Automated 3D Ultrasound Image Analysis for First Trimester Assessment of Fetal Health

Hosuk Ryou

Mansfield College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2017

Acknowledgements

Firstly, I would like to express my sincere gratitude to my supervisor, Prof. Alison Noble for guiding me throughout this project. Your motivation and immense knowledge guided me in all the time of my research. Most of all, you taught me what kind of mental attitude I must have as the researcher and how I should approach the research. My gratitude also goes to my collaborators, Prof. Aris Papageorghiou and Dr. Angelo Cavallaro for providing the immense clinical knowledge to this project.

I would also like to thank my colleagues at the Biomedical Image Analysis Laboratory, in particular Dr. Mohammad Yaqub, Weidi Xie, Dr. Ali Maraci and Dr. Ana Namburete. Thank you for taking your time to have discussion and share the ideas regarding my project. Above all, it was you who made me to enjoy my DPhil experience. I was lucky to have such wonderful people as my labmates.

Thank you to my dearest friend, Eric Ahn for your support and enjoyable life throughout my DPhil.

I also want to thank my love Hyunjin Kim who always supported me and believed in me. Without you, it would be much harder for me to go forward. From deepest of my heart, thank you for your love, support and encouragement.

Finally, I would like to thank my parents, Hongsun Ryou and Chinok Lee for all your love and support. I cannot describe how much I am grateful to you in text. You have been always my life- and spiritual mentor, and encouraged me at every steps I took. Without you, I could not imagine how I went through all the difficult times I had. It was you who made me to achieve this far. I dedicate this work for you.

Abstract

The first trimester fetal ultrasound scan is important to confirm fetal viability, to estimate the gestational age of the fetus and to detect any fetal anomalies early in pregnancy. In particular, there is growing recent clinical interest in shifting aspects of fetal health screening typically performed at the 18-22 week fetal ultrasound scan to the first trimester ultrasound scan for the benefit of the health of the mother and fetus and potentially to reduce cost within healthcare systems if pregnancy problems are picked up early. Ideally, this assessment would be based on 3D ultrasound, given the object of interest (the fetus) fits within a typical transducer field-of-view at this gestational age. The challenge then becomes one of diagnostic plane finding in a volume.

In this thesis, we study, to our knowledge for the first time, how to automate the analysis of first trimester 3D fetal volume screening scans for the purpose of performing a basic fetal health assessment. The developed methods have been integrated into a software application with an interactive user interface to provide clinicians with a fully automatic tool for first trimester scan assessment.

Specifically, this thesis first develops a deep learning algorithms called Fully Convolutional Network-based (FCN) approach for the localization of the whole fetus within a 3D volumetric ultrasound scan. A comparison of the developed approach is presented with a recently published hand-crafted machine learning approach. The proposed FCN-based algorithm was found to have a higher localization accuracy of the whole fetus (Intersection over Union (IoU) of 0.84 compared to 0.74 using Random Forests).

Next, a multi-task FCN network architecture was designed to perform both sagittal plane extraction and the segmentation into head, body and the legs. This achieved 98.9% classification accuracy and better segmentation accuracy compared to a single-task FCN network (89.4% compared to 85.8%). Based on the localization and the segmentation result, we were able to focus attention on the fetal parts in the axial view and by using another FCN-based approach, we achieved a segmentation accuracy of 95.0% and 96.3% for the brain and abdomen, respectively.

Based on the whole fetus segmentation result, the biometric plane for the brain and abdomen can be extracted including the correction of orientation misalignment. Having both the segmentation and corrected orientation, automatic plane extraction, object orientation estimation and measurement can be performed. Automatic analysis is shown to give similar results to manual assessment.

Finally, automatic limb assessment is considered. This is the most challenging part of our research due to the variability in position and appearance of the objects of interest and has not been attempted before. Again, an FCN-based solution is proposed. The developed network architecture has a segmentation accuracy of 72.3%, 76.5%, 81.8% and 82.3% for left and right arms, and left and right legs, respectively and detection rate of 0.65, 0.50, 0.55 and 0.35 for left and right arms, and left and right legs, respectively. Our results are promising but suggest further technical research needs to be done on this topic before clinical utility can be explored.

Contents

List of Figures

List of Tables

List of Abbreviations

2D, 3D . . . . . Two- or three-dimensional
AC . . . . . Abdomen Circumference
ANN . . . . . Artificial Neural Network
APAD . . . . . Anterior-Posterior Diameter
BoVW . . . . . Bag of Visual Words
BN . . . . . Batch Normalization
BPD . . . . . Biparietal Diameter
CFCN . . . . . Cascaded Fully Convolutional Network
CNN . . . . . Convolutional Neural Network
CRL . . . . . Crown-Rump Length
CT . . . . . Computed Tomography
FC . . . . . Fully Connected
FCN . . . . . Fully Convolutional Network
GA . . . . . Gestational Age
HC . . . . . Head Circumference
IoU . . . . . Intersection over Union
ISUOG . . . . . International Society of Ultrasound in Obstetrics and Gynecology
MRI . . . . . Magnetic Resonance Imaging
OFD . . . . . Occipitofrontal Diameter
PCA . . . . . Principal component analysis
ReLU . . . . . Rectified Linear Unit
RF . . . . . Random Forest
SIFT . . . . . Scale Invariant Feature Transform
SRF . . . . . Structured Random Forest
TAD . . . . . Transverse Abdominal Diameter
US . . . . . Ultrasound

1 Introduction

1.1 Motivation

Obstetrics is the medical specialty that deals with pregnancy and childbirth. Since ultrasound has been introduced for screening fetal anomalies in the late 1970s in the UK [1], medical imaging has become a key component of pregnancy care because of its ability to visualize the fetus inside a mother to screen for abnormalities and to check healthy fetal growth. Fetal growth can be affected by many factors such as nutrition, smoking, genetic factors and even the social enviornment [2]. Because of these factors, it is important to monitor the fetus during a pregnancy. This is especially important during early fetal development. In the first trimester, the miscarriage rate is higher than for later trimesters [3]. Miscarriage rate is over 20% in early pregnancy compared to less than 10% in later pregnancy [4]. This is also the stage when a clinician will confirm fetal viability and begin to monitor fetal growth rate [5]. To do this, clinicians use ultrasound imaging to visualize anatomical planes for anatomical assessment and biometric measurements. However, the quality of ultrasound decision-making is heavily dependent on sonographer skill. To reduce this dependency, there is active current interest in medical image analysis research aimed at developing tools to assist human decision-making.

The main goal of this thesis is to research the development of new automated image analysis tools to assist first trimester fetal assessment from 3D ultrasound. This includes

the visualization of the fetal anatomy such as the head, abdomen and limbs, and based on the visualized fetal parts, to perform automated biometry. Due to the recent success of the deep learning methods in both computer vision and medical imaging, this thesis investigates FCN-based approaches for the fetal assessment based on the 3D fetal volume.

1.2 Contribution

The main contributions of this thesis are four fold. Specifically, we present,

1. The first study to propose a FCN-based approach to perform the automated segmentation of the whole fetus into three classes (the head, the body and the legs) within a 3D first trimester ultrasound volume. This is challenging because the boundary between the back of the fetus and non-fetal tissue is not clear, and the fetus might be small compared to the non-fetal parts.

2. The first attempt to automatically correct for orientation misalignment to improve the quality of the head and abdomen images for better automatic anatomical assessment and biometry measurements.

3. The first attempt at automatic fetal limb assessment from a 3D ultrasound volume.

4. The algorithms for each fetal assessments are integrated into a software system for first trimester fetal biometry (Figure 1.1).

Validation is used in each case to support the proposed solution.

1.3 Thesis outline

The thesis is organized as follows: **Chapter 1** explains the motivation for the research and summarizing the contribution of the thesis. **Chapter 2** reviews relevant clinical and technical literature. **Chapter 3** describes the automatic localization of the whole fetus using a two stage machine learning method as a pre-processing step for the subsequent analysis. **Chapter 4** describes how a Convolutional Neural Network (CNN) architecture labels each slices into one of three classes: fetal head, torso and non-fetal parts. Based

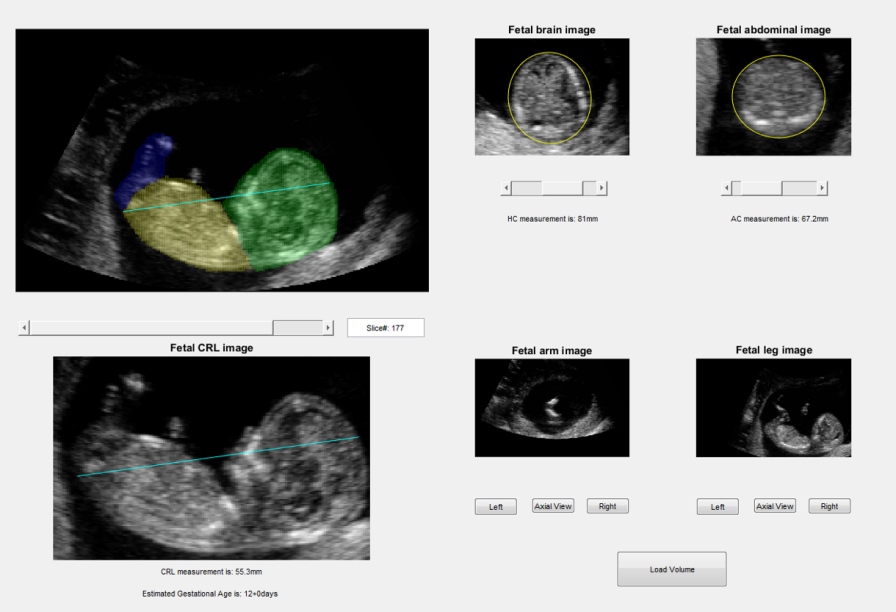


Figure 1.1: Integrated software interface to provide the clinicians the automatic first trimester fetal biometry.

on the classification result, the best plane for the head and for the abdomen is extracted using regression. **Chapter 5** first explains how a FCN-based approach is applied to do the same tasks as Chapter 3 and 4 with faster computation and improved localization. Then using the similar algorithms, fetal biometric assessment for the head, abdomen and limbs is achieved with automatic orientation correction of biometry planes. The developed methods have been integrated into a software user interface (see Figure 1.1). Finally, **Chapter 6** summarizes the contributions of this thesis and discusses possible future work.

1.4 Publications to-date

Related to Chapter 3 and 4:

Ryou, H., Yaqub, M., Cavallaro, A., Roseman, F., Papageorghiou, A, Noble, J. A., "Automated 3D Ultrasound Biometry Planes Extraction for First Trimester Fetal Assessment," *International Workshop on Machine Learning in Medical Imaging*, pp. 196-204, 2016.

2
Literature Review

This chapter reviews relevant clinical and technical literature. Specifically, Section 2.1 provides an overview of ultrasound imaging. Section 2.2 describes the first trimester ultrasound scan. Section 2.3 overviews relevant machine learning methods in the computer vision literature and how these techniques have been applied in the medial imaging. Conclusions are presented in Section 2.4.

2.1 Ultrasound Imaging

2.1.1 Overview of Ultrasound Imaging

Ultrasound has been used as an imaging modality for over 40 years. It has the advantage of being non-invasive and is commonly used in diagnosing a number of medical conditions, including heart disease [6], [7] and breast cancer [8], [9], [10]. The use of ultrasound imaging in fetal health assessment was one of the early applications [11], [5], [1], [12] and this method is used routinely to visualize different fetal structures such as brain, heart , abdomen, kidneys and spine [11], [5]. Ultrasound can also be used in image-guided interventions such as neurosurgery [13], [14].

To date, 2D ultrasound has been used as the standard method for performing routine clinical scans in the hospital setting. The fundamental mechanism of ultrasound image formation is the pulse-echo. Pulse-echo is a diagnostic technique in which short-duration

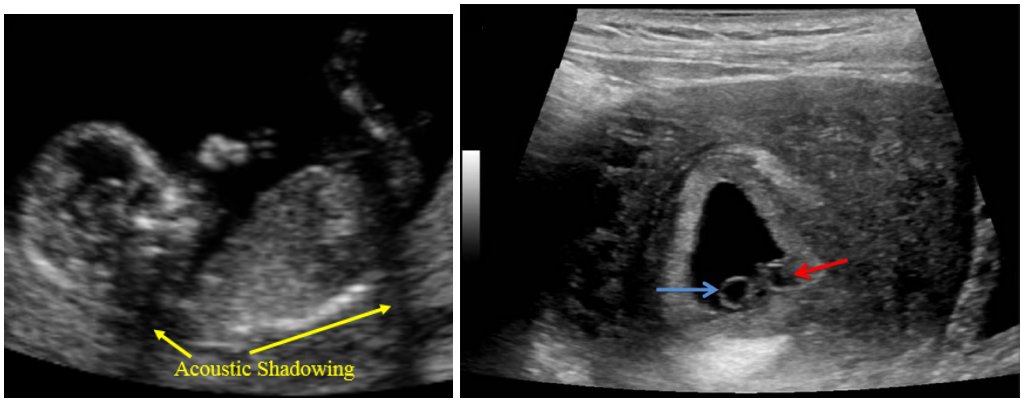
ultrasound pulses are transmitted into the structures, and echo signals resulting from scattering and reflection are detected and displayed [15], [16].

Advantages over other imaging modalities as fetal assessment

Ultrasound has several important advantages over other imaging modalities such as CT and X-ray, especially in obstetrics. Firstly, ultrasound scanning has a relatively low cost in terms of both equipment price and service running which is an important consideration for any clinical screening unit. It can also be done in real-time and it is portable. The most important advantage is its safety. Unlike X-ray or CT, ultrasound does not use any ionising radiation which might cause high risk to both mother and fetus. MRI, on the other hand, use electromagnatic (EM) radiation, which is not ionising radiation. It is also emerging as another fetal imaging modality, especially for fetal central nervous system, due to its high tissue resolution and wide field of view [17], [18], [19], [20]. Although MRI is rising as a potential fetal imaging modality, fetal MRI has higher running cost, and standards do not exist for fetal MRI in terms of standardised views and protocols for detection and evaluation of fetal anomalies [18], [21]. Also, MRI still has not been approved for use in the first trimester because there are concerns about potential disturbance on fetal development due to magnetic field exposure [18], [22], and does not usually provide information additional to that obtained on ultrasound examination [23].

Disadvantages of 2D ultrasound

Although ultrasound has shown good and potential application in the medical fields, ultrasound has some well-documented limitations relative to other imaging modalities. Ultrasound suffers from various artefacts caused by various factors such as the wavelength reflection and refraction [24], [25], [26]. Reflection happens when the ultrasound wave passes from one tissue to another with different acoustic impedances and a fraction of the wave bounces back. The wave always reflect in such a way that the angle at which they approach the surface equals the angle at which they reflect off the surface. Refraction means the direction of the sound is changed because the speed of the sound is different in different media [27]. Acoustic shadowing, for example, is one of the most



(a) **(b)**

Figure 2.1: The example of ultrasound artefacts: a) acoustic shadowing and b) mirroring artefact. (a) shows the acoustic shadowing where the left shadow is caused by the nasal bone. (b) shows that a duplicate of yolk sac (red arrow) appeared adjacent to the actual yolk sac (blue arrow). (b) is obtained from Figure 6 of the work by [25]

common artefacts in 2D ultrasound. Acoustic shadowing is caused by a strong reflector (where there is a significant difference in acoustic impedance between neighbouring tissue layers) that blocks the visualization of the behind object [28]. An example of acoustic shadowing is shown in Figure 2.1 (a).

As shown in Figure 2.1 (a), acoustic shadowing is present behind the bones since the ultrasound waves cannot penetrate structures. This affects the visualization of the certain fetal anatomies because some internal organs might be missing due to the acoustic shadowing which makes the image quality worse. Another example is mirror artefact which produces duplicates of structures or wider structures in the image. The mirror artefact is due to the misregistration caused by the path of the ultrasound pulses changing direction due to either refraction or reflection [25], [29]. Because of this, the same structure appears at incorrect location along with the structure at the correct location as shown in Figure 2.1 (b) that a duplicate of yolk sac appeared adjacent to the actual yolk sac. In this particular case, the transducer received both echoes from the direct and changed path, and therefore displayed two same structures in the image. The reflection and refraction affect image appearance, hence might lead to misdiagnosis such as abnormalities or over-estimation in the measurement.

Moreover, ultrasound image quality is heavily dependent on the skill of the person (the sonographer) performing the scan. Highly skilled sonographers are in great demand. The

physicians who perform and interpret diagnostic examinations, for example, are required to perform over 300 examinations and have at least 3 years of experience according to the American College of Radiology [30]. Even when a sonographer is trained to a high level, there can be significant inter- and intra-observer acquisition variability [31] [32]. Dudley et al. [33] also mentions that ultrasound image quality is one of the main reasons for intra- and inter-observer variability (coefficient of variation in the range 2-3% [34], [35]) in measurement because the same scan of specific anatomical structures itself can have variance between different sonographers. This variability could potentially lead to a different diagnosis depending on the sonographers and because accurate measurement is especially important in fetal screening. This might yield a problem for gestational age estimation or measurement of fetal growth. Furthermore, the fetal anatomy or orientation may restrict the imaging angle that lead to the inaccessibility of the good 2D image plane necessary for diagnosis [36], [37], [38]. 3D ultrasound imaging generates a 3D representation of structures to overcome the limitation of 2D ultrasound.

2.1.2 3D ultrasound

Types of acquisition

There are three common ways of 3D ultrasound acquisition: the mechanical 3D ultrasound scanning device, the free-hand scanning and the 3D ultrasound scanning using 2D arrays [36], [37], [38], [39], [40]. Mechanical 3D ultrasound system uses motorized mechanisms to translate, tilt or rotate a 2D ultrasound transducer with a linear/1D transducer array, and then a computer acquires the series of 2D ultrasound images. Using predefined relative locations and orientations, the acquired 2D ultrasound images are then used to reconstruct 3D images without missing regions. However, the clinicians/sonographers are required to hold the relatively heavy mechanical probes stationary while acquiring images, which might lead to errors for the data acquisition [40]. Also, the transducer might needs particular mechanical motor to be used [40], which is specialised equipment. The second type of acquisition is the free-hand scanning system with no tracking. The operator holds an assembly composed of the transducer and manipulates it over the anatomy. 2D images are then digitized as the transducer is moved with appropriate speed. Although

the free-hand acquisition offers flexibility and does not require mechanical motors, there can be distortions in the data due to the probe-pressure variation and there is no guarantee that the there are no significant gaps between the images [39]. Both mechanical and free-hand 3D ultrasound systems generate a 3D ultrasound image as the transducer is moved to sweep out the anatomy, but both of these approaches limit the speed of volume acquisition [40]. To overcome these, the transducer with 2D arrays generating 3D images have been developed. The elements of 2D array transducer generate a diverging beam in pyramid shape and returned echoes detected by 2D arrays are processed to integrate 3D images in real-time. The advantages of this approach are that the transducers can remain stationary while being used to scan, and it is capable of 3D visualization of dynamic structures [38], [40]. However, this type of acquisition requires relatively high cost of the 2D transducer compared to other types of acquisition [37], [40].

The 3D ultrasound is typically displayed using the following techniques: surface- and volume rendering [41], [42] and multiplanar reformatting [43], [44], [45]. Surface rendering is the technique to display surfaces of the organs [41] and volume rendering presents a display of the entire 3D image [42]. Multiplanar reformatting provides the viewer with planar cross-sectional images extracted from the 3D image [43]. In this thesis, orthogonal-plane approach, one of the multiplanar reformatting technique, has been used in open source applications called ITK-SNAP [46] and MITK [47] as the display technique. Orthogonal planes are displayed together with three dimensional cues such as lines on each extracted plane as shown in Figure 2.2. The operator can then select and extract any planes, and move them within the 3D volume to provide the cross-sectional view at any desired location and orientation that resembles a 2D ultrasound images which is already familiar to the clinicians [48], [49].

Advantages and limitations of 3D ultrasound

The biggest advantage of 3D scanning is that it is possible to visualize the anatomical views that are very difficult to be viewed with 2D scanning [38], [39], [40]. Because of this, the typical application of 3D ultrasound is to view and diagnose the fetal parts such as the face and the limbs which can be difficult to view in 2D. For example, it has been found

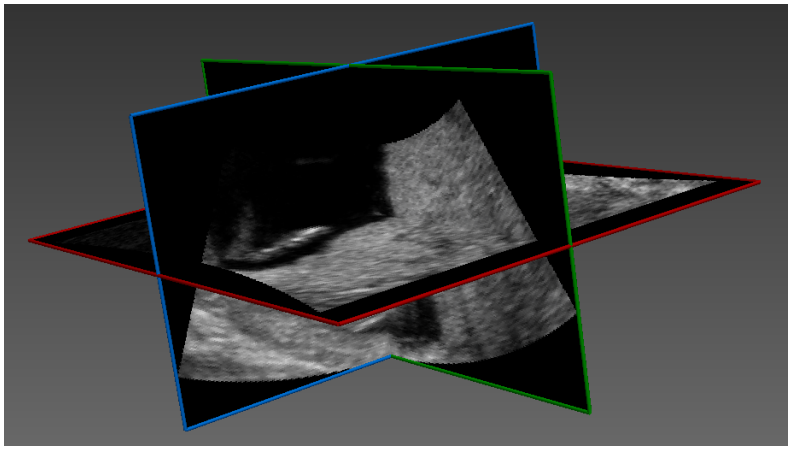


Figure 2.2: The example of orthogonal plane display, one of the multiplanar reformatting techniques.

that cleft palate can be seen more clearly in 3D than in 2D scan [50]. 3D ultrasound can also be applied to assist the diagnosis in rib abnormalities and abnormal spine curvature [51]. Acar et al. [52] also reported that utilizing 3D ultrasound allows to examine the fetal heart in real time. Dyson et al. [53] showed that using 3D ultrasound images provided additional information in 51 % of the anomalies when compared with 2D ultrasound alone. Michailidis et al. [54] reported that 3D ultrasound is a more effective and faster way to identify fetal gender. Other than fetal imaging, 3D ultrasound is used in cardiology [55], musculoskeletal tissue [56], [57], vascular imaging [58], [59] and surgical guidance [60], [61]. In cardiology, for example, the real-time 3D echocardiography is used to visualize the anatomy of the cardiac chambers entirely and to measure left ventricular volume and dynamic changes of chamber volume to diagnose congenital heart disease [55]. Also, 3D ultrasound can contribute to reduce the operator dependence in the scanning process. 2D ultrasound requires the clinician sweep the probe across the anatomy of interest while visualizing the anatomy from multiple images. 3D, however, requires only a single volume. After the acquisition of the volume, the anatomy can be visualized from the slices extracted from the volume at any angle, which enables better visualization of anatomy, and a variety of measurements can be taken from it [43], [38]. Finding good planes and acquiring good images are also possible with 2D acquisition, but it is not as reproducible because of high dependencies on sonographer [43]. Furthermore, 3D ultrasound provides more accurate and repeatable way of evaluating anatomic structures and it is more suitable

for monitoring the progression of a disease [38]. For these reason, 3D fetal ultrasound scan offers some interesting potential opportunities over 2D scanning.

