, and cavum septum pellucidum in normal fetuses in the second and third trimesters of pregnancy". In: *Journal of clinical ultrasound* 31.4 (2003), pp. 194–200.

- [47] Gustavo Carneiro et al. “Semantic-based indexing of fetal anatomies from 3-D ultrasound data using global/semi-local context and sequential sampling.” In: *CVPR*. 2008.
- [48] A Namburete, Bahbib Rahmatullah, and J Alison Noble. “Nakagami-based choroid plexus detection in fetal ultrasound images using adaboost”. In: *MIUA*. 2012, pp. 31–36.
- [49] Michal Sofka et al. “Automatic detection and measurement of structures in fetal head ultrasound volumes using sequential estimation and integrated detection network (IDN)”. In: *IEEE TMI* 33.5 (2014), pp. 1054–1070.
- [50] Synnøve Lian Johnsen et al. “Longitudinal reference charts for growth of the fetal head, abdomen and femur”. In: *European Journal of Obstetrics & Gynecology and Reproductive Biology* 127.2 (2006), pp. 172–185.
- [51] ASM Vinkesteyn, PGH Mulder, and JW Wladimiroff. “Fetal transverse cerebellar diameter measurements in normal and reduced fetal growth”. In: *Ultrasound in obstetrics & gynecology* 15.1 (2000), pp. 47–51.
- [52] Haim Azhari. “Appendix A: typical acoustic properties of tissues”. In: *Basics of Biomedical Ultrasound for Engineers* (2010), pp. 313–314.
- [53] Mohammad Yaqub et al. “Volumetric segmentation of key fetal brain structures in 3D ultrasound”. In: *International Workshop on Machine Learning in Medical Imaging*. Springer. 2013, pp. 25–32.
- [54] Benjamin Gutierrez-Becker et al. “Automatic segmentation of the fetal cerebellum on ultrasound volumes, using a 3D statistical shape model”. In: *Medical & biological engineering & computing* 51.9 (2013), pp. 1021–1030.
- [55] Mark J Gooding, Stephen Kennedy, and J Alison Noble. “Volume segmentation and reconstruction from freehand three-dimensional ultrasound data with application to ovarian follicle measurement”. In: *Ultrasound in medicine & biology* 34.2 (2008), pp. 183–195.
- [56] Wei Hong et al. “Database-guided simultaneous multi-slice 3D segmentation for volumetric data”. In: *Computer Vision–ECCV 2006*. Springer, 2006, pp. 397–409.
- [57] Piotr A Habas et al. “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.
- [58] Ali Gholipour et al. “Multi-atlas multi-shape segmentation of fetal brain MRI for volumetric and morphometric analysis of ventriculomegaly”. In: *NeuroImage* 60.3 (2012), pp. 1819–1831.
- [59] Zhuowen Tu and Xiang Bai. “Auto-context and its application to high-level vision tasks and 3d brain image segmentation”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32.10 (2010), pp. 1744–1757.
- [60] Simon Eskildsen et al. “BEaST: Brain extraction using multiresolution nonlocal segmentation.” In: *MICCAI Workshop on Multi-Atlas Labeling and Statistical Fusion*. 2011, pp. 97–108.
- [61] Pierrick Coupé et al. “An optimized blockwise nonlocal means denoising filter for 3-D magnetic resonance images”. In: *IEEE transactions on medical imaging* 27.4 (2008), pp. 425–441.

- [62] Robert Wright et al. “Age dependent fetal MR segmentation using manual and automated approaches”. In: *MICCAI workshop on Perinatal and Paediatric Imaging*. 2012, pp. 97–104.
- [63] Jeanne A Haimovici et al. “Clinical significance of isolated enlargement of the cisterna magna (> 10 mm) on prenatal sonography.” In: *Journal of ultrasound in medicine* 16.11 (1997), pp. 731–734.
- [64] Antonio Criminisi et al. “Regression forests for efficient anatomy detection and localization in CT studies”. In: *International MICCAI Workshop on Medical Computer Vision*. Springer. 2010, pp. 106–117.
- [65] Xiangrong Zhou et al. “Automatic organ localizations on 3D CT images by using majority-voting of multiple 2D detections based on local binary patterns and Haar-like features”. In: *SPIE Medical Imaging*. International Society for Optics and Photonics. 2013, 86703A–86703A.