However, 3D ultrasound has limitations. 3D ultrasound is prone to the same types of artefacts as 2D [24]. For example, acoustic shadowing in 3D ultrasound can present in a variety of ways because 3D ultrasound can display the image planes in orientation different from the original acquisition orientation [62], [24]. There are also artefacts that are specific to 3D ultrasound [24]. Unlike 2D ultrasound, 3D suffers from the motion effects. The fetal movement or any involuntary movements such as cardiac motion can lead to the poor visualization of the anatomy of interest [24]. Another artefact is caused by the signal attenuation for the structures away from ultrasound probe because of existence of obscuring structures nearer the probe e.g., bone [63].

2.2 Ultrasound first trimester scan

Human fetal gestation usually takes about 40 weeks and it is divided into 3 trimesters. Second trimester ultrasound scanning, which, in the UK, is typically performed between 18 and 22 weeks, is considered the optimal time window to obtain required views and to define normality [1], [11]. However, miscarriage mostly happens at the first trimester and this questions whether more research should be done to assess the pregnancy risk in this trimester. Also, there is growing recent clinical interest in shifting aspects of fetal health screening typically performed at the 18-22 week fetal ultrasound scan to the first trimester scan for the benefit of the health of the mother and fetus and potentially to reduce cost on healthcare systems if pregnancy problems are picked up early. Saltvedt et al. [64], for example, concluded that there was not significant difference between 12 and 18 week scan groups in the detection rate of fetuses with a major abnormality. Rossi and Prefumo also reported that about half of all structural malformations can be diagnosed at 11-13 week gestation [65].

The first trimester covers to week twelve, with the first trimester screening recommended between 11 and 14 weeks [5], [12]. The first trimester ultrasound scan is recommended in several clinical standards. However, there are other clinical tests to assist diagnosis. Maternal serum screening, which checks levels of pregnancy-associated hormone and

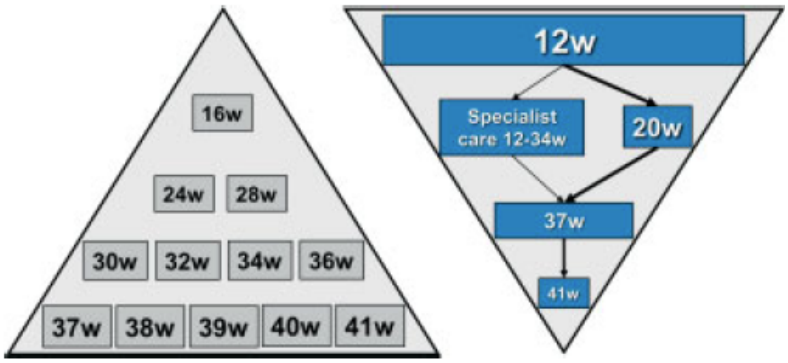


Figure 2.3: Suggested model/pyramid from [72]. In the past (left pyramid), the women had to take the first ultrasound scan at 16 weeks (second trimester), then at 24 and 28 weeks, fortnightly until 36 weeks, then weekly until delivery. Unlike the prenatal care from the past (left), the majority of women will be required to have an the ultrasound scan in the first trimester to be classified as low-risk or high-risk since most major aneuploidies can be identified at this trimester. Due to this, after the first trimester, In the low-risk group, one visit at around 20 weeks will re-evaluate the fetal anatomy and growth, and reassess risk for complications. Then at 37 weeks, another visit will assess fetal health and determine the best time and method of delivery and will be repeated at 41 weeks. For the high-risk group, the specialist care will be performed and then re-assess whether the fetus remains high-risk or low-risk in which case the intensity of their care can be reduced (The figure is obtained from Figure 1 of the work by Nicolaides [72].)

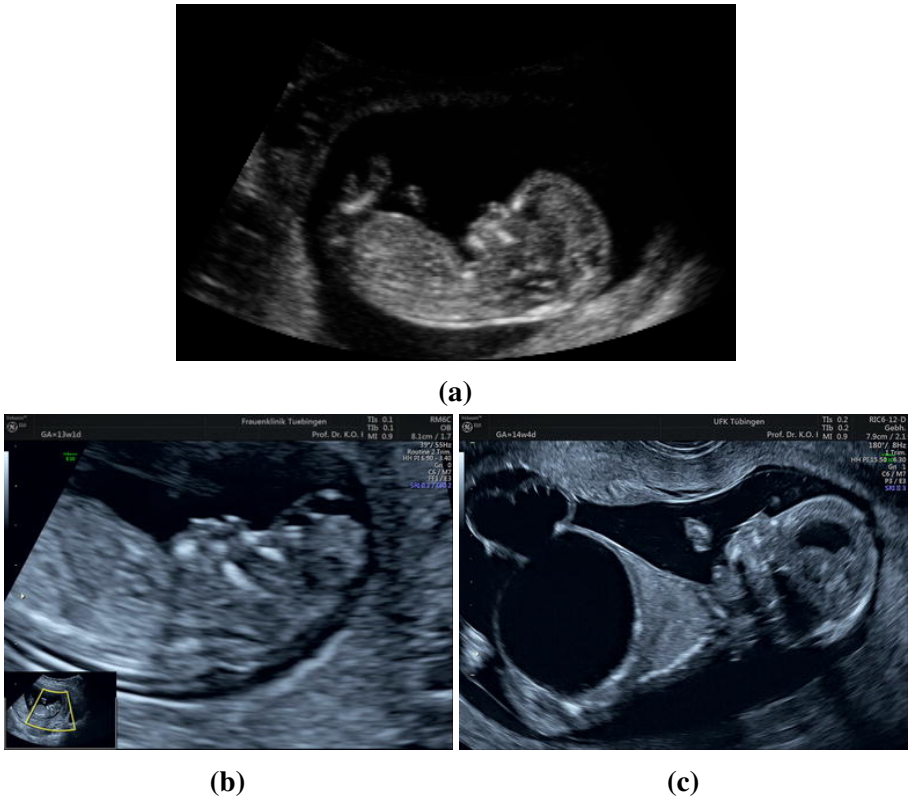
protein from the maternal blood, is performed with nuchal translucency measurement from the ultrasound scan to diagnose chromosomal abnormalities such as trisomy-21 [66], [67]. Spencer et al. [67] reported that the rate of detection of chromosomal abnormalities using both maternal serum screening and nuchal translucency measurement achieved 90% for 5% false positive rate which is far superior to the average 65% achievable by second trimester biochemical screening. Sequencing fetal DNA from the maternal blood [68] or retrieval of trophoblast cells [69] can also assist ultrasound scan to idnetify fetal aneuploidies. Other tests such as listening to fetal hearbeat via an external monitor placed on the outside of the abdomen are also performed. Early amniocentesis or chronic villus sampling [70], which a needle is inserted through the mother's abdominal wall to extract amniotic fluid to check the cells to diagnose chromosomal abnormalities, is also possible, but may have a higher risk of fetal injury [71]. The techniques involved in the prenatal tests, including ultrasound, are still in development to be able to do more and accurate prenetal diagnosis, especially in the first trimester.

Why first trimester scan?

As shown in Figure 2.3, Nicolaides [72] encouraged the new model/pyramid of prenatal care and argues that in the past (left pyramid in Figure 2.3), the first ultrasound scan had to be taken at 16 weeks (second trimester), then at 24 and 28 weeks, fortnightly until 36 weeks, then weekly until delivery. However, as the technology develops, many abnormalities can be detected in the first trimester to classify whether the fetus is at low-risk or high-risk. In the low-risk group, one visit at around 20 weeks will re-evaluate the fetal anatomy and growth, and reassess risk for complications. Then at 37 weeks, another visit will assess fetal health and determine the best time and method of delivery and will be repeated at 41 weeks. For the high-risk group, the specialist care will be performed and then re-assess whether the fetus remains high-risk or low-risk in which case the intensity of their care can be reduced. Therefore, Nicolaides [72] proposed inverted pyramid as the gestational age increases, the number of medical visits will be reduced, and due to this, the future research will inevitably expand the number of conditions that can be identified in early pregnancy and define the biomarkers that can lead to early identification of high-risk groups and prevent the disorders of pregnancy.

The purpose of the first trimester scan is to accurately establish gestational age, to confirm viability, to determine fetal number and to detect any gross fetal abnormalities [5], [73]. In the case of multiple pregnancies, the first trimester scan is also used to assess the chorionicity [5], [74], which is to check whether each fetus has a separate placenta with its own outer membane or they share the same placenta and amnioicity, which is to check the number of inner membranes that surround the fetus. Also, if the fetus is at a high risk of miscarriage due to the severe structural abnormalities, the first trimester scan offers the earlier and safer termination of the pregnancy with less economic and emotional costs [75], [76], [77], [78], [79]. Vintzileos [79], for example, concluded that the first trimester ultrasound could save society $22 million per year.

The first trimester scan is also used for the early detection of chromosomal defects/genetic syndromes [80], [81], [82], [83]. Nicolaides [83], for example, showed that the detection rate of Down's Syndrome and other aneuploidies is 85-90% with a false positive rate of 5% by the combined screening test including the analysis of the increased



(a) **(b)** **(c)**

Figure 2.4: The fetal anomalies possible to be visualized in the first trimester. (a) One of the planes extracted from the automated algorithms developed in this thesis (b) Exencephaly/Anencephaly (c) Megacystis (The figures (b) and (c) are obtained from Figure 1 and 3 of the work by Kagan et al. [90].)

thickness of the nuchal translucency measured from ultrasound. Various studies have shown that the first trimester offers the early diagnosis of non-chromosomal anomalies such as the skeletal abnormalities, cleft palate and cardiac defects [84], [85], [86], [87], [88], [89]. For example, from Syngelaki et al. [87], increased nuchal translucency was found in 35% of the fetuses with major cardiac defects and about 65% of the fetuses in the first trimester were observed compared to 15% in the second. What is common in these studies is that the detection rate depends on factors such as the acquisition protocol and sonographer skill. Also, due to the natural growth of the internal organs, some abnormalities are impossible to detect in a first trimester ultrasound. However, these studies also mention that the first trimester ultrasound scan does have potential for the early detection of key fetal anomalies.

Figure 2.4(a) shows an extracted ultrasound plane for both anatomical assessment and CRL measurement. Fetal anomalies such as the exencephaly/anencephaly and the

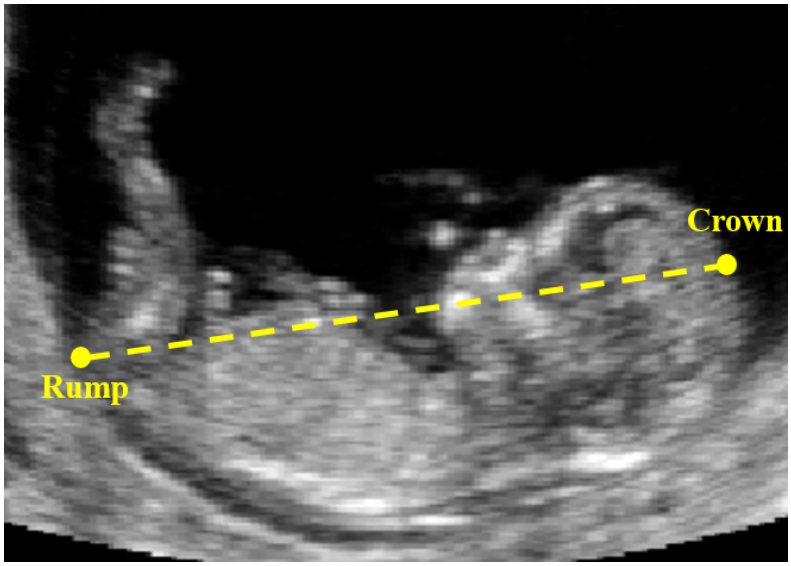


Figure 2.5: The Crown-Rump Length (CRL) measurement in a fetus in sagittal view.

megacystis can be visualized in the extracted plane as shown in Figure 2.4(a).

2.2.1 First trimester fetal assessment

According to the International Society of Ultrasound in Obstetrics and Gynecology (ISUOG) guidelines [5], the first trimester scan is used for two purposes: first-trimester fetal measurement and the assessment of fetal anatomy. Fetal measurements are used to assess the normality of fetal growth and to estimate the *gestational age* (GA). The anatomical assessment is used to check the structural anomalies.

Crown-Rump Length (CRL)

There are a number of fetal biometry measurements performed in the first trimester. One of the most important measurements is the *Crown-Rump Length* (CRL). The data used in this thesis is also based on the CRL criteria which are explained in Appendix A. The CRL is defined as the length between the endpoints of the crown and rump as shown in Figure 2.5. The CRL measurement is the major biomarker for gestational age prediction and the assessment of early fetal growth [91], [92], [93], [94]. There have been various studies for deriving CRL for the estimation of GA based on the population [95]. In this thesis, the estimation derived from based on the INTERGROWTH-21st data has been used [93],

$$GA(Days) = 40.9041 + 3.21585 \times CRL^{0.5} + 0.348956 \times CRL. \quad (2.1)$$

Anatomical Assessment

Anatomical assessment aims to check whether there are any structural abnormalities in the fetus based on visual appearance and the presence of internal structures. According to ISUOG [5], there are number of anatomical areas that make up an anatomical assessment in the first trimester scan as shown in Table 2.1. However, in current 3D ultrasound, some

Organ/Anatomical Area	Present and/or normal?
Head	Present Cranial bones Midline falx Choroid-plexus-filled ventricles
Neck	Normal appearance Nuchal translucency thickness
Chest	Symmetrical lung fields No effusions or masses
Heart	Cardiac regular activity
Abdomen	Stomach present in left upper quadrant
Abdominal wall	Normal cord insertion No umbilical defects
Extremities/Limbs	Four limbs each with three segments
Placenta	Size and texture

Table 2.1: Anatomical assessment at 11 to 13+6-week scan based on ISUOG guidelines [5]. This thesis covers the shaded organs: the head, the abdomen and the limbs

anatomical areas such as the heart and chest are difficult to visualize. In this thesis, three anatomical areas are considered: the head, the abdomen and the limbs. As shown in Table 2.1, the diagnostic image of the head needs to have a visualization of key internal structures: *cranial/skull bone*, *choroid plexus* and *midline falx*. One of the anatomical landmarks for the brain that clinicians mostly look at is the choroid plexus, which has a butterfly shape appearance as shown in Figure 2.6 (a). This is one of the most notable characteristics of the brain and makes it distinguishable from the other fetal structures for not just clinical studies, but also for automatic pattern recognition. Other landmarks like the cranial/skull bone which surround the brain are used for growth assessment. Three fetal head biometric measurements are made: the *biparietal diameter* (BPD), the *occipitofrontal diameter* (OFD) and the *head circumference* (HC). BPD is obtained by measuring the widest diameter from the outer borders of the head. OFD is measured

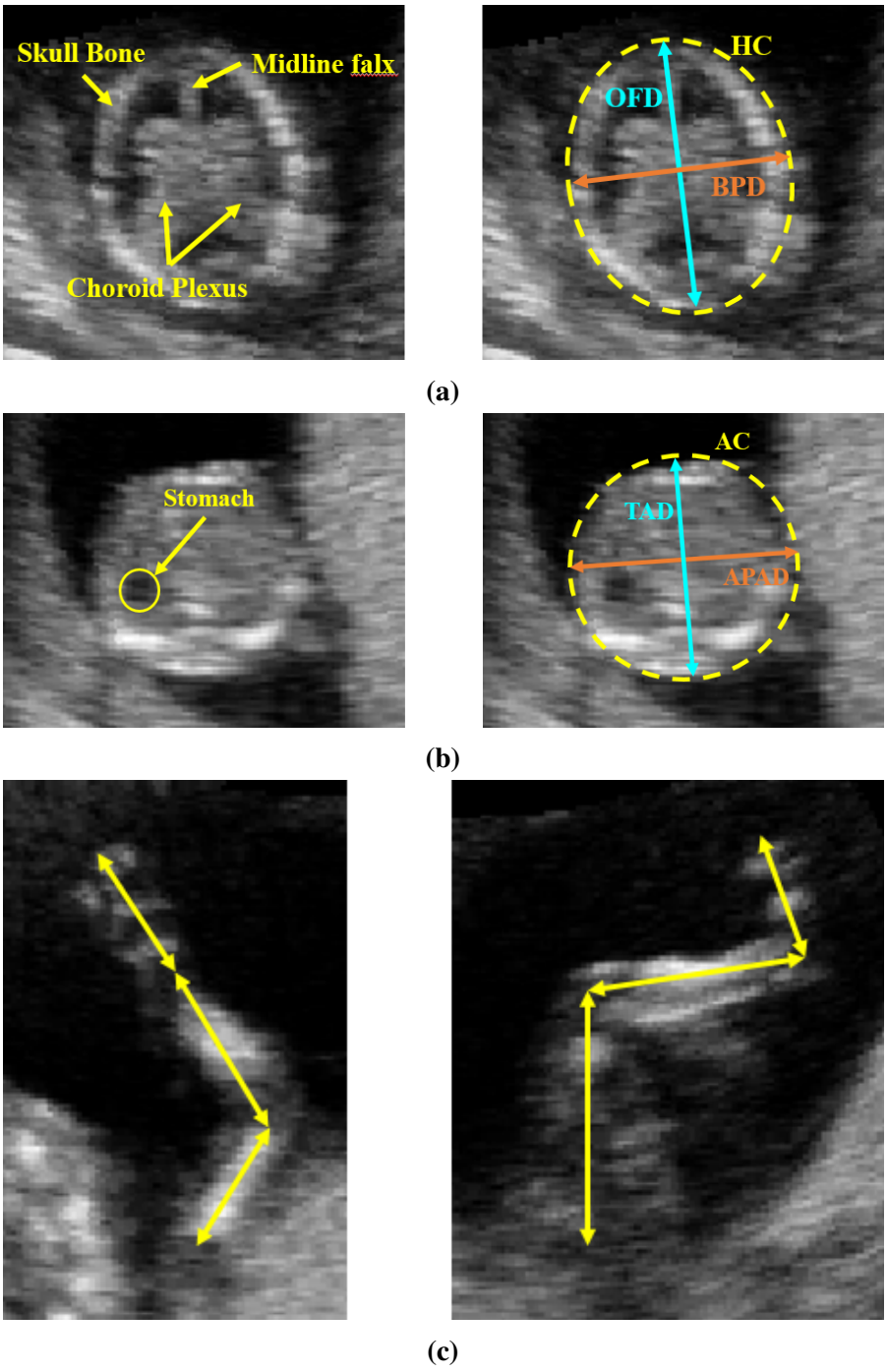


Figure 2.6: The anatomical assessment in the first trimester. (a) The head/brain image with the visualization of the internal structures: Cranial/skull bone, choroid plexus (butterly shape) and midline falx. Based on this image, the HC is measured by either draw the ellipse around the head and measure the circumference or measure OFD and BPD and obtain HC from the equation 2.2. (b) The abdomen image with the visualization of the stomach. Similar to HC, AC is measured either by the circumference of the ellipse around the abdomen or using the equation 2.3 based on the length of TAD and APAD. (c) The limb image with the visualization of three segments (upper, lower and hand/foot).

as the longest diameter from the outer borders of the head. HC is obtained either by drawing an ellipse around the head and measuring the circumference, or by deriving HC from the equation based on BPD and OFD measurement,

$$HC = \frac{\pi \times (BPD + OFD)}{2} \tag{2.2}$$

HC can also be used to estimate the GA, though it has been claimed to be more suitable for the later trimesters [96]. The abdomen plane is relatively simpler than the head. It only requires the visualization of the stomach as shown in Figure 2.6 (b). The umbilical vein can also be visualized, but this is not a required structure for abdomen assessment. Three abdominal image measurements are typically made: the *anterior-posterior diameter* (APAD), the *transverse abdominal diameter* (TAD) and the *abdominal circumference* (AC). APAD is obtained by measuring the widest diameter from the outer borders of the abdomen. TAD is measured from the line from the outer borders at 90° to the APAD. Similar to the HC measurement, AC can also be measured either by measuring the circumference of the ellipse surrounding the abdomen or by deriving AC from the equation based on the APAD and TAD,

$$AC = \frac{\pi \times (APAD + TAD)}{2} \tag{2.3}$$

Both HC and AC measurements are based on the ISUOG guidelines [11].

As shown in Table 2.1, the only requirement of the limbs is whether they are present with three segments (upper, lower and hand/foot) as shown in Figure 2.6 (c). However, unlike the head and abdomen, the limbs can be difficult to visualize due to their unpredictable pose. The limbs are defined by three segments. Each segment can have a different orientation which adds to the complexity of visualization. They are also small relative, not just to the non-fetal parts, but also to the fetus itself.

2.2.2 First trimester scan image analysis

Image analysis in the first trimester is challenging. As mentioned in Section 2.2.1 and 2.2, ultrasound is heavily dependent on the skill of the sonographer. Also, the obesity of the mother and the variance in the fetal pose contribute to imaging challenges. Even when the

correct plane has been found, there are different ways to derive the image measurements. GA, for example, has different equations derived from the CRL [97], [93]. This is due to the different population used to regress the equation. As already mentioned before, the HC and AC can be obtained by using the equation derived from the diameter or by measuring the circumference of the ellipse drawn around the head/abdomen. Therefore, biometric measurements can be different depending on the observer and also which equations/parameters are used.

Compared to the second trimester, the first trimester is challenging due to the small size of the organs and it can be difficult to visualize fine detail of internal structures. In the second trimester, internal organs have more distinguishable characteristics that can be used to inform automatic pattern recognition. In the first trimester, however, the fetus is smaller and partially grown organs can be more difficult to detect. As already mentioned, the data taken for this thesis is based on the whole fetus. Unlike the data focused on specific fetal parts, it is more difficult to visualize the internal structures.

One solution to assist in the interpretation of the first trimester scans would be to provide an automated way to detect/segment and do the measurements on parts of the fetus. Some studies based on the second trimester fetus reported the feasibility of automated analysis motivated by the desire to improve work flow efficiency by reducing the requirement of the high demand from the sonography practices [98], [99]. To do this, a number of studies have shown good performance of automated analysis using machine learning techniques and the next section reviews some of the most successful approaches to-date.

2.3 Image analysis

This section introduces the relevant literature on automated medical image analysis based on machine learning methods.

2.3.1 Machine learning in medical image analysis

Machine learning is starting to be widely applied in medical image analysis. Current recent examples are for the detection and the segmentation of anatomies [100], [101],

[102], [103], [104], [105], measurements [106] [107], [108], [109] and standardized plane findings [110], [111], [112]. Some of the early machine learning methods proposed included *Support Vector Machines* (SVMs) [113], *Adaboost* [114] and *Random Decision Forests* (RFs) [115], [116]. SVMs is a learning algorithm where the classifier constructs a hyperplane that has the least generalization error. Maraci et al. [100], for example, used an SVM classifier based on dense SIFT features [117] with Fisher Vector encoding [118] to identify the fetal skull, heart and the abdomen from each frame in a second trimester ultrasound video with the 98.9% mean classification accuracy. Deng et al. [119] used SVM with *histogram of oriented gradients* (HOG) features to classify the region of nuchal translucency from the first trimester scan. Adaboost is a so-called strong classifier that combines a number of *weak* classifiers to improve classification performance [114]. In computer vision, Viola and Jones [120] used Adaboost to detect the faces from images with Haar-features. In medical image analysis, Rahmatullah et al. [101] used an Adaboost framework to detect the anatomical landmarks of the abdomen such as stomach bubble and umbilical vein with the detection rate of 82.8% and 72.55%, respectively from the second and third trimester scan. Namburete et al. [102] used Adaboost combined with the Nakagami distribution to detect the choroid plexus in second trimester fetal brain images. RFs were first introduced by Breiman [116] and became popular for classification and regression problems. Compared to classic decision forests [115], RFs combine a number of trees in which tree nodes are created in a random manner to reduce over-fitting. Dollar et al. [121] extended RFs to *Structured Random Forests* (SRFs.) This approach considers interdependencies between structures that form the patterns. This turns out to be a very useful automatic edge detector for the boundary of the fetal parts as discussed in **Chapter 3**. Lempitsky et al. [103] used RFs to automatically delineate the myocardium from the 3D ultrasound of the adult hearts. Yaqub et al. [104], [105] proposed weighted voting in the RFs framework to put higher weight on the tree that has higher individual accuracy instead of treating each tree equally. They applied their weighted voting RFs using their 'feature set' to detect the local brain structures from 3D second trimester ultrasound fetal brain images [104] and to do the segmentation of the fetal femur from 3D ultrasound datasets and the brain structures from MRI adult brain

images [105]. Compared to second trimester scan, there are less number of studies done in the first trimester. Since the first trimester, espeically the 3D first trimester scan, is not the routine scan, it is difficult to have large dataset to train on. In **Chapter 3 and 4** will describe how two consecutive RFs (SRFs + RFs classifier) can be used to localize the whole fetus from the 3D first trimester volumes [122].

2.3.2 Deep learning in medical imaging

Deep learning has enjoyed wide publicity and dramatic uptake of use in academic literature in the past few years. Here we provide a brief historical review. The idea of a Convolutional Neural Network (CNN) was first proposed by Fukushima [123] who designed neural networks called *Neocognition* with convolutional layers and multiple poolings. CNN was also applied in medical image analysis by Lo et al. [124] in 1995 for lung nodule detection. In 1998, LeCun et al. [125] proposed the neural network called *LeNet* which is considered to be the first successful deep learning on the real-world application for hand-written digit recognition. However, deep learning did not receive much attention until 2012, when Krizhevsky et al. [126] proposed CNN called *Alexnet* and won the ImageNet challenge [127]. CNN is the deep learning method that has shown the state-of-art performance over many applications. Since then a particular attractive feature of deep learning is that it does not require hand-crafted features. The handcrafted features refer to the properties derived using various algorithms such as *scale invariant feature transform* (SIFT) [128] and *histogram of oriented gradients* (HOG) [129] using the information present in the image itself. One crucial step required for handcrafted feature methods is that the humans have to define the features. However, this step is not explicitly required in deep learning, and instead, the algorithm learns the features that represent the data. Because of this, adapting a network architecture to a new problem can be done systematic way. However, deep learning also has some disadvantages. First, it is computationally expensive to train and more complex models might require weeks for training on a standard PC. For this reason, powerful *Graphic Processing Unit* (GPU) clusters are often used. Second, deep learning requires a large amount of data. As mentioned earlier, deep learning does not require handcrafted features

because it learns features from the data directly. However, this also means that it requires a large amount of data for good performance. This is one of the big challenges for medical imaging because compared to natural images, a typical data size is much smaller due to the need for ethical approval and the availability of applicable data because of scarcity of diseases in a given dataset. This means over-fitting is a major concern. To overcome data size limitations, data augmentation [130], and *Transfer Learning* which initializes the model parameters from the pre-trained models from the huge number of natural images in computer vision and then fine-tune the network parameters with the task-related samples [131], [132], [133] are commonly used.

After the success of the Alexnet, many computer vision groups proposed deeper networks to improve the performance [134], [135]. For instance, Long et al. [136] proposed *Fully Convolutional Networks* (FCNs) that outputs pixel-wise predictions and also combines the predictions from not just the last layer but also that from the lower layers to improve the segmentation performance. Deep learning has very rapidly become a standard technique in medical imaging [137], [138], [139]. One notable work is by Ronneberger et al. [140]. They took the idea of a FCN and proposed the network architecture called *U-net* for medical image applications. The overview of the neural network is explained in the next section and the specific explanation of CNN and FCN architectures will be explained in **Chapter 5**.

2.3.3 Brief overview of Neural Networks

In computer vision, a neural network is inspired by how the human brain learns about the visual world. Biologically, brain processing is done by *neurons*. A neuron is a unit of information processing in the brain. When the neuron receives signals from the other neurons, it processes the data and decides whether it will send the output to other neurons or not. An Artificial Neural Network (ANN) is the artificial model that processes data similar to this model of how a biological neuron processes. An example of a neural network structure is shown in Figure 2.7.

As illustrated in Figure 2.7 (a), the network is composed of a input layer, the hidden layers, and a output layer. The neuron first receives the input signals by the input

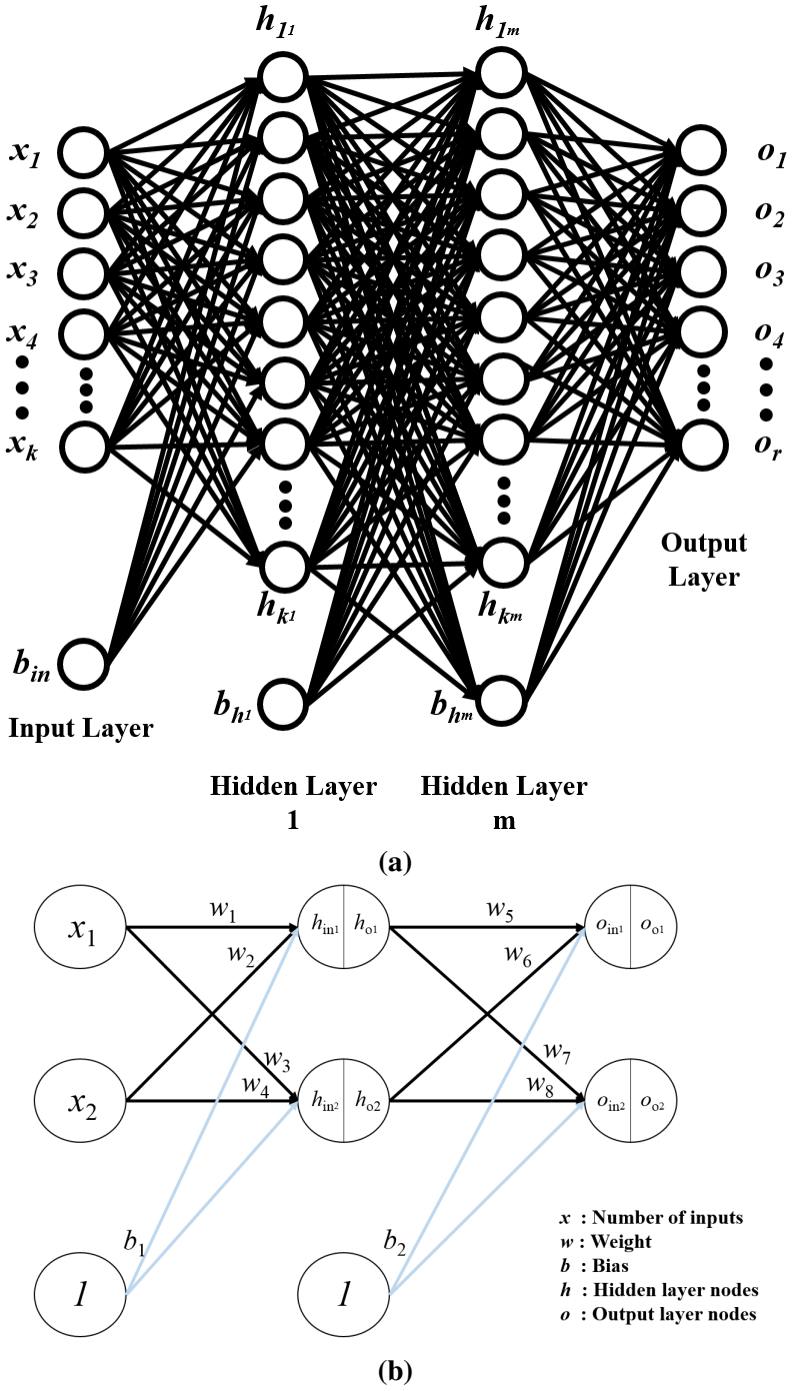


Figure 2.7: Examples of a neural network structure. (a) The input layer has nodes $x = \{x_1, x_2, ..., x_k\}$ and receives the data, and sends it to the nodes inside the $M^{th} \in \{1, m\}$ hidden layers, $h = \{h_1, h_2, ..., h_k\}$, after processing with weights, $w = \{w_1, w_2, ..., w_{k_m}\}$, and biases, b. Repeat this with different w and $b = \{b_1, b_2, ..., b_m\}$ to output layer, $o = \{o_1, o_2, ..., o_r\}$, which outputs the prediction. (b) As an example of how neural network works, simple three layer network is used as an example.

layer containing input nodes $x = \{x_1, x_2, ..., x_k\}$, which can be the input data or features. Using these inputs, the hidden layers learn the pattern and then the output layers make a prediction, for example, if it is a classification task, the output layer predicts the class that the input belongs to. Between each layer, the weights, $w = \{w_1, w_2, ..., w_{k_m}\}$, and the bias, $b = \{b_1, b_2, ..., b_m\}$, are added to the input of each layer. The weight represents contribution of each node. The bias is added to prevent the subsequent error from the case when the previous layer neuron responses are zero. As an example of how neural network works, we consider small network (two nodes for each layer where input layer has x_1 and x_2, hidden layer has h_1 and h_2 where each node has inputs (h_in) and outputs (h_o), and output layer has o_1 and o_2 where they also have input (o_in) and outputs (o_o)) as shown in Figure 2.7 (b). One of the input of hidden layer, h_1, is then:

$$h_{in_1} = w_1 x_1 + w_2 x_2 + b_1 \tag{2.4}$$

Then the output of the hidden layer, h_{o1} can be calculated using, for example, a logistic function,

$$h_{o1} = \frac{1}{1 + e^{h_{in_1}}} \tag{2.5}$$

Using h_{o1} as the input to the output layer, similar to the process between input and hidden layer, the input of the output layer, o_1, is then:

$$o_{in_1} = w_5 h_{o_1} + w_6 h_{o_2} + b_2 \tag{2.6}$$

As the output of the output layer, o_{o1}, another logistic function is used,

$$o_{o1} = \frac{1}{1 + e^{o_{in_1}}} \tag{2.7}$$

which is the prediction result. Other parameters including h_{o2} and o_{o2} can be calculated in a similar way. With this process, the neural network is the combination of the number of neurons to output the prediction. However, since it is difficult to find out which input signals are more important for the prediction, a training method based on *gradient descent* is applied on the weights by repeating the feedforward and backward propagation.

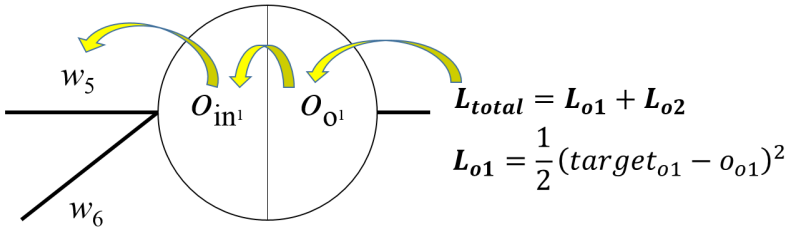


Figure 2.8: The backpropagation in the output layer from the output that shows how much w_5 affects the loss, L_{total}

Backpropagation

Feedforward is basically doing the prediction using the certain weights and biases that has been explained previously. Once the network outputs the prediction, this prediction needs to be compared with the groundtruth to get the error. This is called the *loss*, and the loss function is applied. The *Squared Error Loss*, L_{total}, is used as an example in this section,

$$L_{total} = \sum_{i=1}^{n} \frac{1}{2}(Y_i^{'} - Y_i)^2 \tag{2.8}$$

where L is the loss, n is the number of predictions (classes or regression value), $Y_i^{'}$ is the groundtruth, and Y_i is the prediction. As equation 2.8 shows, the loss represents how close the prediction and groundtruth are. Therefore, as the loss gets smaller, the better is the prediction. As mentioned earlier, to make the loss smaller, finding good weights is required. To do this, *Backpropagation* is employed, which is to keep updating the weights until the loss value reaches the global minimum by finding out how much changes in weights affect the total loss.

Output layer As shown in Figure 2.8, in the output layer, each loss is affected by the weights applied to the outputs from the hidden layer. Therefore, finding how much these weights affect the loss needs to be considered first. For example, $\frac{\partial L_{total}}{\partial w_5}$. $\frac{\partial L_{total}}{\partial w_5}$ can be calculated by applying the chain rule since the weights affect the loss indirectly. By applying the chain rule, $\frac{\partial L_{total}}{\partial w_5}$ is

$$\frac{\partial L_{total}}{\partial w_5} = \frac{\partial L_{total}}{\partial o_{o1}} \times \frac{\partial o_{o1}}{\partial o_{in_1}} \times \frac{\partial o_{in_1}}{\partial w_5} \tag{2.9}$$

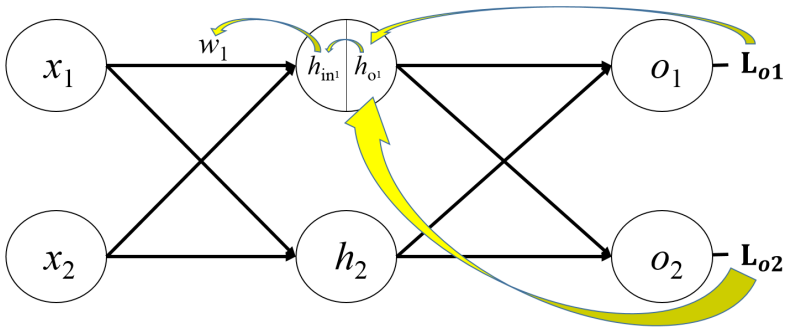


Figure 2.9: The backpropagation in the hidden layer from the output that shows how much w_1 affects the loss, L_{total}. Note that the backpropagation in the hidden layer needs to compute the multiple outputs.

After calculating $\frac{\partial L_{total}}{\partial w_5}$, the current weight is updated by subtracting $\frac{\partial L_{total}}{\partial w_5}$ multiplied by the learning rate η, which controls how quickly the network updates the weights to prevent the overcorrection that will cause the loss to increase. Namely,

$$w_5^* = w_5 - \eta \times \frac{\partial L_{total}}{\partial w_5} \tag{2.10}$$

where w_5^* is the updated weight.

Hidden layer Similar to the output layer, we need to find how much changes in the weights applied to the input layer affect the outputs of a hidden layer. An example is shown in Figure 2.9. However, unlike the output layer, there is more than one output from each neuron in a hidden layer. Therefore, every loss from each task needs to be considered. For example, $\frac{\partial L_{total}}{\partial w_1}$ can be separated using the chain rule,

$$\frac{\partial L_{total}}{\partial w_1} = \frac{\partial L_{total}}{\partial h_{o_1}} \times \frac{\partial h_{o_1}}{\partial h_{in_1}} \times \frac{\partial h_{in_1}}{\partial w_1} \tag{2.11}$$

where $\frac{\partial L_{total}}{\partial h_{o_1}}$ is,

$$\frac{\partial L_{total}}{\partial h_{o_1}} = \sum_{i=1}^{2} \frac{\partial L_{o_i}}{\partial o_i} \times \frac{\partial o_i}{\partial h_{o_1}} \tag{2.12}$$

With $\frac{\partial L_{total}}{\partial w_1}$, the updated weight, w_1 is,

$$w_1^* = w_1 - \eta \times \frac{\partial L_{total}}{\partial w_1} \tag{2.13}$$

Other weights that have not been covered in this section can be updated by following similar principles. Once all the weights are updated, feedforward is applied again, but

with the updated weights to check if the loss is decreased. This process is repeated until the loss becomes as small as possible.

Deep learning in medical image analysis

Recently, deep learning has established itself as a mainstream technique in the medical image analysis literature due to its state-of-art performances over previous methods. Before FCNs, most deep learning methods (mostly CNN) focused on image classification or object detection [141], [132], [142], [143], [122], [144], [145], [146]. For instance, Gao et al. [132] used a CNN to label each frame from the second- and third trimester fetal ultrasound video into one of four classes: fetal skull, heart, abdomen and other structures. What is interesting in that work is that the CNN which is fine-tuned from the transferred parameters from the pre-trained models from the natural images have shown better results compared to the model trained directly from the ultrasound dataset (91.5 % versus 87.9 %. Gao et al. [145] extended the classic CNN architecture to a two-stream CNN that combines temporal and spatial learning to improve fetal heart detection and localization performance. [146] used a CNN to localize the fetal brain, to detect the regions that contain structures of interest and to learn the acoustic pattern in the regions that enable plane verification based on the second trimester fetal brain. According to [146], they achieved 96.9 % Dice coefficient for detecting the head region. [132] and [146] are closely related to the goal of this thesis because we also use deep learning approach to classify the 2D slices extracted from a first trimester 3D volume. However, [132] and [146] used deep learning approach on the second trimester and their data is not from the 3D data. Also, the images they used are focused on the fetal anatomies (in axial view) whereas our data is focused on the whole fetus which makes the fetal anatomies in the axial view much smaller compared to the non-fetal parts.

Image segmentation using deep learning also became a popular application after Long et al. [136] proposed a FCN. Introducing a pixel-wise output that is useful for the semantic segmentation and combining the output with the prediction from the lower layers to obtain finer boundaries inspired the medical image analysis community for the better anatomical segmentation [147], [148], [149]. [147] used FCN to perform semantic

segmentation of the fetal heart and achieved the error rate of 23.48 %. [148] proposed *Cascaded FCN* (CFCN) which combines multiple FCNs to keep refining the boundary of the segmentation result of the second- and third trimester fetal head and abdomen (Jaccard Index of 0.97 and 0.97 respectively). [147] and [148] showed good performance on the fetal segmentation, however, both used 2D images that are specifically focused on the fetal anatomies in later trimester, and [148] requires more than two networks to acheive their performance. [149] used implementation of FCN called FovFCN which uses the image patches with different resolutions and then combine those features to predict the segmentation on the second trimester 3D abdomen volumes with the Dice coefficient of 0.86. Raynaud et al. [150] used a FCN combined with morphological filters for fetal spine segmentation to automatically align the 3D ultrasound volume to better localize multiple organs such as the heart, the stomach, the umbilical vein and the bladder. However, both [149] and [150] only focused on the abdominal volumes. Ronneberger et al. [140] extended a FCN and built a U-net that combined opposing convolution and deconvolution layers. Other application of U-net include [151], [152], [153]. Cicek et al. [151], built a *3D U-net* for the full 3D segmentation of the *Xenopus* kidney which was achieved by learning from 2D sparse annotation. Yang et al. [153] performed 3D segmentation of the whole fetus, gestational sac and the placenta from the first trimester fetal volumes using a 3D FCN as the initial segmentation of each part and an RNN to refine the segmentation result. In our knowledge, this is the only study that used deep learning approach to do the segmentation that includes the whole fetus in the first trimester. Although this interesting work proposed volumetric segmentation of the whole fetus, they performed whole fetus segmentation only, but not segmentation of each fetal part such as the head, the torso and the lower limbs.

Performing the segmentation task using the deep learning has not been studied well enough, especially for the first trimester ultrasound. Even for the second triester fetal ultrasound, there is only a few number of studies. The deep learning approach to the semantic segmentation for the medical image analysis is recently growing, and the manual groundtruth for the semantic segmentation is much harder to get compared to the classification task. This is because to get the groundtruth, the human needs to do the

semantic label as accurate as possible, whereas the classification just requires the class number. Like [153], there has been research about the segmentation of the whole fetus, but still, no studies has been done on the semantic segmentation of the whole fetus in the first trimester. Also, there were less studies that involved the automated fetal assessments including anatomy visualization and the measurement at the same time from a 3D volume.

2.4 Conclusion

Ultrasound is the preferred imaging modality in prenatal care. Especially, the first trimester scan is important to monitor the fetal growth rate and it has potential to perform early diagnosis of certain anomalies that are usually diagnosed in the second trimester. However, the first trimester scan is challenging compared to the second trimester due to the small size of the fetal organs. Moreover, ultrasound image quality depends heavily on the sonographer skill which yields inter- and intra-observer variability. Therefore, it is interesting to consider how to develop automated method to overcome these limitations. There are many studies using deep learning for automated fetal image analysis, but they mostly concern the second trimester fetal data and focus on specific fetal parts such as brain and abdomen. Also, no studies have been done on the 3D fetal ultrasound volumes that were acquired based on the whole fetus to perform a multiple fetal health assessments automatically. Yang et al. [153], for example, only performed segmentation task on the whole fetus, but not on each fetal parts. In this work, we developed the automated tools based on the deep learning approach for the automated first trimester fetal assessment that includes the visualization of the fetal anatomy and performs the biometry.

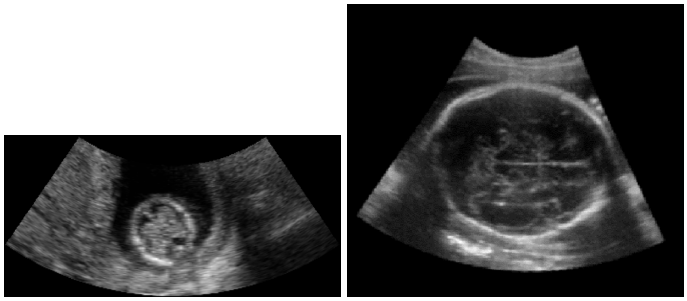
3

Automated Localization of the Whole Fetus from the 2D Images Extracted from a 3D Volume

In Chapter 2, clinical motivation was provided for automated first trimester fetal analysis, and relevant research that has been done on related problems in computer vision and medical applications was also reviewed. As described in Chapter 1, the first step in automatic analysis is to localize the whole fetus from the sagittal view, which is the good anatomical plane for fetal localization. In the current chapter how to localize the whole fetus from the sagittal view using a machine learning based method is described. Section 3.1 introduces the problem and proposed solution. Section 3.2 describes two main consecutive methods that have been designed: one uses edge features and the other uses the whole fetal profile. Section 3.3 presents the experimental setup and describes the validation methodology employed. Section 3.4 presents results. A summary is provided in Section 3.5.

3.1 Introduction

The localization of the whole fetus within a 3D US volume is an important step in automation of analysis. The data employed in this research has been acquired based on



(a) **(b)**

Figure 3.1: Examples of a) First trimester fetal brain image (b) Later trimester fetal brain image. Both are from 3D, but (a) is scanned based on whole fetus, and (b) is scanned based on fetal brain.