- [66] Xiaoguang Lu, Daguang Xu, and David Liu. “Robust 3D Organ Localization with Dual Learning Architectures and Fusion”. In: *International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis*. Springer. 2016, pp. 12–20.
- [67] Bob de Vos et al. “ConvNet-Based Localization of Anatomical Structures in 3D Medical Images”. In: *IEEE Transactions on Medical Imaging* (2017).
- [68] Christian F Baumgartner et al. “SonoNet: Real-Time Detection and Localisation of Fetal Standard Scan Planes in Freehand Ultrasound”. In: *IEEE transactions on medical imaging* 36.11 (2017), pp. 2204–2215.
- [69] Florin C Ghesu et al. “Marginal space deep learning: efficient architecture for volumetric image parsing”. In: *IEEE Transactions on Medical Imaging* 35.5 (2016), pp. 1217–1228.
- [70] Paul Viola and Michael J Jones. “Robust real-time face detection”. In: *International journal of computer vision* 57.2 (2004), pp. 137–154.
- [71] Christian Wachinger, Tassilo Klein, and Nassir Navab. “Locally adaptive nakagami-based ultrasound similarity measures”. In: *Ultrasonics* 52.4 (2012), pp. 547–554.
- [72] Girshick Ross. “Fast r-cnn”. In: *IEEE international conference on computer vision*. 2015, pp. 1440–1448.
- [73] Shaoqing Ren et al. “Faster r-cnn: Towards real-time object detection with region proposal networks”. In: *Advances in neural information processing systems*. 2015, pp. 91–99.
- [74] Kaiming He et al. “Mask r-cnn”. In: *Computer Vision (ICCV), 2017 IEEE International Conference on*. IEEE. 2017, pp. 2980–2988.
- [75] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. “Deep learning”. In: *nature* 521.7553 (2015), p. 436.
- [76] Ian Goodfellow et al. *Deep learning*. Vol. 1. MIT press Cambridge, 2016.
- [77] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. “Imagenet classification with deep convolutional neural networks”. In: *Advances in neural information processing systems*. 2012, pp. 1097–1105.

- [78] Karen Simonyan and Andrew Zisserman. “Very deep convolutional networks for large-scale image recognition”. In: *arXiv preprint arXiv:1409.1556* (2014).
- [79] C Szegedy et al. “Going deeper with convolutions, CoRR abs/1409.4842”. In: URL <http://arxiv.org/abs/1409.4842> (2014).
- [80] Barbara Zitova and Jan Flusser. “Image registration methods: a survey”. In: *Image and vision computing* 21.11 (2003), pp. 977–1000.
- [81] Sayan Nag. “Image Registration Techniques: A Survey”. In: *arXiv preprint arXiv:1712.07540* (2017).
- [82] Ardeshir Goshtasby, George C Stockman, and Carl V Page. “A region-based approach to digital image registration with subpixel accuracy”. In: *IEEE Transactions on Geoscience and Remote Sensing* 3 (1986), pp. 390–399.
- [83] Simon Moss and Edwin R Hancock. “Multiple line-template matching with the EM algorithm”. In: *Pattern Recognition Letters* 18.11-13 (1997), pp. 1283–1292.
- [84] Wen-Hao Wang and Yung-Chang Chen. “Image registration by control points pairing using the invariant properties of line segments”. In: *Pattern Recognition Letters* 18.3 (1997), pp. 269–281.
- [85] J Alison Noble. “Finding corners”. In: *Image and vision computing* 6.2 (1988), pp. 121–128.
- [86] William K Pratt. “Correlation techniques of image registration”. In: *IEEE transactions on Aerospace and Electronic Systems* 3 (1974), pp. 353–358.
- [87] Daniel I Barnea and Harvey F Silverman. “A class of algorithms for fast digital image registration”. In: *IEEE transactions on Computers* 100.2 (1972), pp. 179–186.
- [88] Colin Studholme, Derek LG Hill, and David J Hawkes. “An overlap invariant entropy measure of 3D medical image alignment”. In: *Pattern recognition* 32.1 (1999), pp. 71–86.