CRL measurement criteria. Refer to the Appendix A for more detail about the 3D data acquisition based on CRL criteria. As shown in Figure 3.1, the fetal parts cover a small area in the axial plane. Organs, such as brain and abdomen, are not fully developed in this first trimester scan and smaller compared to later trimesters. Also, the data is 3D which is focused on the whole fetus. This makes the fetal parts cover a much smaller area in the axial view. Therefore, to localize the whole fetus from the sagittal view, the search field of the fetal parts in the axial view is reduced by removing non-fetal parts that are out of the localized region while preserving the fetal parts that needs to be focused on.

One of the key things to localize the whole fetus is that there is always amniotic fluid between the fetal front and non-fetal tissue as shown in Figure 3.2. Due to the fact that the fluid echogenic, this region is visualized as a black region in the image. Therefore, edge-detection method to detect the boundary between the fetal front and amniotic fluid is used. However, the difficult part for localization of the whole fetus is that there is less space between the back of the fetus and non-fetal tissue which yields an ambiguous intensity boundary. Also, it is important to note that since there are so many intensity edges that can be detected, small edges, especially those from the inside the fetus, should be ignored. Threfore, using only intensity-related features such as intensity values and intensity gradient, like how Canny edge detection method [154] uses, is not suitable. The example using classic Canny edge detection method [154] is shown in Figure 3.3. As shown in Figure 3.3 (d), having larger threshold can remove some internal edges, but some parts of the boundaries of the whole fetus are also removed. Most of all, the value of optimal threshold is dependent on the data, and hence cannot be generalized. Therefore,

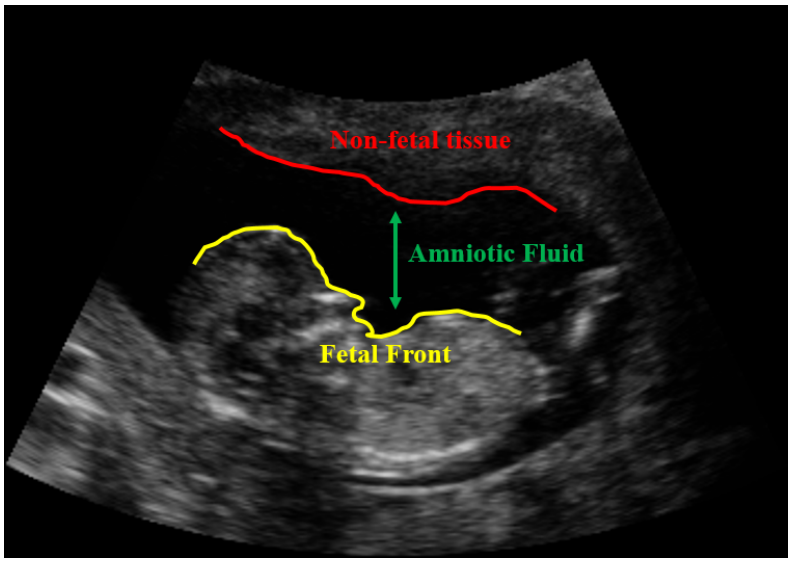


Figure 3.2: The example of the whole fetus in the sagittal view. Fetal front (yellow) is where the fetus faces and amniotic fluid (green) is the black region between fetal front and non-fetal tissue (red)

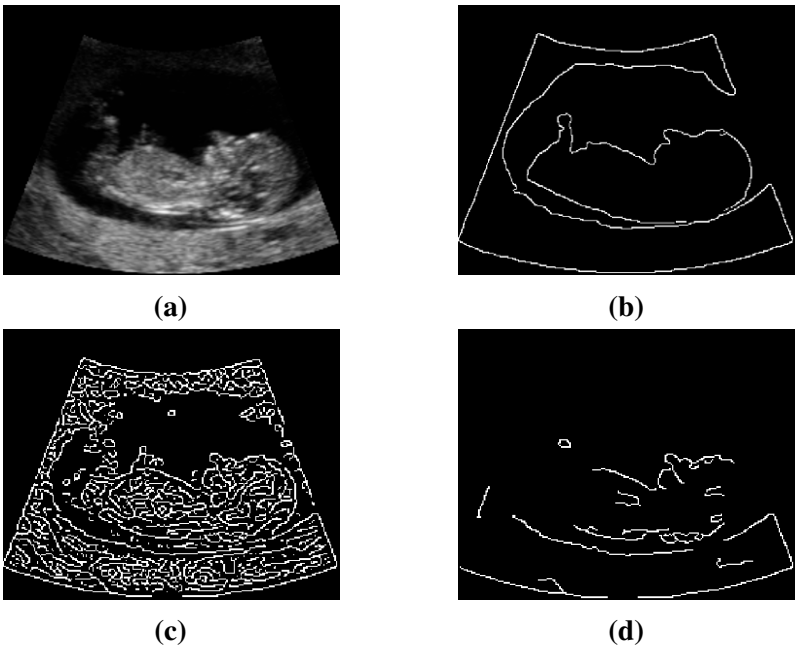


Figure 3.3: Examples of edge map using Canny edge detection method (a) Raw data (b) Groundtruth, c) Edge-map using Canny d) with more threshold. There are many small edges that can be detected due to the internal fetal structures and non-fetal tissue.

methods that can give more strength on the fetal boundary than other edges are needed. There has been a number of studies which applied edge detection for medical images [155], [156], [157], [158]. [155] and [156] applied the speckle noise reduction before edge detection. For example, [155], used adjusted Canny edge detection that removes the speckle noise by using median filter instead of Gaussian smoothing on the whole-fetus ultrasound phantom image. [156] used clustering method called fuzzy c-means [159] to distinguish the edges from the speckle noise. [157] and [158] applied phase-based edge detection method. [157], used phase symmetry [160] to detect the edges of the prostate ultrasound images, and [158] used phase-based probabilistic gradient vector flow [161] to extract the edges to detect the nerves from the ultrasound images. Both studies proved that phase-based approach is robust to the speckle noise and able to identify the edges. [155], [156], [157], [158] focused on the speckle noise since they can weaken the edges or yield the unnecessary edges. However, these methods require substantial degree of user-interaction and the threshold/parameters have to be carefully chosen. Due to this, other studies applied learning method to do the edge-detection/segmentation [162], [163], [164]. The advantage of the learning-based method is that it can identify which pattern can distinguish the region of interest from the background and which edges are more important than others. [162] used marginal space learning to detect the robust edge and refine them using 3D shape model with probabilistic boosting tree for the CT liver. However, they require 3D mesh model which is difficult to annotate manually. [163] used graph-based learning for the segmentation of 3D ultrasound phantom and reported that their method outperformed fuzzy c-means clustering method. However, due to the complex graph structures, it is computationally expensive. [164] applied AdaBoost algorithm with Haar features followed by boundary detection using phase-based method to detect the head. [164] also used gestational age to guide the sliding detection window to decrease the detection speed. They reported that their automatic measurement based on the detected boundary had no significant difference compared to the manual measurement and they also reported that they were able to achieve with faster computational speed. However, their method is limited to the head which the skull is visualized as bright ellipse, and the gestational age has to be provided. Therefore, for the objective of this

thesis, we need to apply learning-based method to have less parameter-dependency and does not require shape- and graph-based learning method to have faster computational speed. Inspired by [116] and [121], we used structured random forests. Random Forests is one of the machine-learning method that is relatively robust to overfitting and noise compared with Adaboost, and the training is faster and better for multi-class problem [116]. Structured random forests is the extension of the random forests which is more suitable for the segmentation and edge detection because it considers interdependencies between labels that can help to extract clearer edges. We will explain more in detail in Section 3.2. Once the edges are extracted, the fetus needs to be localized by having a bounding box around using the detected edges. Using the fact that the object, which is the fetus in this case, has an enclosed boundary, the bounding box is likely to contain the object if the enclosed boundaries are wholly contained in the bounding box.

However, using only an edge-detection method cannot localize the object because it is highly possible that many edges can be detected and it is difficult to know which edges belong to which objects. This is also why there will be so many bounding boxes that can be proposed, which indicates the bounding boxes that the whole object is likely to exist within, based on the detected edges. This is critical for whole fetal localization because the bounding box will likely miss out the fetal parts. Therefore, another method is needed to reduce or select the proposed objects which contain the whole fetus by recognizing its particular image pattern within the object proposal region.

In this work, a method with two components is proposed, each based on Random Forests, to localize the whole fetus in the sagittal view. First, to get the object (fetus) proposals, Structured Random Forests [121] and the Edge-box approach [165] are proposed. Then a second Random Forest is approached to classify whether the proposed region contains the whole fetus or not. This second Random Forests is shown to be important to reduce the number of object proposals which improves the accuracy of whole fetus localization.

3.2 Methods

Figure 3.4 shows the block diagram representation of the resulting algorithm. Using the structured random forests to detect the edges around the whole fetus from the input/raw

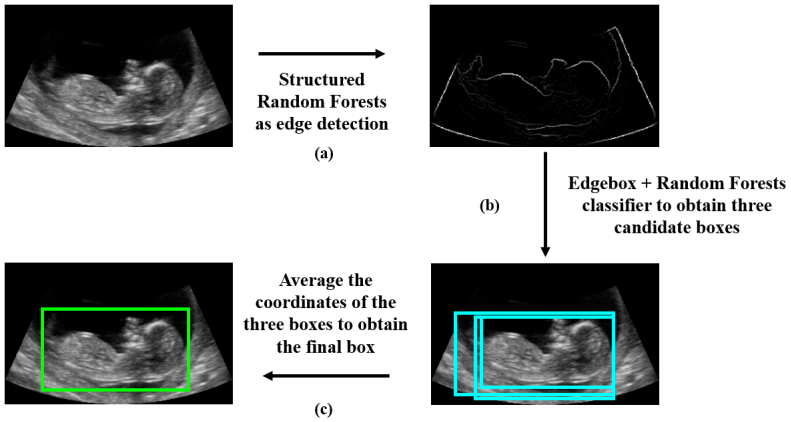


Figure 3.4: The block diagram representation of the resulting algorithm. (a) Using the structured random forests to detect the edges around the whole fetus from the input/raw ultrasound image. (b) Using the Edgebox algoithm and Random forests classifier, three candidate boxes around the whole fetus are obtained. (c) The coordinates of the three boxes are averaged to get the final box around the whole fetus.

ultrasound image (Figure 3.4 (a)). Then based on these edges, Edgebox and Random Forests classifier are used to get the three candidate boxes (Figure 3.4 (b)). These boxes' coordinates of each corner are averaged to obtain the final box (Figure 3.4 (c)).

3.2.1 Structured Random Forests

Structured Random Forests [121] uses Random Forests framework to learn structured information. The big difference between a traditional learning method, which is generally pixel-based, and structured learning is that while traditional learning outputs a class label or regression value, the structured learning outputs the structures. The motivation of learning the mapping between input and output is that many visually salient edges do not always correspond to the intensity gradient which classic edge detection method used, such as Canny. As shown in Figure 3.3, the salient edges are the boundary around the fetus, but there are also many internal edges as well. Therefore, by learning from manually annotated data, the most important edges can be learned. As the application for this work, this is important for whole fetus localization because internal fetal structures yield many edges which are less important than the edges around the whole fetus. A further advantage of learning the structured objects is that it is able to exploit the interdependencies within a local patch. For example, the edge pixels in a local patch are highly interdependent that they form the patterns such as straight lines or Y-shaped lines [166], [121]. In

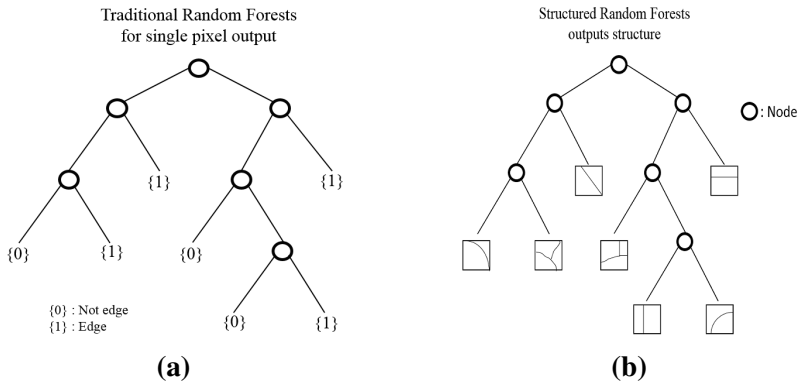


Figure 3.5: Example of decision tree from a) Traditional Random Forests of single pixel prediction, b) Structured Random Forests which does structure prediction

classification approaches such as traditional Random Forests, the input data are assigned to a single class labels such as $\{0,1\}$ for binary classification from the centre pixels of the local patch whether that pixel belongs to the edge or not. However, since they only consider the centre pixel of the local patch, they do not exploit any dependencies or edge structure patterns. This will yield edges with no coherent structures. Therefore, by exploiting this interdependencies, the extracted edges will have coherent structures which outputs a clearer edge map. A comparison of the one decision tree from traditional Random Forests and Structured Random Forests is shown in Figure 3.5.

As shown in Figure 3.5, Random Forests is the collection of decision trees. Each tree contains a number of nodes which each split into the group by similarity as structures. This similarity is determined by Information Gain which will be explained shortly. Instead of having output as $\{0,1\}$, Structured Random Forests outputs edge structures. Using Random Forests framework with structured labels has two main challenges. First, the output spaces are complex and high-dimensional. The traditional Random Forests used for edge detection does single pixel prediction. The prediction is the binary classification of the centre pixel in the local patch that indicates whether that pixel belongs to the edge or not. Structured Random Forests, however, needs to compute structured output which has a higher dimension than pixel output. Second, the information gain over structured labels are not well defined. In each node j, the binary split function h is applied,

$$h(x, \theta_j) \in \{0,1\} \tag{3.1}$$

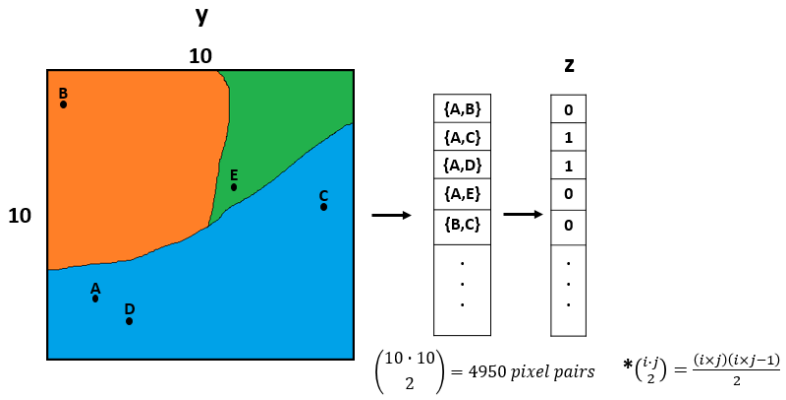


Figure 3.6: The example of intermediate mapping from y to z by encoding whether every pair of the pixels in y belong to the same segments or not. In this example, y is 10×10, so there will be 4950 pairs that was encoded to y.

where x is sample and θ_j is the parameters. The goal of training the tree is to find the parameter θ_j that split the data into a group with as much similarity as possible, which is called a *good* split of the data. As a simple example, if the data is the collection of sky and grass, the parameter that is related to the object color will be strong parameter that can do a *good* split. To do this, the information gain is used which is based on measuring the similarity. The information gain is defined as,

$$I_j = H(S_j) - \sum_{k \in L,R} \frac{|S_j^k|}{|S_j|} H(S_j^k), \tag{3.2}$$

where $H(S)$ is Gini impurity $H(S) = \sum_y p_y(1 - p_y)$, S is the training set in node j, and p_y is the probability of class y in S. The parameter θ_j is chosen to maximize the information gain, $\theta_j^* = \arg\max_{\theta_j} I_j$.

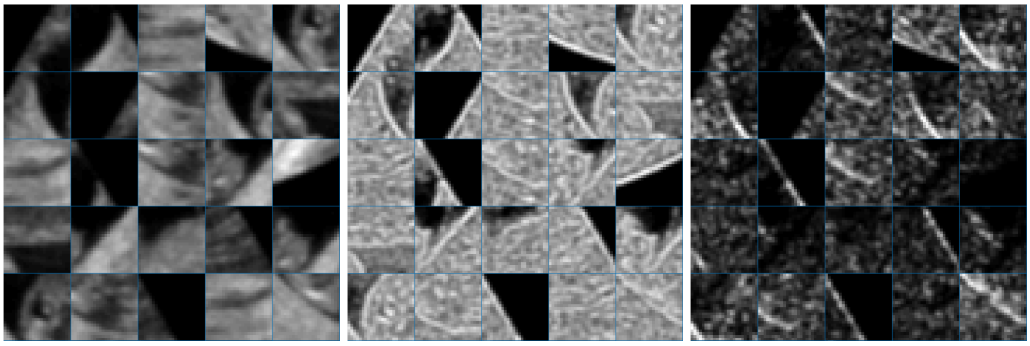
However, the similarity in structure has not been defined well enough because there needs to be some additional information such as spatial relationship between labels inside the patch. Dollar et al. [121] address this issue by a two-stage approach of first mapping $Y \rightarrow Z$ followed by mapping of $Z \rightarrow C$, where Y is output space, Z is an intermediate space and C is discrete labels. Mapping between Y and Z is called the *intermediate mapping*, $\Pi : Y \rightarrow Z$. Intermediate mapping is computed by encoding whether every pair of the pixels in y belong to the same segments or not. An example is shown in Figure 3.6.

However, as shown in Figure 3.6, as the size of the patch increases, computing $z \in Z$ for every $y \in Y$ is getting more expensive. Therefore, to reduce the dimensionality,

the dimensions of Z is first randomly sampled. Randomly sampling from Z gives two advantages. First, the computation is faster since the full dimension of Z is not analysed. Second, by adding another source of randomness during learning process, it helps to have more diversity of trees. Then Principal Component Analysis (PCA) is used to further reduce the dimensionality of Z. This reduced Z is then assigned to discrete labels $c \in C$ where $C = \{1, ..., k\}$ and k represents the number of classes, which, in this case, the number of different edge structure pattern that are mapped from structured labels $y \in Y$ which is annotated manually. During training, z with similar edge structure pattern is assigned to c. With the discrete label, the information gain can be calculated in a similar fashion to traditional Random Forests. The set of n labels $y_1 ... y_n \in Y$ are combined into a single prediction for both training and testing. To do this, m dimensions of Z are sampled and $z_i = \Pi_\phi(y_i)$ computed for every i. Finally the label y_k is selected whose z_k minimizes $\sum_{ij}(z_{kj} - z_{ij})^2$. Since Random Forests use multiple decision trees, there will be multiple predicted edge maps from each image patch. Therefore, the predicted edge maps are averaged as the final prediction.

In this work, the method for downsampling and the parameters similar to [121] is applied except the size of image patch which predict 24 $\times$ 24 segmentation mask from larger 48 $\times$ 48 image patch from the input images to capture the bigger edges. As input features, there are two main types of features from image patch: pixel and pairwise differences. The pixel is the accumulation of channels of features extracted from the image patch which are grey-intensity (pixel values), 2 normalized gradient magnitude and 4 orientation from each gradient magnitude, which makes total 11 channels of features. An example is shown in Figure 3.7.

Another feature type called pairwise differences or self-similarity [121] is also used. With the observation that the contour occurs not only at intensity but also at texture boundaries, the pairwise difference features can capture the region inside the image patch that contain similar texture. Each channel is blurred using a linear filter called triangular filter, and downsample to a resolution of 5 $\times$ 5 pixels which gives $\binom{5\cdot5}{2}$ = 300 additional features for each channel. Blurring before the downsampling is to avoid



(a) (b) (c)

Figure 3.7: Examples of features used in Structured Random Forests a) Intensity value b) Normalized gradient magnitude c) Normalized gradient orientation.

the distortion and make smoother boundaries when the image is downsampled. In total, 9636 features are used per image patch.

3.2.2 Edge-box

Given the edges from Structured Random Forests, bounding boxes can be obtained based on a score which measures the number of edges enclosed within the region of interest (ROI) compared to the number of edges overlapping the ROI boundary [165]. The edges are grouped together by combining the edge pixels until the sum of their orientation difference is above a threshold, which in this case is set to $\frac{\pi}{2}$. $\frac{\pi}{2}$ is set as threshold because edge pixels connected by a contour with high curvature will have low affinity to be grouped. With these edge groups, another affinity measure, α, is computed, which is defined between edge groups, based on their mean position and mean orientation,

$$\alpha\left(s_i, s_j\right) = \left|cos(\theta_i - \theta_{ij})cos(\theta_j - \theta_{ij})\right|^{\gamma} \tag{3.3}$$

where s_i and s_j are a pair of edge groups, θ_i and θ_j are the orientation of each s_i and s_j, θ_{ij} is the angle between the mean positions of two groups, and γ is called the affinity sensitivity; $\gamma = 2$ in this work. In each group, a value v_b is computed, which indicates whether the edge group s is wholly contained or not in the box. v_b lies within the range 0 and 1; $v_b = 1$ when the edge group is wholly contained in the box, and 0 otherwise. For a partially contained edge group in the box, v_b is computed as follows:

$$v_b(s_i) = 1 - \max_{T} \prod_{j}^{|T|-1} a\left(t_j, t_{j+1}\right) \tag{3.4}$$

where T is the ordered path of edge groups with the length of $|T|$ that begins with $t_1 \in S_b$ and ends at $t_{|T|} = s_i$. Using v_b, the scoring, h_b, is defined as,

$$h_b = \frac{\sum_i v_b(s_i) m_i}{2(b_w + b_h)^k} \tag{3.5}$$

where m_i is the summed magnitude of s_i edges, each b_w and b_h indicates box's width and height, and k is the bias. The larger the score, the more that an object is enclosed within a box. Based on these scores, the candidate boxes can be ranked. Also, according to [165], the edges in the center of the box are less important than the edges near the boundary. Therefore, the final score, h_{bb} can be computed by subtracting h_b (equation 3.5) from the score of edges that are in the centred box, b^{in}, which has width and height $\frac{b_w}{2}$ and $\frac{b_h}{2}$ respectively,

$$h_{bb} = h_b - \frac{\sum_{p \in b^{in}} m_p}{2(b_w + b_h)^k}, \tag{3.6}$$

where p is the edges inside b^{in} and m_p is the summed magnitude of p edges.

Since the box is based on detected edges, there are many possible candidate boxes proposed. Therefore, the number of boxes is decreased by introducing the constraints. The constraints are on the expected position and size of the boxes. It is highly possible that the centre of the irrelevant box and the centre of the whole image are not similar. To do this, we first get the box, which area is half of the area of the image and is located at the centre of the image. This box will be called *half-box* in this work. Then if the centre of the candidate box is located outside of the half-box, then it is highly possible that this candidate box does not contain the fetus because we expect that the position of the fetus is located at the centre of the image. Also, the area of the irrelevant box is smaller than 30% of the area of the image which follows the criteria of good sagittal plane as mentioned in Appendix A. These irrelevant boxes are removed to decrease the number of candidate boxes. These irrelevant boxes will be called *out-of-range boxes* in this work. However, relying on the box with the largest score to identify the best enclosing box around the fetus still does not guarantee a good localization of the fetus since the bounding box itself is focused on the enclosed edges, not on the whole fetus. Therefore, another method is developed to classify whether a box contains the whole fetus or not as described next.

3.2.3 Weighted Random Forests classifier for best bounding box

Inspired by [105], weighted Random Forests are used to classify whether a candidate box contains the whole fetus or not. In the classic Random Forests classifier, it treats each vote from each tree as equal. However, it is possible that some individual trees give a poor accuracy. Yaqub et al. [105] proposed weighted voting that puts higher weight on the vote from the tree that has higher individual accuracy so that a tree with low accuracy has less effect on final classification. In a traditional Random Forests classifier, the probabilities from each tree are averaged to get the final probabilistic decision, p of image, I, belonging to the class c,

$$p(c|I) = \frac{1}{T}\sum_{t}^{T} p(c|I) \tag{3.7}$$

where T is the number of trees. Instead of averaging, which also means equal vote, Yaqub et al. [105] used another term α which involves weighting on the tree with higher accuracy as explained before,

$$p(c|I) = \sum_{t}^{T} \alpha_t p(c|I) \tag{3.8}$$

$$\alpha_t = \frac{\frac{1}{F}\sum_{f}^{F}(Q_f)_t}{\sum_{t}^{T}(\frac{1}{F}\sum_{f}^{F}(Q_f)_t)} \tag{3.9}$$

where $f \in F$ is the number of features from root to the leaf node and $(Q_f)_t$ is the training score of feature f used in tree t.

For features, the method from [105] is followed which extracts three main types of features: Unary, binary, average. Unary is the intensity value of random pixel in the random size window around the ROI. Binary is the sum, difference, absolute difference and log of two random windows. Average is based on the sum of all pixels of a random size divided by the number of pixels. The difference in these average features is taken. For fast computation for calculating these values, the integral image is estimated. Three types of features were extracted from two types of images. One is the raw image and the other is the pre-processed image that has feature symmetry measure, inspired by [167], local phase-based method that is intensity invariant and less sensitive to the speckle [167]. In addition to the three main types of features from raw images, Haar feature are

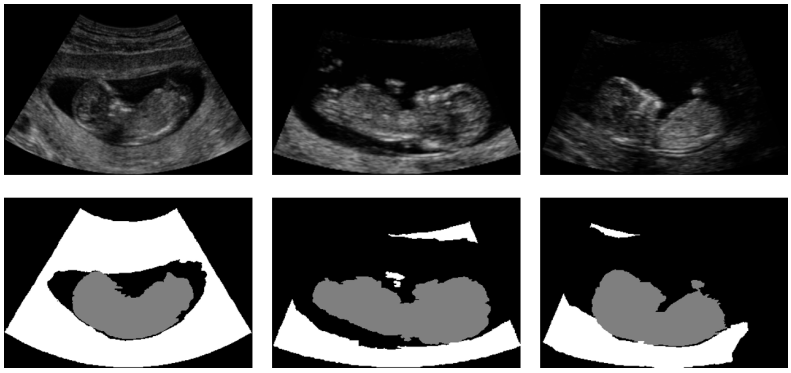


Figure 3.8: The training examples of Structured Random Forests. The top row represents the original images and bottom row represents the corresponding labelled images.

extracted [105]. Haar takes the difference between sum of the pixels from two adjacent windows, and this has been used to capture the edge regions. Because of this property of Haar-feature, Haar features were not extracted from feature-symmetry image. In this work, for parameters of the classifier, the maximum size of the window was set as 30 $\times$ 30, the number of trees as 20 and maximum depth of 10.

3.3 Experimental Setup

3.3.1 Datasets

Details of 3D data acquisition and criteria of good sagittal plane are given in Appendix A. From this data, the slices were extracted manually that visually define a good sagittal plane from each volume. An assumption in this work is that good sagittal planes have been given for this task. The volumes were randomly distributed into the training or test sets, but all volumes from same fetus were only in one of the sets. 33 volumes were used for training and 32 for testing. To train the Structured Random Forest, each image from each volume was first labelled into three classes: the whole fetus, the non-fetal tissue and the background, as illustrated in Figure 3.8 using ITK-SNAP [46].

For the box-classifier, the bounding box around the whole fetus was manually cropped as a positive example and flipped as augmentation to increase the size and variability in the training set. As the negative set, the bounding box was cropped across the whole image with randomized size and location with the constraint that the Intersection over Union (IoU) with the corresponding positive box must be under 0.7. As mentioned in

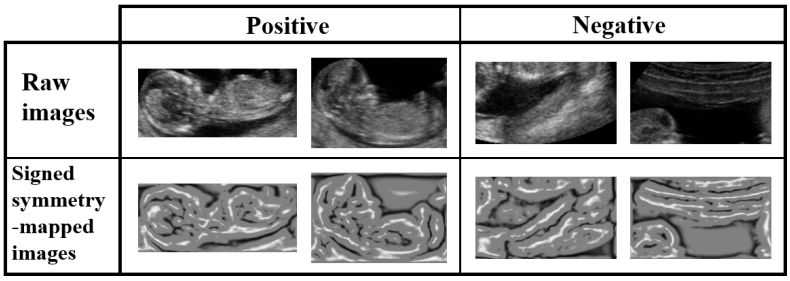


Figure 3.9: Training examples of positive and negative boxes. The top row represents the original images and bottom row represents feature-symmetry-mapped images.

section 3.2.3, the raw image was pre-processed into feature-symmetry-mapped images from each candidate box obtained from the Edge-box algorithm. Examples of positive and negative training examples are shown in Figure 3.9 with the corresponding signed symmetry-mapped images.

3.3.2 Validation Methodology

As a qualitative evaluation, Structured Random Forest results were compared with edge detection methods, Canny [154] and Sobel. This was to show how structured learning can weaken the small edges compared to other methods by learning the edge structures directly from the annotated data. The ground-truth is obtained from manually drawn boundary around the fetus and non-fetal tissue. To evaluate the boxes that contain the whole fetus, *Intersection over Union*, (IoU), precision and recall are applied, which are defined as

$$IoU = \frac{M \bigcap A}{M \bigcup A} \tag{3.10}$$

$$Precision = \frac{M \bigcap A}{A} \tag{3.11}$$

$$Recall = \frac{M \bigcap A}{M} \tag{3.12}$$

where M and A are the area of manual box and automatically selected box respectively. IoU is similar to *Jaccard Index* which quantifies the similarity between manual and automatically selected box. precision represents how much overlapped area covers the A, and recall represents how much overlapped area covers the M. precision and recall are used to tell whether the automatically selected box is under- or over-estimation since it is difficult to be clarified from IoU. For example, if two automatically selected boxes

have same IoU but one has small precision and large recall, that means this box is underestimated. The threshold of IoU was 0.7, precision was 0.7 and recall was 0.9. According to Everingham et al. [168], the threshold of 0.5 is typically used to determine whether the bounding box was correct, but it can be quite loose. Therefore, in this work, 0.7 is used as the threshold of IoU and precision to filter more number of inaccurate boxes. For recall, it is more important to have the whole fetus as much as possible inside the box, so the threshold of 0.9 is used. To measure the accuracy of the final whole fetal localization, similar to the criteria for positive box, same threshold of IoU and precision was used, but a larger recall >= 0.95 were used as a validation threshold because it is more important to have the whole fetus within the bounding box without any missing fetal parts. Note that the purpose is to localize the whole fetus and remove the search area of axial slices as much as possible. This is why higher recall is added instead. The results were also compared between choosing one box and averaging the three boxes' coordinates of each corner that received largest class probabilities from the classifier.

3.4 Results

3.4.1 Structured Random Forests and Edge-box

A qualitative comparison between Structured Random Forests, Canny and Sobel is shown in Figure 3.10. Unlike the other detection methods, Structured Random Forests can ignore the edges inside the fetus. This is important for fetal localization because having more small edges can yield more irrelevant boxes. Canny and Sobel just extract the edges that are above a pre-defined threshold. However, Structured Random Forests gathers and learns from the fetal boundary with the structure pattern based on training data which gives more strength on the fetal boundary, especially the boundary between the fetal front and amniotic fluid.

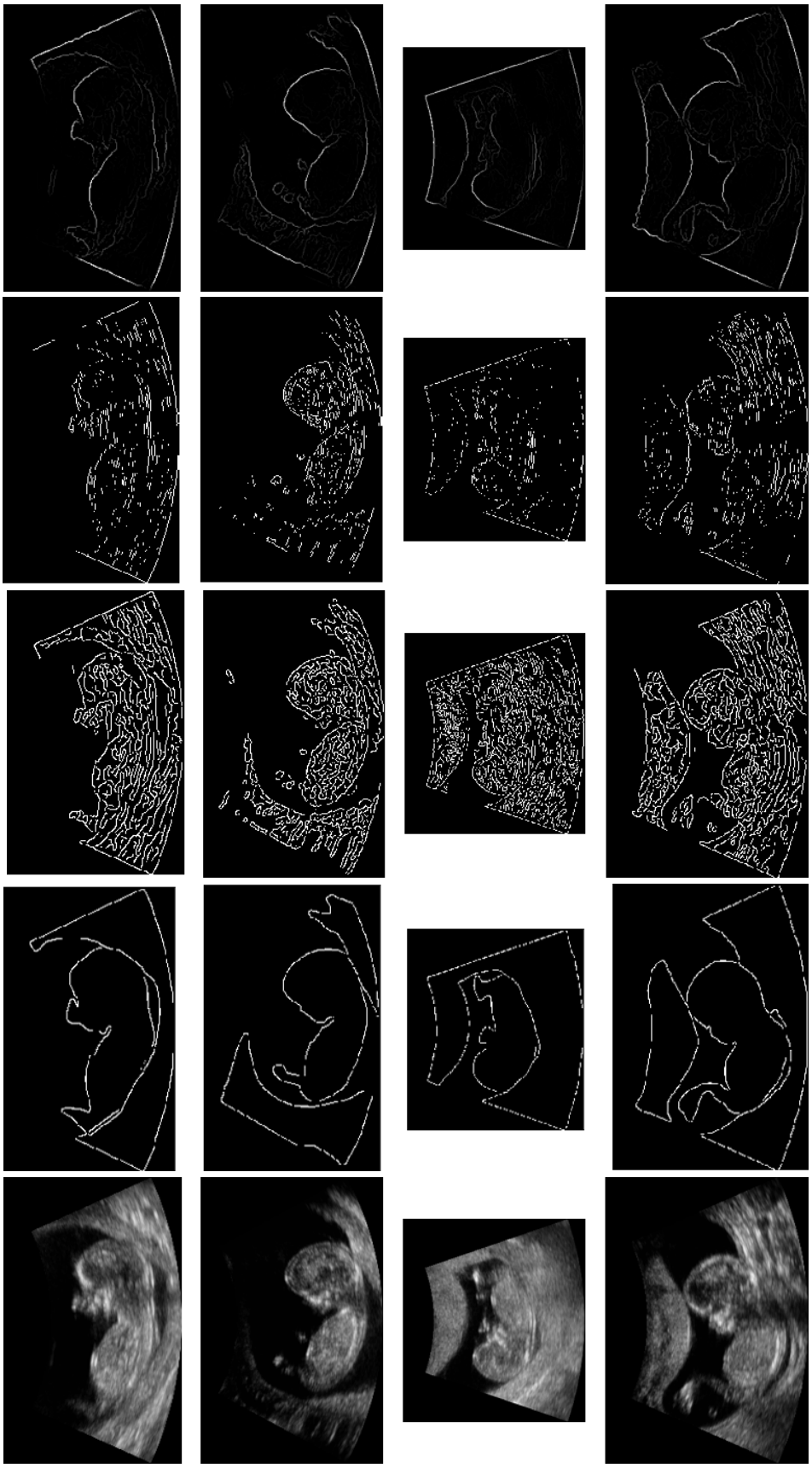


Figure 3.10: Edge Results: By column: a) Raw Image b) Ground-Truth c) Canny d) Sobel e) Structured Random Forests.

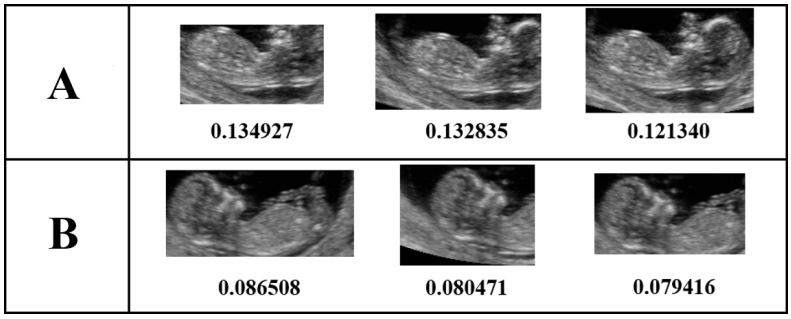


Figure 3.11: Top 3 Candidate boxes based on edge-scores ranking shown for two subjects.

As shown in Figure 3.10 (e), the edge strength is strong between the fetus and amniotic fluid, but it is weak between the fetus and non-fetal tissue. This shows that relying only on edge can mislead to wrong object localization. Figure 3.11 shows the 3 boxes that received top edge scores based on the Edge-box algorithm. However, as shown in Figure 3.11, even though the box obtained high scores among the candidates, it is possible that the box can miss some fetal parts.

3.4.2 Box-Classifier

Selected Features

The feature sets used in the weighted Random Forests classifier for positive boxes were investigated. Figure 3.12 and 3.13 show the total counts for each feature, and how much each feature is used at each tree depths. There are two interesting things to note: (1) the average has been used for first separation and (2) the most used feature is the difference of average while unary is used least. This means that the algorithm first separates the boxes simply by looking at whether the box contains more or less background intensity. After the first split, it then splits based on the pattern from the features provided inside the bounding box.

The classifier for correct and incorrect bounding box achieved an accuracy of 84.1%, sensitivity 59.2% and specificity 88.3%. The full classification is shown in Table 3.1.

Analysis shows that every data has at least one good bounding box which is also classified as positive. Among the positive boxes, the fetal localization estimated from choosing only one box and from the box by averaging the three boxes' coordinates of each corner that received largest class probabilities from the classifier are compared.

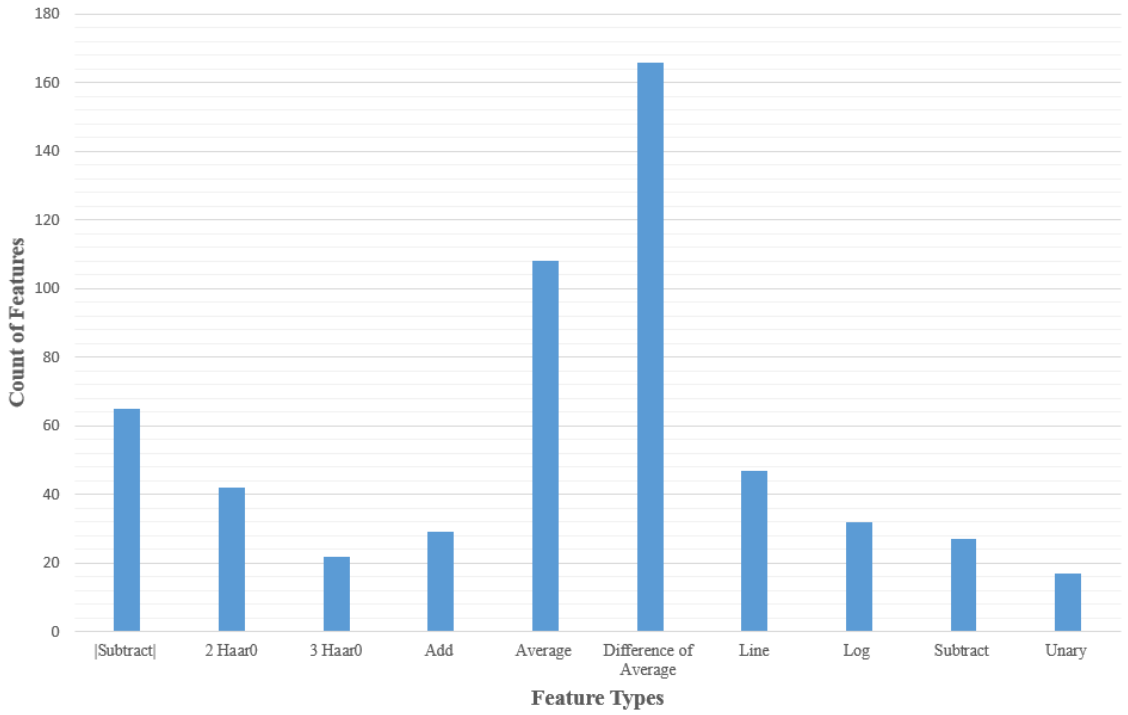


Figure 3.12: The number of each feature type used during the training of Random Forests classifier

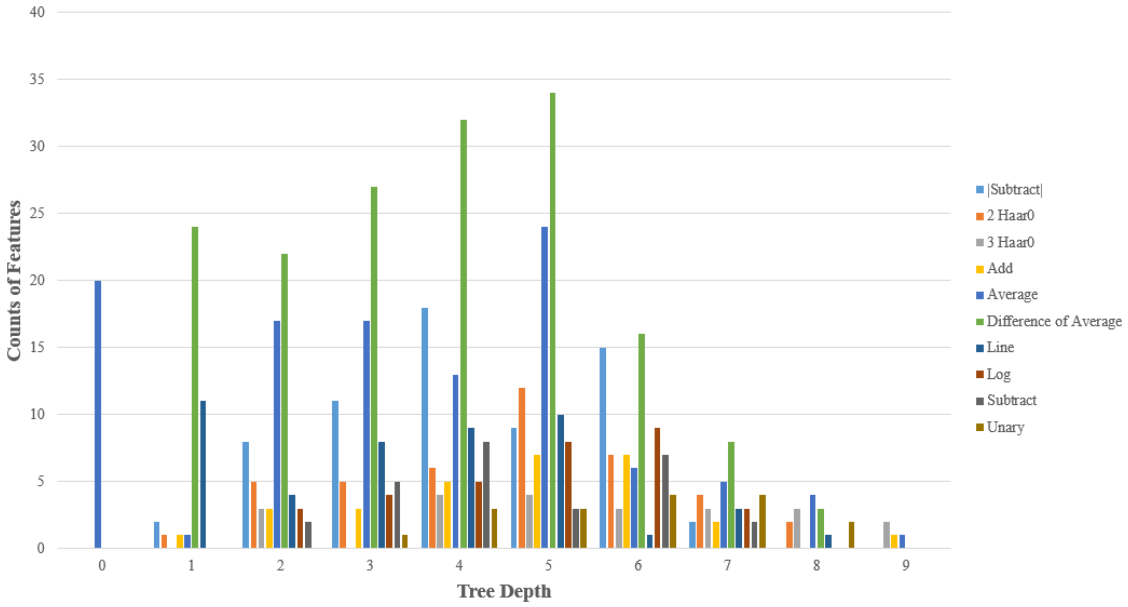


Figure 3.13: The number of each feature types used in each tree depth during the training of Random Forests classifier

	Result		Total
Auto / Manual	Positive	Negative	
Positive	**125**	**86**	**211**
Negative	**145**	**1095**	**1240**
Total	**270**	**1181**	**1451**

Table 3.1: The number of correct and incorrect classifications of bounding boxes

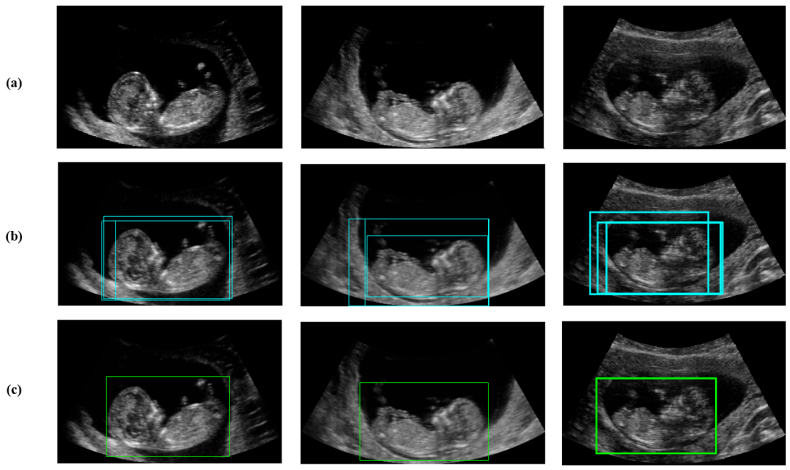


Figure 3.14: The final bounding box by averaging three boxes. Top row represents raw image, middle row represents top 3 boxes with highest class probability and last row represents final box by averaging 3 boxes.

Based on the criteria for final localization as mentioned in section 3.3.2, one-box achieved 68.8% accuracy and averaged box achieved 84.4%. Results from the averaged-three-box are shown in Figure 3.14. Although it received the largest class probability, choosing one box as the final localization might still miss some fetal parts. By averaging three boxes, it allows more boxes to fit on the whole fetus. Averaging five boxes was also analyzed. This achieved 81.3% which is a lower accuracy than that from three averaged boxes. This is because as more boxes are added to be averaged, more irrelevant boxes can be included in the average.

3.5 Summary

In this chapter, a two-stage method based on the Random Forests to localize the whole fetus is described. The first stage need Structured Random Forest is able to put more importance on boundary around the fetus than on internal structures. The second stage

used Random Forests to re-rank the boxes by class probabilities of each candidate boxes, and as a final box, averaging top three boxes than one box achieved better localization. With this localization result, not just the localization of the whole fetus is obtained, but also the search area of fetal parts in axial view is reduced by reducing the non-fetal parts, which will be explained in next chapter.

4

Fetal Partitioning and Biometry Plane Extraction

As mentioned in the literature review of Chapter 2, the assessment of fetal anatomy is the main purpose of the first trimester scan and standardized plane extraction is needed to do this. Therefore, automated fetal partitioning, which is referring to the classification of fetal ultrasound images into 3 classes: head, body and background, and biometry plane extraction are the subject of this chapter. Specifically, Section 4.1 introduces the problem and the proposed solution. Section 4.2 describes a deep learning based algorithm for the diagnostic plane extraction. Section 4.3 presents the experimental setup and validation methodology employed to test the new method. Section 4.4 presents experimental results. A chapter summary is given in Section 4.5.