- [89] J Michael Fitzpatrick and Jay B West. “The distribution of target registration error in rigid-body point-based registration”. In: *IEEE transactions on medical imaging* 20.9 (2001), pp. 917–927.
- [90] Ardeshir Goshtasby. “Piecewise linear mapping functions for image registration”. In: *Pattern Recognition* 19.6 (1986), pp. 459–466.
- [91] JB Antoine Maintz and Max A Viergever. “A survey of medical image registration”. In: *Medical image analysis* 2.1 (1998), pp. 1–36.
- [92] Piotr A Habas et al. “A spatiotemporal atlas of MR intensity, tissue probability and shape of the fetal brain with application to segmentation”. In: *Neuroimage* 53.2 (2010), pp. 460–470.
- [93] François Rousseau et al. “A novel approach to high resolution fetal brain MR imaging”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2005, pp. 548–555.
- [94] Colin Studholme, Derek LG Hill, and David J Hawkes. “An overlap invariant entropy measure of 3D medical image alignment”. In: *Pattern recognition* 32.1 (1999), pp. 71–86.

- [95] Maria Kuklisova-Murgasova et al. “Reconstruction of fetal brain MRI with intensity matching and complete outlier removal”. In: *Medical image analysis* 16.8 (2012), pp. 1550–1564.
- [96] Hessam Sokooti et al. “Nonrigid image registration using multi-scale 3D convolutional neural networks”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2017, pp. 232–239.
- [97] Bob D de Vos et al. “End-to-end unsupervised deformable image registration with a convolutional neural network”. In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer, 2017, pp. 204–212.
- [98] Wolfgang Wein et al. “Automatic CT-ultrasound registration for diagnostic imaging and image-guided intervention”. In: *Medical image analysis* 12.5 (2008), pp. 577–585.
- [99] Aris T Papageorghiou et al. “International standards for fetal growth based on serial ultrasound measurements: the Fetal Growth Longitudinal Study of the INTERGROWTH-21 st Project”. In: *The Lancet* 384.9946 (2014), pp. 869–879.
- [100] J Villar et al. “The objectives, design and implementation of the INTERGROWTH-21st Project”. In: *BJOG: An International Journal of Obstetrics & Gynaecology* 120.s2 (2013), pp. 9–26.
- [101] David Jacobs. “Correlation and convolution”. In: *Class Notes for CMSC* 426 (2005).
- [102] Qingde Li and John G Griffiths. “Least squares ellipsoid specific fitting”. In: *Geometric modeling and processing, 2004. proceedings.* IEEE. 2004, pp. 335–340.
- [103] Jack L Lancaster et al. “A modality-independent approach to spatial normalization of tomographic images of the human brain”. In: *Human brain mapping* 3.3 (1995), pp. 209–223.
- [104] Christopher P Bridge and J Alison Noble. “Object localisation in fetal ultrasound images using invariant features”. In: *Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on.* IEEE. 2015, pp. 156–159.
- [105] Michael Felsberg and Gerald Sommer. “The monogenic signal”. In: *IEEE Transactions on Signal Processing* 49.12 (2001), pp. 3136–3144.
- [106] Martin A Fischler and Robert C Bolles. “Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography”. In: *Readings in computer vision*. Elsevier, 1987, pp. 726–740.
- [107] Ana IL Namburete, Richard V Stebbing, and J Alison Noble. “Diagnostic plane extraction from 3D parametric surface of the fetal cranium”. In: *Medical Image Understanding and Analysis (MIUA)* (2014), pp. 27–32.
- [108] John C Gower. “Generalized procrustes analysis”. In: *Psychometrika* 40.1 (1975), pp. 33–51.
- [109] Jiri Matas et al. “Robust wide-baseline stereo from maximally stable extremal regions”. In: *IMAGE VISION COMPUT* 22.10 (2004), pp. 761–767.
- [110] Carlos Arteta et al. “Learning to detect cells using non-overlapping extremal regions”. In: *MICCAI 2012*. Springer, 2012, pp. 348–356.

- [111] Chris Seiffert et al. “RUSBoost: A hybrid approach to alleviating class imbalance”. In: *IEEE T SYST MAN CY A* 40.1 (2010), pp. 185–197.