4.1 Introduction

In Chapter 3, a method for whole fetal localization has been described that uses two consecutive machine learning methods based on the Random Forests framework. Although the whole fetus has been localized, the location of each fetal part is still unknown. Unlike second trimester ultrasound scans which are typically focused on specific fetal structures, the data used in this thesis is based on the whole fetus. Thus, it is essential to estimate where each fetal part is. A typical approach is to do fetal partitioning by

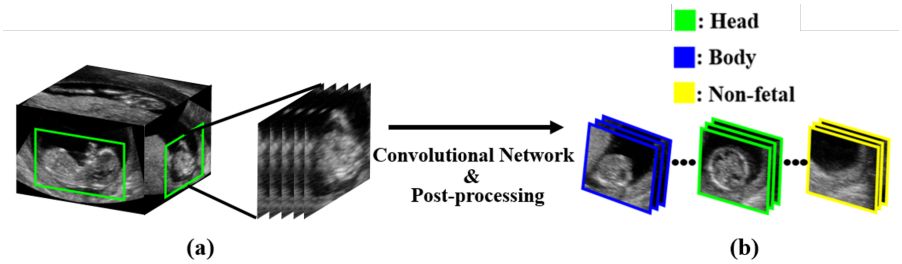


Figure 4.1: Whole algorithm for fetal classification. (a) Based on the bounding box obtained from chapter 3, the axial slices are cropped as pre-processing. (b) Cropped images are then input to a CNN for the classification of the axial slices into three classes (the head, the body and the non-fetal tissue) with subsequent post-processing step.

image classification of each axial slice into the head, body and non-fetal parts. However, image classification throughout the axial slices is challenging in our application because the fetal parts are much smaller than non-fetal parts, as shown in Figure 3.1 (a), and the location of the fetal parts are unknown. Also, the internal organs can appear rotated which can give more complexity. Therefore, the features that are invariant to translation and rotation need to be computed for good classification performance. In this chapter, the whole fetus localization results from Chapter 3 are used to reduce both the area of the non-fetal parts from the axial view and the number of the axial slices to classify. Also, a CNN is used for classification because of its ability to extract the translational and rotational invariant features which can capture both local and global patterns of the anatomies of fetal parts. Based on fetal partitioning, the plane extraction for the brain and abdomen is performed by linear regression.

4.2 Methods

Figure 4.1 shows a flow diagram of the whole algorithm for fetal classification that is explained in this chapter. Based on the localization result from chapter 3, the axial slice is cropped (Figure 4.1 (a)). These cropped images are then used as the input to a CNN and finally outputs the fetal partitioning result (Figure 4.1 (b)).

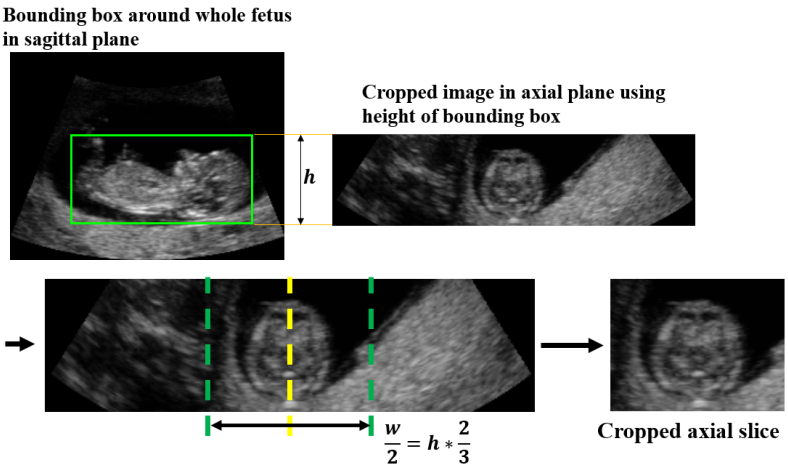


Figure 4.2: Based on the box height, the width of the image in axial plane can be calculated as $\frac{4}{3}$ times the height centered at the middle of the image. Using these, final cropped axial slice is obtained.

4.2.1 Pre-processing Image

An image is first cropped based on the bounding box result from chapter 3. The height of the cropped image is set to be the same as the height of the bounding box. For the width, however, the length has to be longer than the bounding box height because it is possible that the fetus might not lie straight. Because of this, if the length of the width is same as the height, then it is possible that the part of the fetus can be removed from the cropped image. Therefore, the width has to be larger than height. However, it cannot be much larger since the cropped image will then have large area of the background. Here, the width of the cropped image is empirically chosen as $\frac{4}{3} \times height$ as shown in Figure 4.2, which has been found to contain the fetal parts without having a large area of the background. Cropping the image was found to simplify CNN network learning because of less unrelated non-fetal parts.

4.2.2 Convolutional Neural Networks

What is a CNN?

As explained in Section 2.3.1, an artificial neuron simply calculates the weighted sum of the input with the addition of bias, and then decides whether it should be activated or not. A Convolutional Neural Network (CNN) is a multi-layer network which uses convolution as a filter to extract features [123], [126]. Depending on how many filters are

applied, a number of different types of features can be extracted. The difference between a CNN and a simple Fully-connected network is that the CNN does weight sharing, has a local receptive field, and has spatial pooling. Recall from Section 2.3.1, in each hidden layer, the neurons are fully connected to the neurons in the next layer. However, full connection does not exploit the local structure since each neuron is connected to every neuron from the previous layer. Using local structure is based on the strong and general assumption that in an image, as the distance between two pixels increases, the correlation between the two pixels decreases. In a CNN, each neuron is constrained to be connected to a set of neurons that are in a neighbourhood to capture the local structure. This connectivity is called the *receptive field* which affects the activation of the neurons in the next layer. For example, with the certain size of the filter, this subwindow sweeps through the input to extract features. Generally, the size of the receptive field is small relative to the size of the input which means that the neurons generally form a local rather than a global receptive field. With such a receptive field, a CNN learns the set of filters $w_{i|i=1,2,3,\ldots,n}$ across the multiple subwindows within the input. Since the same filter is applied over the multiple subwindows, the weights are shared among the neurons in the hidden layer. The weight sharing has two main advantages. One is that the network can learn a general representation of the input data which allows for better generalization. Another is that it can reduce the number of parameters which makes training easier and more efficient. For example, if a certain feature is useful in a certain position in the input, it can also be useful in other positions. Based on this idea, instead of computing every possible parameter, the number of parameters can be substantially reduced. Finally, a CNN is different from a simple fully-connected network in that it uses a pooling layer for sub-sampling. The purpose of a pooling layer is to reduce the number of parameters of a convolutional response and to provide translational invariance in a model, which helps reduce the sensitivity of the learnt features to object locations in the image. Max-pooling and average-pooling are the two common pooling methods used in a CNN [169]. In this thesis, max-pooling is used. Max-pooling takes the maximum among the values inside the subwindow and passes this value to the next layer as shown in Figure 4.3. Repeating this *convolution+pooling* process, the dimensionality

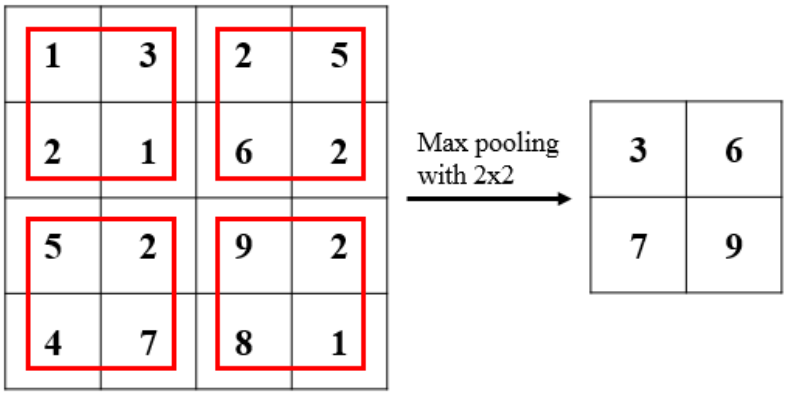


Figure 4.3: The example of max-pooling with 2×2 filter size with stride 2

of a feature map is reduced while having features that represent a whole image. Since it only considers maximum activation no matter where that activation is, the extracted features are translation and rotation invariance. These two invariances are important for fetal data, especially for the data in this thesis. Since the data has been taken based on the whole fetus without a constraint on fetal pose, the location of the fetal parts in axial view is unknown and the fetal parts can also be rotated. Therefore, translational and rotational invariance are important for the detection of the fetal parts.

CNN Architecture and Training

There has been a recent historical trend in CNN architecture design for computer vision: AlexNet [126], VGGnet [134] and ResNet [135]. They have shown state of art results on natural image classification and object recognition. The base architecture, such as types of layers, is similar for all popular CNN architectures. In this chapter, the network designs are based on the AlexNet. The AlexNet is shallow compared to other networks such as ResNet. The architecture used in this thesis is shown in Figure 4.4. The AlexNet is composed of several types of layers: a convolutional, an activation, a pooling and a fully-connected layer. The convolutional layer applies a number of convolutional filters to the input to get different types of feature maps. As described earlier, a number of types of features can be controlled by setting up the number of filters. Having a large number of filters means that there will be more features for decision making. Similar to the network described in chapter 2, the output Y_{ij}^{l} of the l^{th} layer of the convolutional

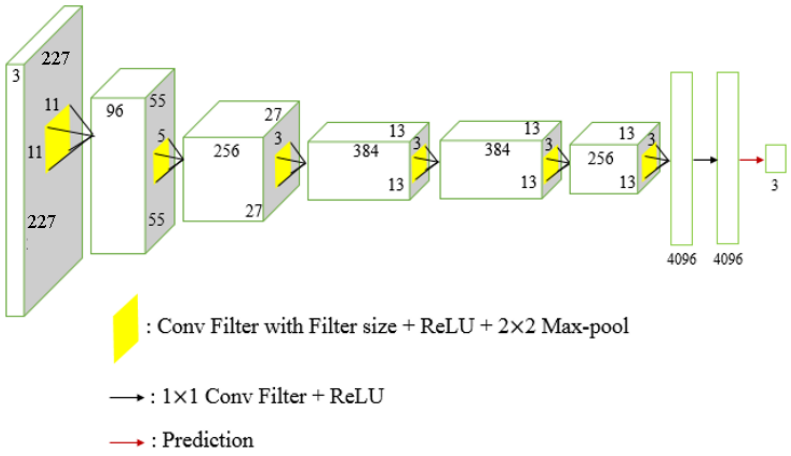


Figure 4.4: The architecture of CNN based on AlexNet

neural network by the feedforward is,

$$Y_{ij}^{l} = \sum_{q=0}^{m-1} \sum_{r=0}^{m-1} w_{qr}^{l} Y_{i+q,j+r}^{l-1} + b_l \tag{4.1}$$

where $m \times m$ is the size of filters, w_{qr}^{l} is the filter which parameter is to be learned during the training and b_l is the bias.

The size of the output is decided based on the three parameters: the filter size, the number of strides and the amount of zero-padding. The filter size is decided based on how many feature types to be used. This is important because depending on the task, the filter size can affect the result. For example, the task that requires more information or guidelines to be achieved is difficult and complex, so it will be helpful to extract and use more features. The stride controls how the filter convolves around the input. For example, if it is 1, the filter moves by one pixel, and when it is 2, the filter jumps 2 pixels as it slides, which will output a smaller volume size than when the stride is 1. The larger the stride, less overlapped will be the filters. However, the number of strides needs to be considered based on the input size. For example, if the input size is 5×5 and 3×3 filter with the number of stride is 3, this will make the filters unable to slide through all the regions of the input as shown in Figure 4.5. Zero-padding pads the input with zeros around the border. The main purpose of zero-padding is to control the spatial dimensions of the output to preserve the size of the output same as the input. With these three parameters, the size of the output volume can be controlled and to calculate the size of the output,

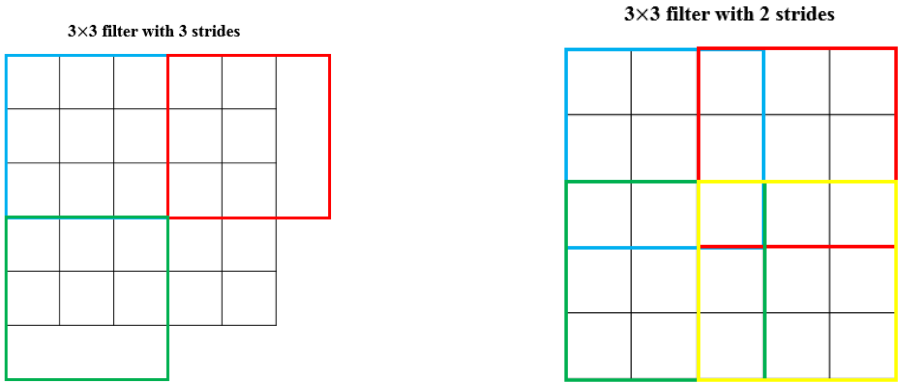


Figure 4.5: 5×5 input and 3×3 filter with a) 3 strides b) 2 strides. If 3 strides is applied, the filter is not applicable because it does not fit to the input size. 2 strides, however, can be applicable.

$\frac{W-F+2P}{S} + 1$ is used, where W is the input size (width and height), F is the filter size, P is the amount of zero-padding and S is the number of strides. As an example, in the architecture that is used in this chapter, the input is $227 \times 227 \times 3$ and 96 11×11 filters have been applied with 4 strides. This yields an output of $55 \times 55 \times 96$ pixels.

Activation Layer The output of a convolutional layer, x^l, is passed to the activation layer, where it decides whether it should be activated or not. Since x^l can take any values ranging from $-\infty to +\infty$, the activation layer introduces an activation function to define the decision boundary. There are several activation functions for example, the *sigmoid* and *the Rectified Linear Unit* (ReLU),

$$Sigmoid : \frac{1}{1+e^{-x^l}}, \tag{4.2}$$

$$ReLU : max(0, x^l). \tag{4.3}$$

Here, x^l is the output, Y^l, from the convolutional layer. It is important to note that these two functions introduce a nonlinearity into the network. If they were linear, no matter how many layers have been used, the network would be the same as one single layer because adding several linear functions still gives a linear function. This also indicates that the network considers the relationship between the input and output as linear, which is a weak model for a real world problem, which are generally non-linear problems.

If the sigmoid function is used as the activation function, it has a problem called

(a) **(b)**

Figure 4.6: Activation function a) Sigmoid b) ReLU

Gradient Vanishing. As can be seen in Figure 4.6(a), a sigmoid function outputs a value between 0 and 1. The weights in the network can be updated by using the gradient (or derivative) of the loss function with respect to the current weight. Since the sigmoid function ranges from 0 to 1, the gradient will be small especially for the saturated region where the output value is almost 1 as shown in Figure 4.6(a). Therefore, as the number of layers increases, the gradient of the loss function will eventually take a value near zero and results in a slow update of the weights. To overcome this problem, the ReLU has been proposed [170]. The ReLU passes the input value if it is bigger than 0, but otherwise, it passes the values as 0 as shown in Figure 4.6(b). ReLU has several advantages. Firstly, it reduces the gradient vanishing problem because it passes the original value if it is larger than 0 which gives no saturation. Second, it has a sparse activation property. Imagine that the network has a lot of neurons. Using an activation function such as a sigmoid will possibly activate all neurons since they are non-zero values although they are really small. However, due to the characteristic of a ReLU, it will not activate since it gives a zero value. This means that only necessary neurons will activate and it leads to less computation due to the smaller number of the activations.

Pooling Layer After the activation layer, a pooling layer is used. As mentioned before, max-pooling is used, where the output, Y^l is,

$$Y^l = max(Y^l(i,j), Y^{l-1}(i+q, j+r)) \tag{4.4}$$

which compares the output value of (i,j) in the patch with the output values of other pixels $(i+q, j+r)$ in the same patch to output the maximum value and also to save the

index to be used during in backpropagation. These three layer types (conv + ReLU + maxpool) are repeated until it reaches the fully-connected layer.

Fully-Connected Layer As the name suggests, all neurons are fully connected to all activations from the previous layer similar to the fully-connected network explained in chapter 2, where the convolutional layers are connected only to the local region. As already mentioned, a convolutional layer extracts the features from the input data. However, the input still needs to be classified, if the task is classification, based on these extracted features. By adding the FC layers, the features are computed as the mapping to the outputs. The size of the output depends on how many classes are classified, which in our case is three: head, body, non-fetal parts. Specifically,

$$Y^l = \sum w_l * Y^{l-1} \tag{4.5}$$

where Y_l is the output of the l^{th} layer and w_l is the weight applied to the $(l-1)^{th}$ layer. If it is the last fully connected layer, the output, Y_l will be the prediction for the network. Also, in every FC layer, except the last layer, ReLU and dropout [171] are added. Dropout is a method to prevent overfitting which occurs when there are too many parameters relative to the number of observations. In every training, the randomly chosen outputs of the previous layer are dropped and their corresponding weights are not updated. However, during the test, the network uses all the weights. This is similar to ensemble learning which the weak classifiers are trained separately and then combine them at test time. Without dropout, the network might become "too confident" within the training data producing poor results in testing.

Loss The loss shows how close the prediction is to the groundtruth. In the classification task, for example, the loss represents how many data are correctly classified. To calculate the loss, there are several types of loss function, and in this work, *softmax log loss* or *multinomial logistic function*, L is used,

$$p_{Y_c} = \frac{e^{Y_c}}{\sum_{k=1}^{3} e^{Y_k}} \tag{4.6}$$

$$L = -log(p_{Y_c}) = -log\left(\frac{e^{Y_c}}{\sum_{k=1}^{3} e^{(Y_k)}}\right) = -Y_c + log\sum_{k=1}^{3} e^{Y_k} \tag{4.7}$$

where p_{Y_c} is the softmax function, $Y_k|_{k=1,2,3}$ is the output from the network and Y_c is the correct prediction.

As explained in chapter 2, once the loss is obtained, a CNN performs backpropagation using the chain rule. The basic process is the same for the example given in chapter 2, except the CNN uses convolutional filters that its parameters need to be updated similar to how the network in chapter 2 updates the weights. Also, a CNN involves the pooling and the activation, but there are no parameter updates in those layers.

Backpropagation Similar to backpropagation as explained in chapter 2, the chain rule is also used in a CNN. Once the loss, L, is obtained, the derivatives of the losses respect to the output of the last FC layer, $\frac{\partial L_{total}}{\partial Y_k}$, is computed first,

$$\frac{\partial L_{total}}{\partial Y_k} = -\frac{\partial}{\partial Y_k}\left(-Y_c - log\sum_{k=1}^{3} e^{Y_k}\right) = -\left(\delta_{k=c} - \frac{e^{Y_c}}{\sum_{k=1}^{3} e^{Y_k}})\right) \tag{4.8}$$

where, δ is *kronecker delta*,

$$\delta = \begin{cases} 0, & \text{if } k \neq c. \\ 1, & \text{if } k = c. \end{cases} \tag{4.9}$$

Backpropagation then computes the derivative with error respect to the weights, $\frac{\partial L}{\partial w_l}$,

$$\frac{\partial L}{\partial w_l} = \frac{\partial L}{\partial Y^l}\frac{\partial Y^l}{\partial w_l} = \frac{\partial L}{\partial Y^l}Y^{l-1} \tag{4.10}$$

where $\frac{\partial L}{\partial Y^l}$ ($\frac{\partial L_{total}}{\partial Y_k}$ if it is from the last layer) has been already calculated, and Y^{l-1} is already known as well since this is the output of the upper layer, $(l-1)^{th}$. This process is repeated until the first FC layer.

Backpropagation of max-pooling layer has to be computed before the convolutional layer. However, since there are no weights to be computed, there is no learning in this layer, which in other words means no weights to be updated. Instead, it passes the derivative of the error with respect to the output of pooling layer which is the maximum value of the corresponding pixels whereas others are 0. Similar to max-pooling, the ReLU does not

have weights to be updated. Therefore, the backpropagation in ReLU, $\frac{\partial L}{\partial Y_{l-1}}$, is simply the derivative of the error with respect to the input of the ReLU,

$$\frac{\partial L}{\partial Y_{l-1}} = \begin{cases} 0, & \text{if } Y_l < 0. \\ \frac{\partial L}{\partial Y_l}, & \text{otherwise.} \end{cases} \tag{4.11}$$

where Y_{l-1} is the output of the convolutional layer.

The backpropagation in the convolutional layer is similar to the FC layer but involves a convolution.

$$\frac{\partial L}{\partial w_{qr}} = \sum_{i=0}^{N-m}\sum_{j=0}^{N-m} \frac{\partial L}{\partial x_{ij}^l}\frac{\partial x_{ij}^l}{\partial w_{qr}} = \sum_{i=0}^{N-m}\sum_{j=0}^{N-m} \frac{\partial L}{\partial x_{ij}^l} Y_{(i+q)(j+r)}^{l-1} \tag{4.12}$$

$$\frac{\partial L}{\partial x_{ij}^l} = \frac{\partial L}{\partial Y i j^l}\frac{\partial Y_{ij}^l}{\partial x_{ij}^l} = \frac{\partial L}{\partial Y_{ij}^l}\frac{\partial}{\partial x_{ij}^l}\left(\sigma(x_{ij}^l)\right) = \frac{\partial L}{\partial Y_{ij}^l}\left(\sigma'(x_{ij}^l)\right) \tag{4.13}$$

where, $\frac{\partial L}{\partial Y_{ij}^l}$ is already known from the backpropagation-output from ReLU, and σ is the activation function. Once all the $\frac{\partial L}{\partial w}$ are obtained, ws are then updated using Equation 4.7 and 4.10.

Training The network follows the AlexNet architecture implemented using MatConvNet [172] and trained on a Intel Xeon CPU at 3.50 GHz with 16.0 GB RAM. The original parameter values, which are weights and biases, from the AlexNet have been imported as initial parameter of the network to reduce the convergence time. Since the AlexNet has been trained with the large number of natural images, it is possible that instead of using zeros or random numbers as the initialization, their parameter values are good initial values to train with. Tajbakhsh et al. [133], for example, showed that a fine-tuned CNN outperformed the CNN trained from the scratch when only limited training data are available. The learning rate was chosen to be the distribution from 1e-3 to 1e-5 in 50 epochs. The weight decay, which regularizes the weight changes, was chosen to be 0.005 and the momentum parameter 0.9. These hyperparameters can be optimized, however, we use the most standard values that have been applied in both natural and medical images [173], [174], [132], [175]

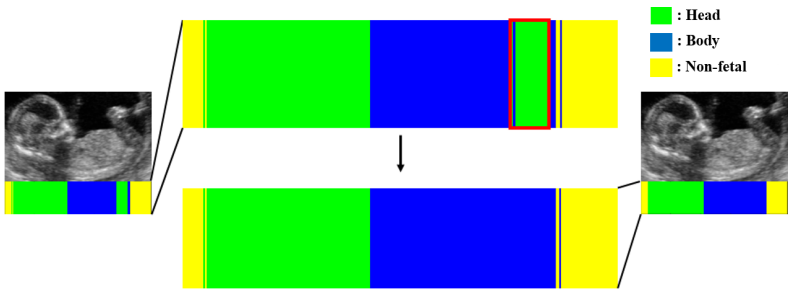


Figure 4.7: Exploiting the anatomical constraint. a) Acquire the set of axial planes which belong to the head and body. b) Misclassified slices (red box) are corrected since the head is on the left according to the partitioning result. c) The final partitioning result.

4.2.3 Extraction of Best Plane for Fetal Brain and Abdomen Biometry

As shown in Figure 4.7, the set of axial planes which belong to the head and body are found by image classification using a CNN. An anatomical constraint is then exploited that if the head is found to be located to the left relative to the body, body slices should not appear on the left side of the image. Using this constraint, image slices located on the left which have been previously classified as body are reassigned to the head class. Based on the partitioned regions of the fetal slices, a greedy approach is used to find the best plane for both the head and abdomen. In this work, the assumption has been made that the positions of the diagnostic plane of the head and the abdomen with respect to the fetal head and torso length are similar among the normal fetus of similar gestational age. Therefore, the head and abdomen fetal biometry planes are estimated based on a linear regression of the head and torso length. Two regression approaches are performed to approximate the best plane of the brain and the abdomen. From the training data, a linear regression is performed to predict the distance of the best head plane from the approximate fetal crown as a function of the length of the head. Similarly, another linear regression is performed for the best abdominal plane from the approximate fetal rump as a function of the length of the body.