- [112] Beryl R Benacerraf, Bernard Harlow, and Fredric D Frigoletto. “Are choroid plexus cysts an indication for second-trimester amniocentesis?”. In: *American journal of obstetrics and gynecology* 162.4 (1990), pp. 1001–1006.
- [113] Sandro Gabrielli et al. “The clinical significance of prenatally diagnosed choroid plexus cysts”. In: *American journal of obstetrics and gynecology* 160.5 (1989), pp. 1207–1210.
- [114] PWA Goodyear et al. “Outcome in prenatally diagnosed fetal agenesis of the corpus callosum”. In: *Fetal diagnosis and therapy* 16.3 (2001), pp. 139–145.
- [115] Fred L Bookstein et al. “Corpus callosum shape and neuropsychological deficits in adult males with heavy fetal alcohol exposure”. In: *Neuroimage* 15.1 (2002), pp. 233–251.
- [116] László Sztriha. “Spectrum of corpus callosum agenesis”. In: *Pediatric neurology* 32.2 (2005), pp. 94–101.
- [117] Shannon R Turner et al. “Sonography of Fetal Choroid Plexus Cysts Detection Depends on Cyst Size and Gestational Age”. In: *Journal of ultrasound in medicine* 22.11 (2003), pp. 1219–1227.
- [118] Lee R Dice. “Measures of the amount of ecologic association between species”. In: *Ecology* 26.3 (1945), pp. 297–302.
- [119] Felix Hausdorff. *Mengenlehre*. Walter de Gruyter Berlin, 1927.
- [120] Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. “Deep inside convolutional networks: Visualising image classification models and saliency maps”. In: *Workshop at ICLR* (2014).
- [121] Kaiming He et al. “Deep residual learning for image recognition”. In: *arXiv preprint arXiv:1512.03385* (2015).
- [122] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. “U-net: Convolutional networks for biomedical image segmentation”. In: *MICCAI*. Springer. 2015, pp. 234–241.
- [123] Amir Jamaludin, Timor Kadir, and Andrew Zisserman. “SpineNet: Automatically Pinpointing Classification Evidence in Spinal MRIs”. In: *MICCAI*. Springer. 2016, pp. 166–175.
- [124] SGM Carroll et al. “Correlation of prenatal ultrasound diagnosis and pathologic findings in fetal brain abnormalities”. In: *Ultrasound in obstetrics & gynecology* 16.2 (2000), pp. 149–153.
- [125] Sarah L Archibald et al. “Brain dysmorphology in individuals with severe prenatal alcohol exposure”. In: *Developmental Medicine & Child Neurology* 43.03 (2001), pp. 148–154.
- [126] Amy L Krain and F Xavier Castellanos. “Brain development and ADHD”. In: *Clinical psychology review* 26.4 (2006), pp. 433–444.
- [127] Ana IL Namburete et al. “Learning-based prediction of gestational age from ultrasound images of the fetal brain”. In: *Medical image analysis* 21.1 (2015), pp. 72–86.

- [128] Özgün Çiçek et al. “3d U-net: Learning dense volumetric segmentation from sparse annotation”. In: *MICCAI*. MICCAI. 2016, MICCAI.
- [129] 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.
- [130] Maxim Tatarchenko, Alexey Dosovitskiy, and Thomas Brox. “Octree Generating Networks: Efficient Convolutional Architectures for High-resolution 3D Outputs”. In: *arXiv preprint arXiv:1703.09438* (2017).
- [131] A Namburete, Bahbib Rahmatullah, and J Noble. “Nakagami-based AdaBoost learning framework for detection of anatomical landmarks in 2D fetal neurosonograms”. In: *Ann. Br. Mach. Vision Assoc.(BMVA)* 2 (2013), pp. 1–16.
- [132] Ruobing Huang et al. “Automated Mid-sagittal Plane Selection for Corpus Callosum visualization in 3D Fetal Ultrasound images”. In: *MIUA* (2015).
- [133] Jonathan Long, Evan Shelhamer, and Trevor Darrell. “Fully convolutional networks for semantic segmentation”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015, pp. 3431–3440.
- [134] Tijmen Tieleman and G Hinton. “Lecture 6.5-RMSProp, COURSERA: Neural networks for machine learning”. In: *University of Toronto, Tech. Rep* (2012).