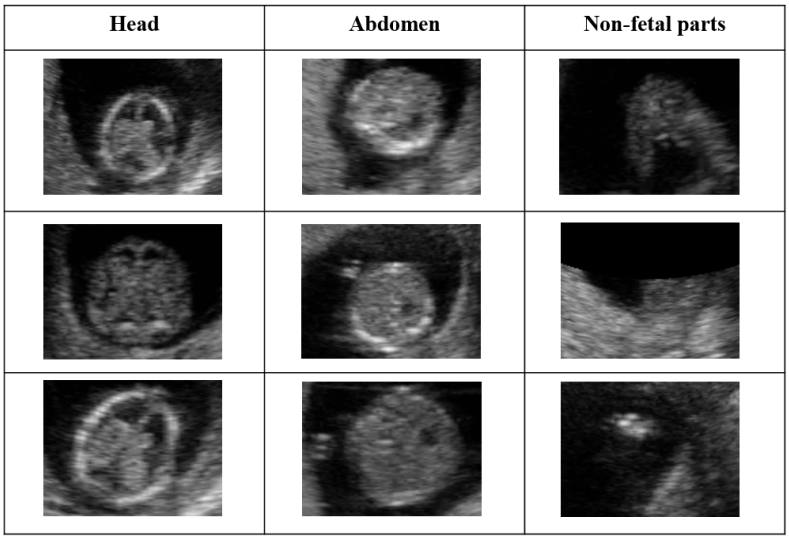


Figure 4.8: Example of axial slices for each classes: Head, body and non-fetal parts

4.3 Experimental Setup

4.3.1 Datasets and Validation Methodology

The data used was introduced in chapter 3 (see Section 3.3.1). As the training set, the axial slices were manually cropped as described in Section 4.2.1. Each axial image was manually labelled as three classes: head, body and non-fetal parts as shown in Figure 4.8. The expert also specified the best manual biometry plane for the head and abdomen which formed the groundtruth to compare with automatically extracted biometry planes.

Data Augmentation

As mentioned in Section 2.3.2, one of the major challenges for medical imaging compared to natural images is its small data size. There are many variances in the medical images, however, if the data size is small, these variances cannot be covered. One solution is *Data Augmentation*, which allows to augment the original data to increase training data size, improve model robustness to geometric variations, and reduce overfitting. Number of papers have shown that the data augmentation increased the accuracy [176], [177], [178]. The data augmentation was performed by flipping left-right, which mimics flipping the probe, hence provide richer data.

4.3.2 Evaluation

To compare the performance, three-fold cross validation was performed. CNN with and without the pre-processing step as shown in Figure 4.2 have been compared. The method was also compared with a hand-crafted method using SIFT features with fisher vector following [100] [1] trained and tested with the same amount of pre-processed images used for CNN as shown in Figure 4.8 (3403 slices for head, 5460 slices for body, 4154 slices for non-fetal parts). Also, the dimensionality of SIFT descriptors was reduced to 80 using PCA and 256 components of Gaussian Mixture Model has been used, which showed the improvement in the accuracy according to [100].

4.4 Results

4.4.1 Classification Results

Classification results for the three methods are summarized in Table 1 and as confusion matrices in Figure 4.9. The average classification accuracy of the traditional method following [100] achieved 75.2% $\pm$2.9 and CNN with the input as raw image achieved 73.0% $\pm$4.8. However, the average classification accuracy using the CNN with pre-processed images based on the auto-cropped volume achieved 76.9% $\pm$2.7 which was the best result. By performing *t-test*, which tells how significant the differences of two means are, the difference was statistically significant with a two-tailed *p-value* of less than 0.05 ($p-value = 0.041$ ([100] and CNN with raw image), 0.00024 (CNN with raw and pre-processed image) and 0.046 ([100] and CNN with pre-processed image)). First, the pre-processing step based on the localization result from Chapter 3 has shown improvements in accuracy. As mentioned before, the raw image contains larger area of non-fetal parts which makes the task to be more difficult. However, by auto-cropping the image to be more focused on the fetal parts, the accuracy result has been improved. This indicates that having the image that is more focused on the fetal parts can help the network to learn better.

[1] As part of the research, the algorithm code is provided by Maraci.

Method	Mean Classification Accuracy (%)	Head Accuracy (%)	Body Accuracy (%)	Non-fetal Accuracy (%)
Maraci et al. [100]	75.2±2.9	86.4	87.1	52.6
CNN using raw image	73.0±4.8	89.4	79.7	52.6
CNN using pre-processed image	76.9±2.7	92.3	83.5	56.6

Table 4.1: Classification accuracy comparison for different methods.

The results indicate that the CNN is able to capture the specific patterns of each important fetal part without any handcrafted features provided. The method following [100] is also able to capture the important patterns of each brain and abdomen. However, the feature types to extract the useful information are provided by the human. A CNN, however, finds and learns the important features straight from the data itself. Specifically, as shown in Table 4.1 and Figure 4.9, CNN using the pre-processed image performed better for the head and the non-fetal parts. The method following [100] shows better results in the body classification. This might indicate that a CNN can perform better for more complex pattern since the brain image has more patterns than the abdomen does. A CNN also performed better in distinguishing between the head and the body according to Figure 4.9. For instance, CNN using pre-processed image has 68 head images that are misclassified as body, whereas [100] has 187 misclassified images. Also, the non-fetal classifications showed bad performance from the three methods. This might be due to the fact that the neck-region is removed from the training set to avoid the confusion. However, since they are included in the test set, the network might classify that region as either the head or body even though they are considered as the non-fetal parts. Also, note that according to Figure 4.9, all three methods confused the body as the non-fetal parts. Since the only characteristic of the abdomen is stomach bubble and if the stomach bubble is not visualized, it might be similar to the non-fetal tissue. In clinical perspective, the misclassification between the head and the body is more serious than misclassification of the non-fetal because each fetal anatomy has their own guidelines for anatomical assessments. If the algorithm misclassifies fetal anatomies, the clinicians need to override the decision, otherwise misdiagnosis happens. Therefore, as clinical perspective, CNN using pre-processed image is more accurate to the clinical application than [100].

The specific example of fetal partitioning is shown in Figure 4.10 (a), and the qualitative results with more examples are shown in Figure 4.10 (b). It was also found that all of the groundtruth biometry planes were within the partitioned range of the head and the body as shown in Figure 4.12 (a).

Auto \ Manual	Head	Body	Non-fetal
Head	92.3 %	2.0 %	6.7 %
Body	2.7 %	83.5 %	13.8 %
Non-fetal	25.6 %	17.8 %	56.6 %

(a)

Auto \ Manual	Head	Body	Non-fetal
Head	89.4 %	5.8 %	4.8 %
Body	4.2 %	79.7 %	16.2 %
Non-fetal	23.9 %	23.5 %	52.6 %

(b)

Auto \ Manual	Head	Body	Non-fetal
Head	86.4 %	5.5 %	8.1 %
Body	3.8 %	87.1 %	9.0 %
Non-fetal	17.1 %	30.3 %	52.6 %

(c)

Figure 4.9: The confusion matrix for a) CNN using pre-processed image, b) CNN using raw image, c) Maraci et al. [100].

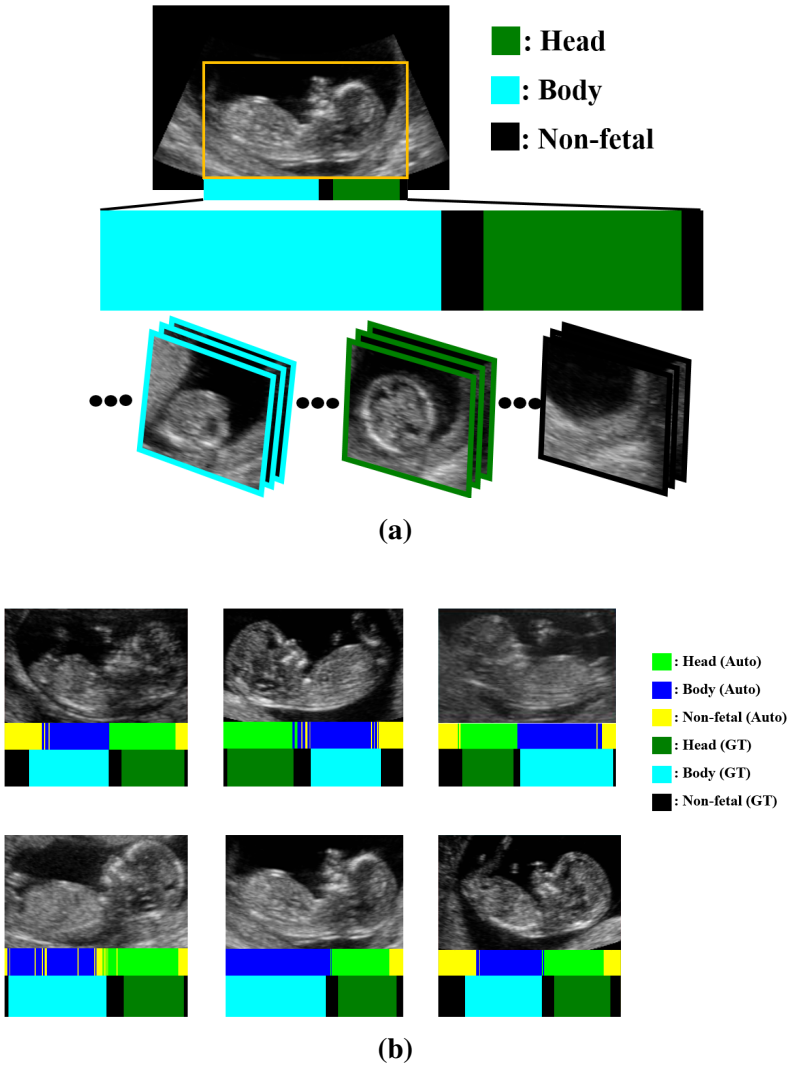


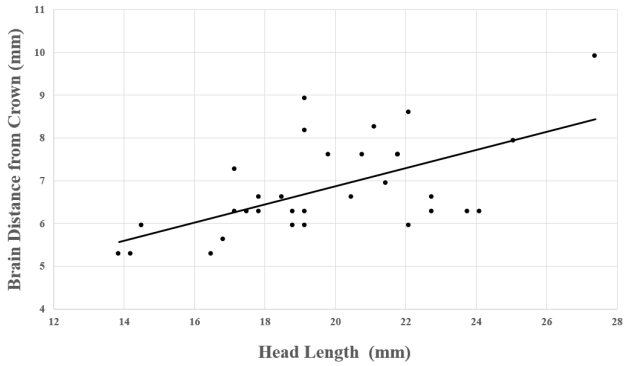
Figure 4.10: a) The example of the partitioning result b) The examples of partition results based on CNN with input as cropped image resulted from the localization.

4.4.2 Extraction of Biometry Planes

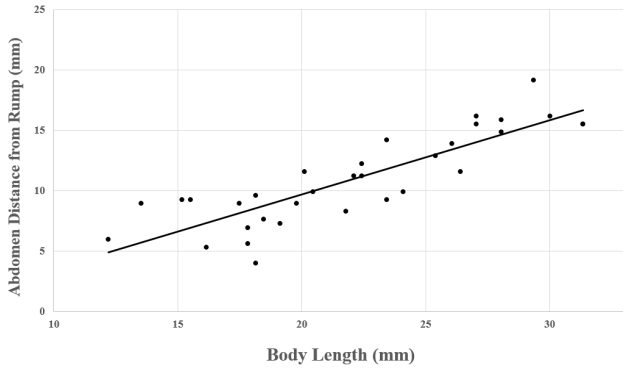
As mentioned in Section 4.2.3, two linear regressions have been performed to approximate the location of the best plane of the brain and abdomen based on the head body length without considering the orientations. As shown in Figure 4.11, linear regression is performed. The resulting equations for brain and abdomen plane location definition are,

$$Plane_{Brain}(mm) = 0.21 \times Head\ Length(mm) + 2.63 \tag{4.14}$$

$$Plane_{Abdomen}(mm) = 0.62 \times Body\ Length(mm) - 2.66 \tag{4.15}$$



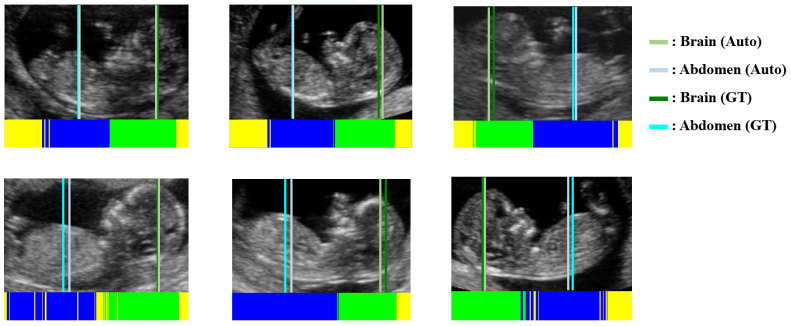
(a)



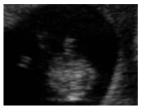
(b)

Figure 4.11: The linear regression approach for a)Brain and b) Abdomen. For example, the brain is predicted from the distance of the best head plane from the approximate fetal crown as a function of the length of the head. Same process for the abdomen.

Based on these equations, the diagnostic plane for brain and abdomen can be chosen. However, as the R^2 value indicates in Figure 4.11, there are more variance, because the orientation is not considered in this chapter. As a result, the mean distance between the manually and automatically extracted biometry planes is 1.6 ± 0.2mm for the fetal brain and 3.4 ± 0.4mm for the fetal abdomen. Some typical results are shown in Figure 4.12. Both extracted plane showed similar results compared to the manually extracted plane, but abdomen showed larger distance compared to the brain. These might be caused by not considering the orientation of the fetal parts.



(a)

	Head		Abdomen	
	Good	Bad	Good	Bad
Manual				
Automatic				
Difference	0.3 mm	2.6 mm	0.3 mm	11.9 mm

(b)

Figure 4.12: The extraction of best plane for brain and abdomen for biometry measurements a) Locations of the best planes from ground-truth and automatically extracted, and b) Manually selected and automatically selected best planes for the fetal head and fetal abdomen with distances of difference between them.

4.5 Summary

This chapter has described initial experiments to classify each axial slice into one of 3 classes (head, body and non-fetal) using a CNN architecture and extracting the best plane for brain and abdomen using the linear regression. Using a CNN has shown better results than using the machine learning method following Maraci, et al. [100] which the type of the feature is introduced by human. Also, by pre-processing the input by using localization result from the work in the chapter 3, the classification accuracy has been improved. This shows two interesting facts that the non-fetal parts affect the classification accuracy that the result has been improved significantly by having images which are more focused on the fetal parts. The second is that the neural network can be improved not only by changing the architecture, but also by introducing pre-processing of the input.

Based on this partition result, extracting the best plane has also shown good results using linear regression compared to the manually extracted plane.

5

Fully Convolutional Network based approach for Fetal Biometry Assessment

In Chapter 3 and Chapter 4, both fetal localization of the whole fetus and the fetal partitioning using a machine learning method have shown promising results. However, the methods used in Chapter 3 and 4 require several steps. This chapter introduces a different approach which we show to have similar performance for the localization of the fetus and fetal partitioning using a single neural network. Also we show, by using a similar neural network architecture, that first trimester fetal assessment of other fetal parts (fetal head, abdomen and limbs) can be performed.

Specifically, Section 5.1 outlines assumptions and limitations of the methods described in the previous chapter. Section 5.2 describes how fetal assessment has been achieved based on a deep learning method. Section 5.3 presents experimental setup. Section 5.4 presents the evaluation, and Section 5.5 presents experimental results. A summary is given in Section 5.6. The overall concept is shown in Figure 5.1.

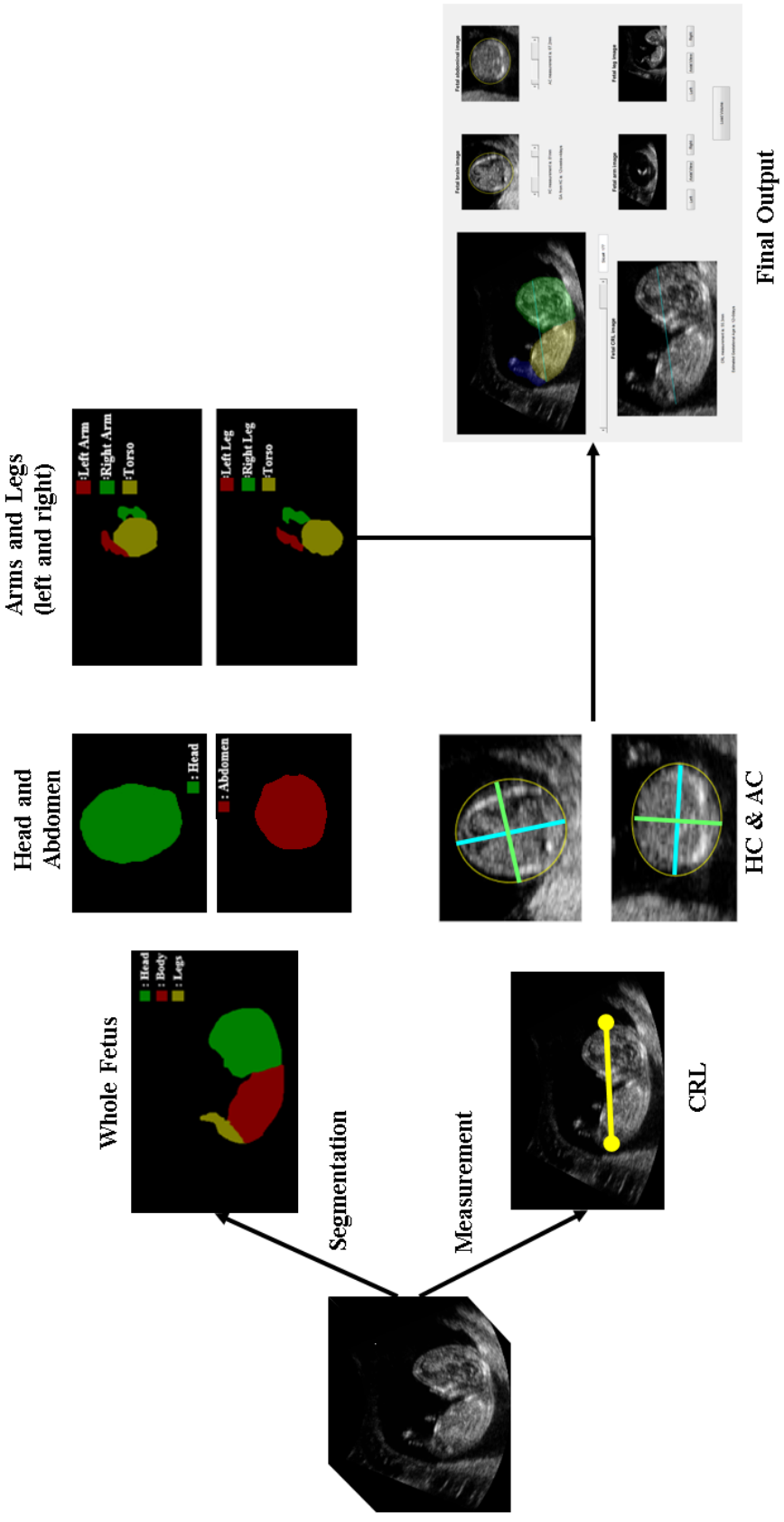
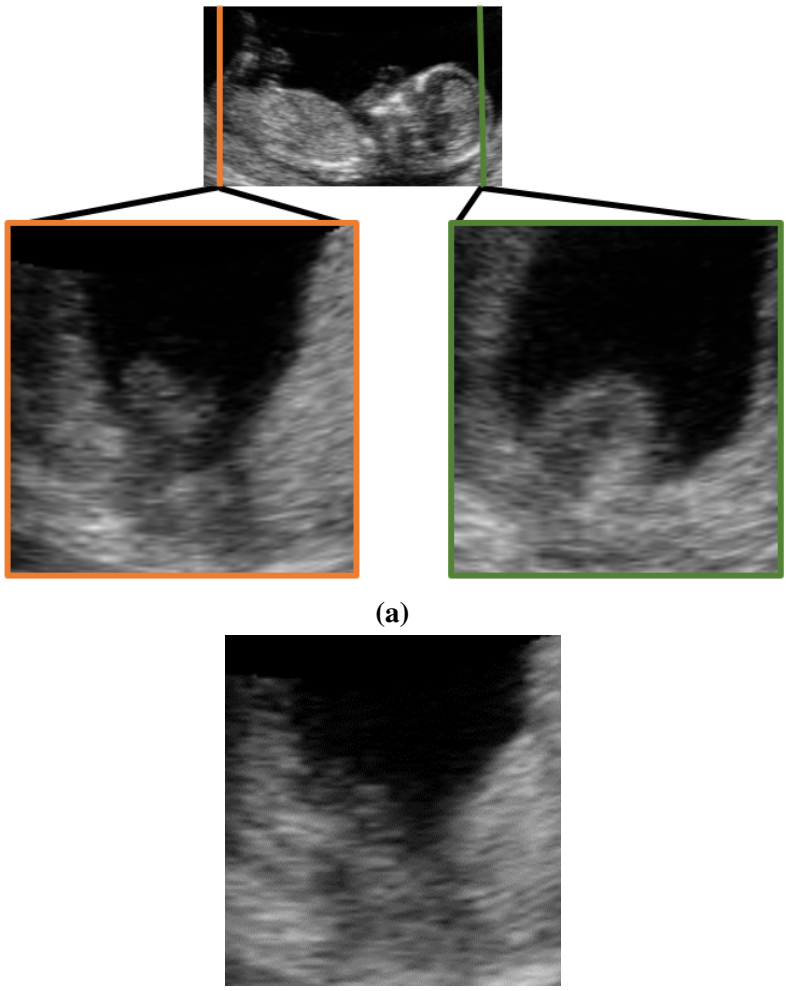


Figure 5.1: The overall concept of Chapter 5. From the 3D volume, the segmentation of the fetal anatomies and automatic measurements are performed. Then finally, the integrated software interface displays these results.

5.1 Introduction

In the previous two chapters, the localization and the visualization of the fetal head and abdomen were achieved by using two Random Forests and a CNN based on the 2D slices extracted from a 3D volume. However, there are two limitations of those methods. First, they require several steps. Moreover, the localization of the whole fetus is first achieved by identifying the bounding box around the fetus based on edge detection, not on the fetus itself. The internal structure of the head contain image patterns that can be used for the localization of the whole fetus. However, edge detection is only focused on the edges, so there is a high possibility that it will perform poorly if the edges around the fetus are not distinct. This is not unusual in ultrasound images where *signal-to-noise ratio* (SNR) and *contrast-to-noise ratio* (CNR) are not always high. Another limitation is that the classification of the axial slices for fetal partitioning cannot identify the fetal parts on the point at the boundary between the fetus and non-fetal part as illustrated in Figure 5.2. As shown in Figure 5.2, sometimes there is no difference in image pattern between the fetal part and the non-fetal part. This is problematic, as the CRL is measured between the crown and rump, but these landmarks are located at the boundary of the fetus as shown in Figure 5.2 (a). Therefore, inaccurate landmark placement can result in an under-estimated CRL measurement, which in turn will lead to an under-estimation of the gestational age. To overcome these problems, a method for more accurate fetal localization, that is based on the fetus itself with fewer steps is described. Adopting fewer step procedure is beneficial due to the less propagation of errors and computational time. As mentioned before, in the previous two chapters, since the each step can highly affect the next step, it is highly possible that if the fist step fails, the large errors will be propagated. Also, since it requires more steps, the computational time will be increased as well.