- [135] Fei Liu et al. “Development of the human fetal cerebellum in the second trimester: a post mortem magnetic resonance imaging evaluation”. In: *Journal of anatomy* 219.5 (2011), pp. 582–588.
- [136] Selami Serhatlioglu et al. “Sonographic measurement of the fetal cerebellum, cisterna magna, and cavum septum pellucidum in normal fetuses in the second and third trimesters of pregnancy”. In: *Journal of clinical ultrasound* 31.4 (2003), pp. 194–200.
- [137] Resul Arisoy and Murat Yayla. “Nomogram of fetal cisterna magna width at 15-24th gestational weeks”. In: *Perinatal Journal* (2010).
- [138] Vincent Vanhoucke. “Lecture notes in Deep Learning by Google”. June 2016.
- [139] Carole H Sudre et al. “Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations”. In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer, 2017, pp. 240–248.
- [140] Olga Russakovsky and Li Fei-Fei. “Attribute learning in large-scale datasets”. In: *European Conference on Computer Vision*. Springer. 2010, pp. 1–14.
- [141] Jean Talairach and Pierre Tournoux. “Co-planar stereotaxic atlas of the human brain. 3-Dimensional proportional system: an approach to cerebral imaging”. In: (1988).
- [142] Alan C Evans et al. “3D statistical neuroanatomical models from 305 MRI volumes”. In: *Nuclear Science Symposium and Medical Imaging Conference, 1993., 1993 IEEE Conference Record*. IEEE. 1993, pp. 1813–1817.
- [143] Vladimir Fonov et al. “Unbiased average age-appropriate atlases for pediatric studies”. In: *Neuroimage* 54.1 (2011), pp. 313–327.

- [144] François Tadel et al. “Brainstorm: a user-friendly application for MEG/EEG analysis”. In: *Computational intelligence and neuroscience* 2011 (2011), p. 8.
- [145] A James Barkovich and BO Kjos. “Normal postnatal development of the corpus callosum as demonstrated by MR imaging.” In: *American Journal of Neuroradiology* 9.3 (1988), pp. 487–491.
- [146] G Malinger and H Zakut. “The corpus callosum: normal fetal development as shown by transvaginal sonography.” In: *AJR. American journal of roentgenology* 161.5 (1993), pp. 1041–1043.
- [147] Giuseppe Rizzo et al. “Assessment of corpus callosum biometric measurements at 18 to 32 weeks’ gestation by 3-dimensional sonography”. In: *Journal of Ultrasound in Medicine* 30.1 (2011), pp. 47–53.
- [148] G Pilu et al. “Diagnosis of midline anomalies of the fetal brain with the three-dimensional median view”. In: *Ultrasound in obstetrics & gynecology* 27.5 (2006), pp. 522–529.
- [149] W Plasencia et al. “Assessment of the corpus callosum at 20–24 weeks’ gestation by three-dimensional ultrasound examination”. In: *Ultrasound in obstetrics & gynecology* 30.2 (2007), pp. 169–172.
- [150] Y Gao, Mohammad Ali Maraci, and J Alison Noble. “Describing ultrasound video content using deep convolutional neural networks”. In: *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on*. IEEE. 2016, pp. 787–790.
- [151] Diego Marcos, Michele Volpi, and Devis Tuia. “Learning rotation invariant convolutional filters for texture classification”. In: *Pattern Recognition (ICPR), 2016 23rd International Conference on*. IEEE. 2016, pp. 2012–2017.
- [152] Gong Cheng, Peicheng Zhou, and Junwei Han. “Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images”. In: *IEEE Transactions on Geoscience and Remote Sensing* 54.12 (2016), pp. 7405–7415.
- [153] Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al. “Spatial transformer networks”. In: *Advances in Neural Information Processing Systems*. 2015, pp. 2017–2025.
- [154] Misha Denil et al. “Learning where to attend with deep architectures for image tracking”. In: *Neural computation* 24.8 (2012), pp. 2151–2184.
- [155] Diederik Kingma and Jimmy Ba. “Adam: A method for stochastic optimization”. In: *arXiv preprint arXiv:1412.6980* (2014).
- [156] Sergey Ioffe and Christian Szegedy. “Batch normalization: Accelerating deep network training by reducing internal covariate shift”. In: *International Conference of Machine Learning* (2015).