Specifically, a deep learning method known as a *Fully Convolutional Network* (FCN) [136] is used to automatically annotate the fetus in the sagittal view for both localization and fetal partitioning. In the previous chapters, we assumed that the best sagittal plane has been given. In this chapter, this assumption is relaxed and the sagittal plane is automatically chosen based on the classification of sagittal slices.



(a)

(b)

Figure 5.2: The example of the slice at the point where the boundary between the fetus and non-fetal part is. a) The axial slice of the fetal crown (green) and rump (orange), b) The non-fetal part.

Based on the segmentation result, the orientations of both the head and the torso are considered to obtain a good visualization of the head and the abdomen. Even though the previous results showed that the distance between the manually and automatically extracted planes is relatively small compared to the fetal size, the visualization of the plane is still not good enough for the diagnosis due to the difference in plane orientation from standardized plane orientation that would be achieved with a 2D acquisition. As shown in Figure 5.3, the butterfly shape of the choroid plexus is clearly visible when the orientation is correct. Since the fetus is not always in the same orientation as the standardized plane orientation, it is necessary to find the standardized plane. This is easy

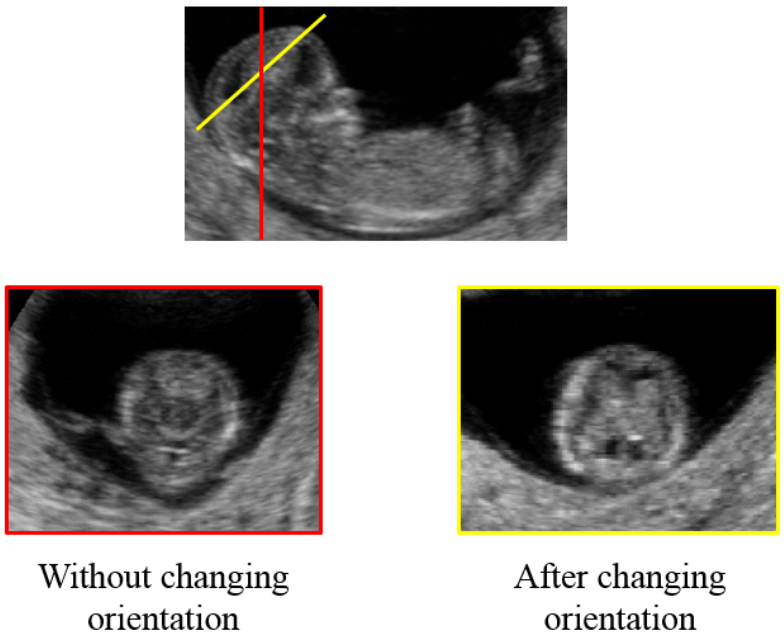


Figure 5.3: An example of the axial slice of the fetal head a) without changing orientation (best axial slice), b) after changing orientation. The internal structure is much clearer when the orientation is changed according to the head orientation.

to do if you can estimate head orientation. Finally, the automatic measurements (CRL, HC and AC) can be performed.

Automatic limb detection and assessment are harder as the limbs are much smaller with varied image appearance. A FCN-based network was developed for the localization/segmentation of the limbs and based on the segmentation result, the orientations of the limbs are corrected to obtain a good visualization of the limbs.

5.2 Methods

5.2.1 Fully Convolutional Network architecture

We first review the principles of FCNs. A Fully Convolutional Network (FCN) [136] is based on a CNN. However, the main difference is that the FCN outputs pixel-wise prediction, while the CNN for the classification outputs the classification probability or scores as shown in Figure 5.4.

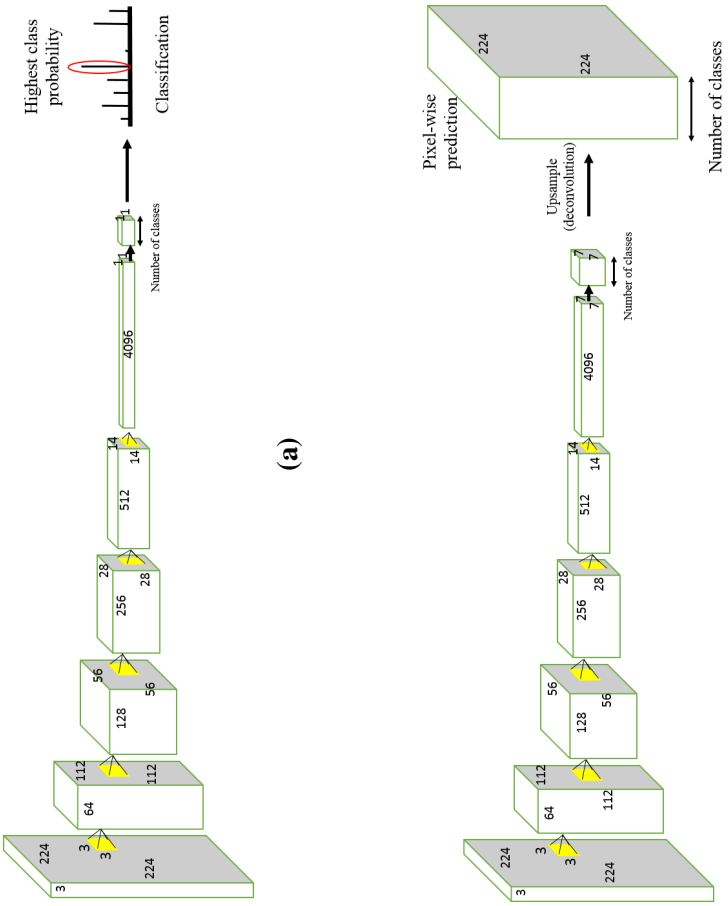


Figure 5.4: An example of the architecture of a) CNN and b) FCN. While CNN outputs the classification prediction, a FCN outputs the pixel-wise prediction or segmentation.

As shown in Figure 5.4, the basic architecture of the FCN is similar to that of the CNN, but by predicting what classes each pixel belongs to, it is more suitable for segmentation prediction. For the basic FCN architecture, VGG-16 nets [134] have been used. Segmentation requires local information, which tells *where*, as well as the global information which tells *what*. However, the classic CNN considers the global information in the image only and partially neglects spatial information due to the pooling. To overcome this, a FCN links the final prediction layer with the prediction from the previous layers to add local information to refine the output. For example, an edge feature is extracted in the lower layer and by linking this layer to the prediction layer, the boundary of the segmentation result can be finer. According to [136], the layer called the *skip* layer can be added which uses 1×1 convolution to match the depth with the prediction layer. The FCN architecture (FCN-8s) used in this work is shown in Figure 5.5.

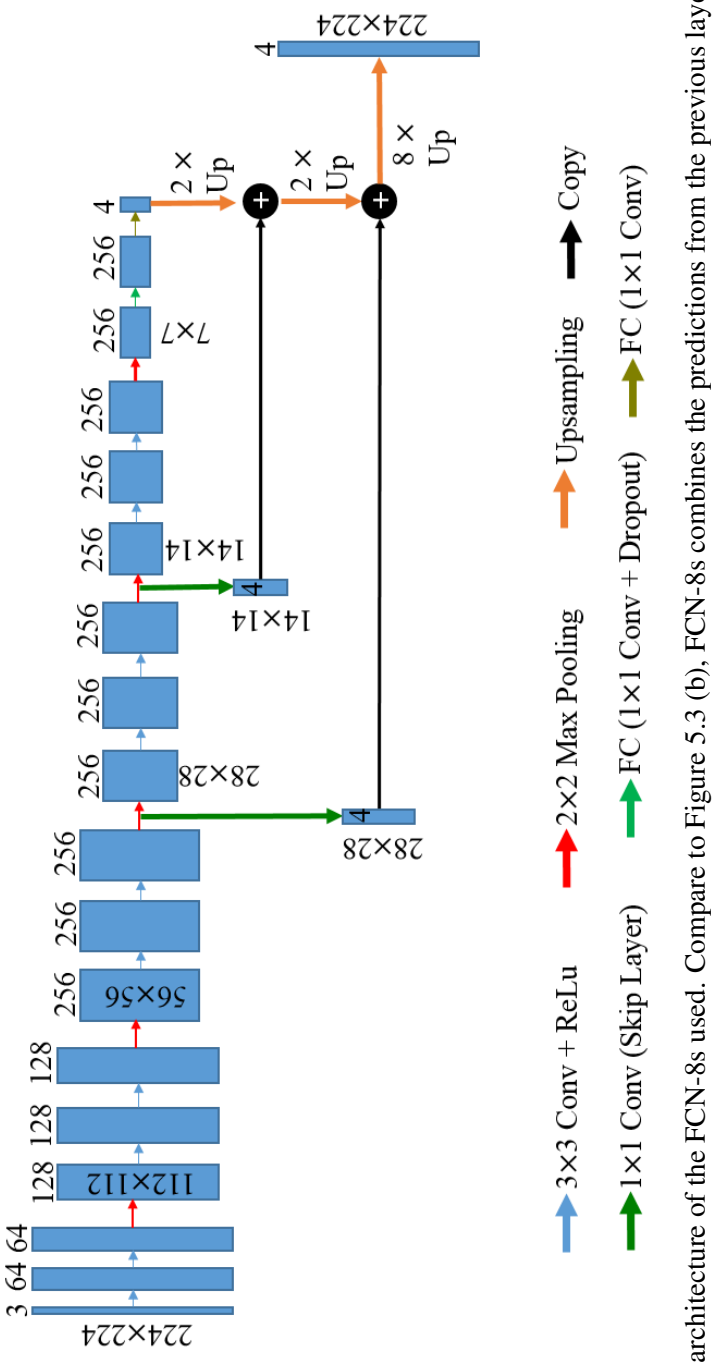


Figure 5.5: The architecture of the FCN-8s used. Compare to Figure 5.3 (b), FCN-8s combines the predictions from the previous layers using the skip layer.

As shown in Figure 5.5, upsampling from the output of the fully connected layers is employed to make it the same size as the output of the lower layers to do the summation. The upsampling is achieved by using simple bilinear interpolation to match the same size as the output of the lower layers.

The progress of the feed-forward and the backpropagation is the same as for a CNN as explained in Chapter 2. However, since the output is a pixel-wise prediction, the loss (softmax log-loss) is computed for each pixel by comparing with a manually annotated label. Once the loss for each pixel is computed, they are all added to get the loss value for the image. However, the number of pixels for each class is unbalanced. Due to the larger area of the non-fetal part compared to the fetal part, the number of pixels of background (non-fetal) is larger than that of the fetal part. According to [179], [180], it is possible that the minority class is likely to be predicted as noise or outliers which causes more misclassification. This is important because the limbs are much smaller compared to the torso and the non-fetal tissue. To reduce this imbalance, one way is to put more weights on the loss of the minor classes to have more influence on the loss than the major class has. In other words, if there is a misclassification on the minor class, it is more important to reduce this error rather than the mislcassication on the major class. The weight is multiplied with each pixel-loss, and the weight is calculated based on the number of pixels. Therefore, the weight for the fetal part is considered as 1 and the background is computed as follows,

$$\text{Weight for background} : \frac{\text{Number of Fetal pixels}}{\text{Number of Background pixels}} \tag{5.1}$$

By doing this, the fetal parts affect more on the loss than the background. Once the weight is computed, the rest of the process is the same as for the classic CNN.

From the previous chapters, one of the assumptions was that a good sagittal plane is provided as input. However, choosing a good sagittal plane can be achieved using the classification network to identify whether a slice satisfies the CRL criteria or not. Moreover, one further thing to note is that this classification network is using the same images that have been used for whole fetus segmentation in the sagittal view. Thus, the classification and segmentation networks have shared features. Therefore, these two

networks can be merged as one network with two outputs or tasks. This is called a *Multi-task Network* and the architecture is shown in Figure 5.6.

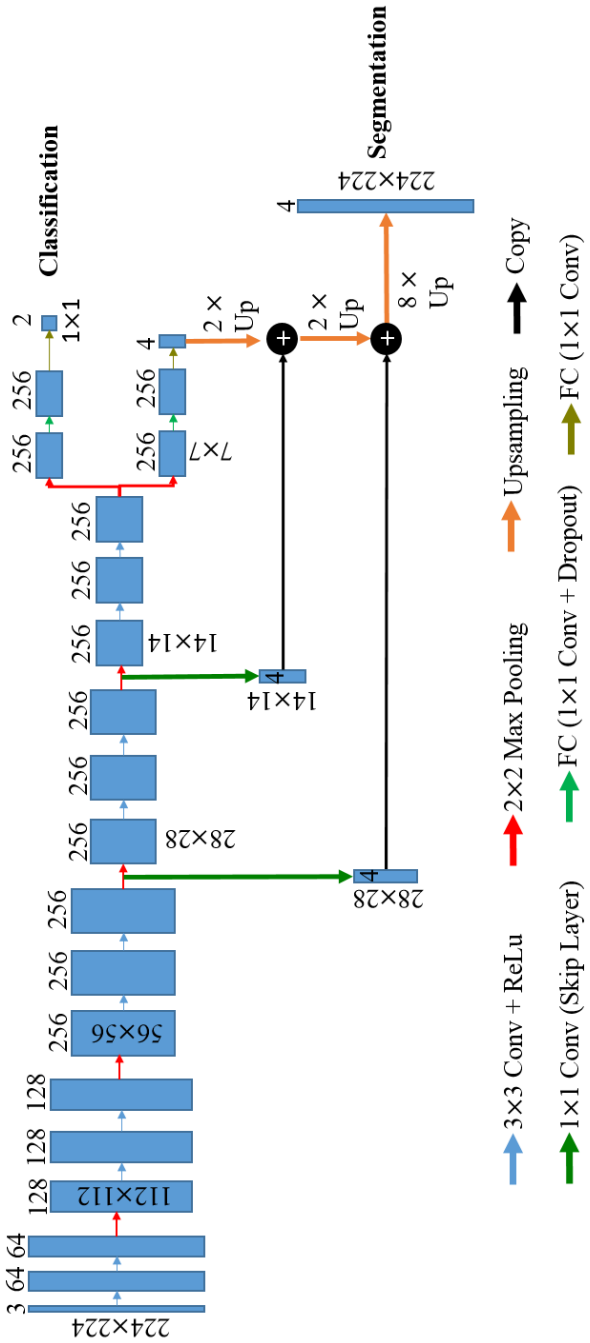


Figure 5.6: The architecture of multi-task FCN used for the classification and the segmentation of the whole fetus in the sagittal view. Compare with Figure 5.5, the FCN-8s is divided before the fully connected layers (FC) to perform both the classification and the segmentation.

For limb segmentation, a refined boundary is required due to its smaller size compared to the fetal parts. Therefore, a U-Net [140] has been used for this task. As explained in Chapter 2, a U-Net is similar to a FCN, but uses all lower layers and uses the concatenation instead of the summation while linking with the output of the higher and the lower layers. The architecture used for the limb segmentation in this work is summarized in Figure 5.7,

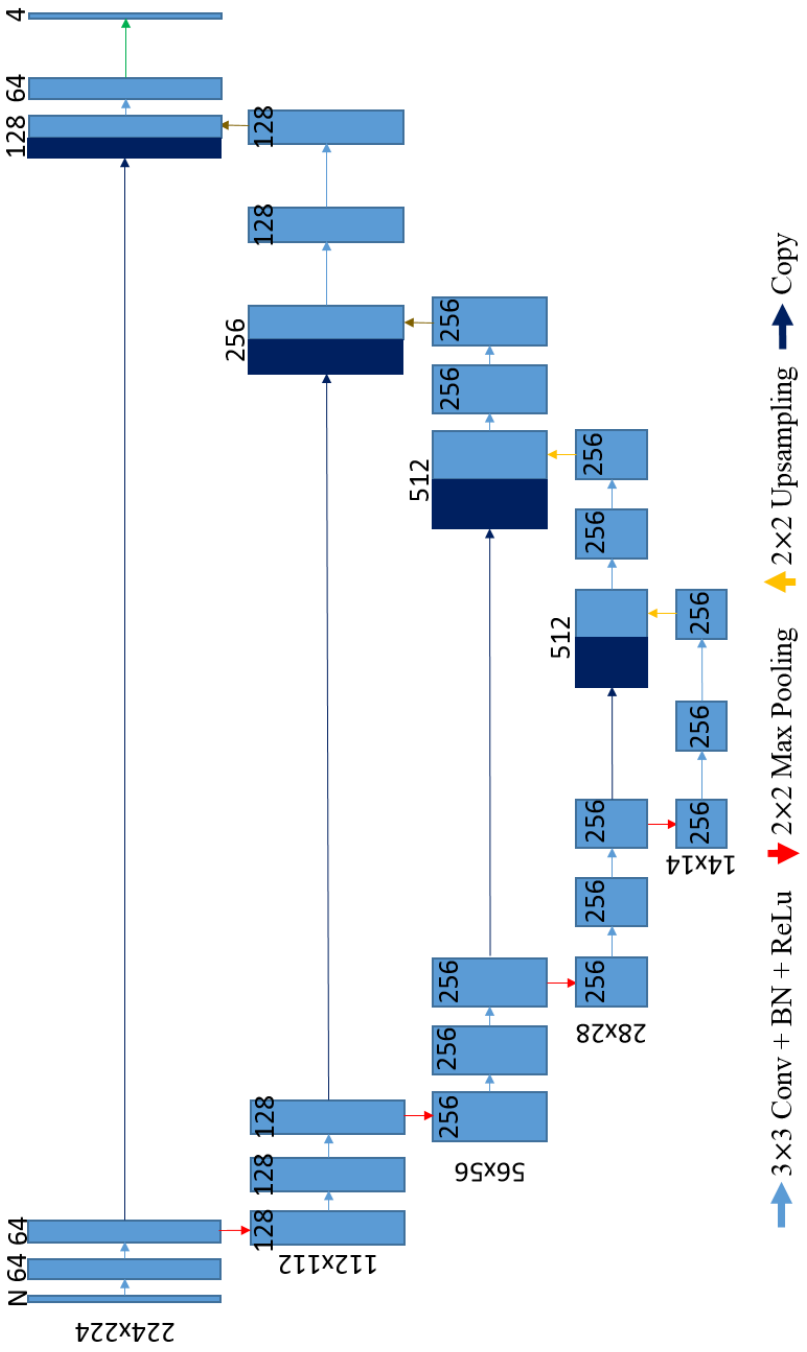


Figure 5.7: The architecture of U-Net for the limb segmentation. Compare to the FCN-8s shown in Figure 5.6, U-Net combines all the previous layers. The differences between FCN-8s and U-Net are that FCN-8s sums the predictions whereas a U-Net concatenates the features from the previous layers.

As shown in Figure 5.7, a U-Net combines the features from the higher layers with the features from the low layers by using skip layers. As mentioned in Section 4.2, the ReLU has been proposed to reduce the problem of vanishing gradients. However, the vanishing gradient problem was still remain with a deeper network. Also, as the network gets deeper, it has been found that small changes in the input distribution from the previous layer can affect the later layers significantly [181]. This phenomenon is called *Internal Covariate Shift* [181]. To prevent this, Ioffe et al. [181] proposed Batch Normalization (BN). BN normalizes the inputs of each layer to have a mean output activation of 0 and variance of 1. This is helpful because it can adjust the values which are too big or too small which will lead to Internal Covariate Shift. BN can also help to accelerate the learning rates due to the reduction of outlier activations which affect the learning [181]. This is the reason that the learning rate has to be small. However, since BN reduces the outlier activations, a higher learning rate can be used. Unlike the FCN based on the VGG-16, there are no pre-trained U-Nets available, so the parameters/weights are learned from scratch.
Also, the method called *Batch Normalization* [181] has been used within the U-net. Specifically, let

$$\begin{aligned} \mu_{BN} &= \frac{1}{m}\sum_{i=1}^{m} x_i, \\ \sigma_{BN}^2 &= \frac{1}{m}\sum_{i=1}^{m}(x_i-\mu_{BN})^2, \end{aligned} \tag{5.2}$$

Then,

$$\begin{aligned} \hat{x_i} &= \frac{x_i-\mu_{BN}}{\sqrt{\sigma_{BN}^2,+\varepsilon}}, \\ y_{BN} &= \gamma\hat{x_i}+\beta. \end{aligned} \tag{5.3}$$

where μ_{BN} is the mean of the mini-batch, m is the number of the mini-batch, x_i is the i^{th} input to the Batch Normalization layer, σ_{BN}^2 is the mini-batch variance, $\hat{x_i}$ is the i^{th} normalized input, ε is a small value to prevent dividing by 0, y_{BN} is the output of the Batch Normalization layer, and γ and β are the parameters to be learned during the training.

5.2.2 Standardization of plane orientation

Head and Abdomen

As mentioned in Section 5.1, plane orientation is an important factor to consider due to the varying fetal pose in acquisition. As Figure 5.3 shows, the image is more suitable for anatomical assessment after standardization of orientation. This is particularly important for biometry measurement. For example, the CRL plane requires identification of specific points on the crown and rump. However, if fetal orientation is not considered, these two points are hard to find. One of the ways to determine fetal orientation is to rotate each slice by several candidate orientations and to choose the best plane automatically by performing image plane classification. However, this is computationally expensive because it requires several changes in the orientation for every slice. Also and unlike many medical image analysis tasks, the range of orientations over which to sample is unknown. Therefore, in this work, segmentation of the whole fetus in the sagittal view has been used to find the orientation of the head and the torso as shown in Figure 5.8. Since the segmentation of the head and torso in the sagittal view has been performed, the orientation of each part can be approximated by a mid-line and the back of the whole fetus based on the segmented region. To obtain the mid-line, skeletonization is used. The skeletonization is done by using the Fast Marching Method [182] [1]. The mid-line is extracted in two steps. First, the whole fetus segmentation is skeletonized to obtain the mid-line of the whole fetus. Second, the whole fetus segmentation is partitioned into the head and torso, and skeletonized lines are fit to each partition using second order polynomial curve fitting. However, since these mid-lines are obtained from the skeletonized line based on the whole fetus, this mid-line might not represent the mid-line of each head and body. To obtain a more accurate mid-line, either the front or back boundary of the head and torso can be used. However, the front part of the fetus has too many "spurs". For example, the fetal head has the nose and mouth. Therefore, instead of using the front part of the fetus, a second order polynomial fitting is applied to the boundary of the back of the fetus to define the *back-curve*. Then from each segmented region of the head and torso, the slope of each mid-line curve and back-curve are averaged to estimate the final orientations

[1] The public code is provided by Howe, http://www.cs.smith.edu/ nhowe/research/code.

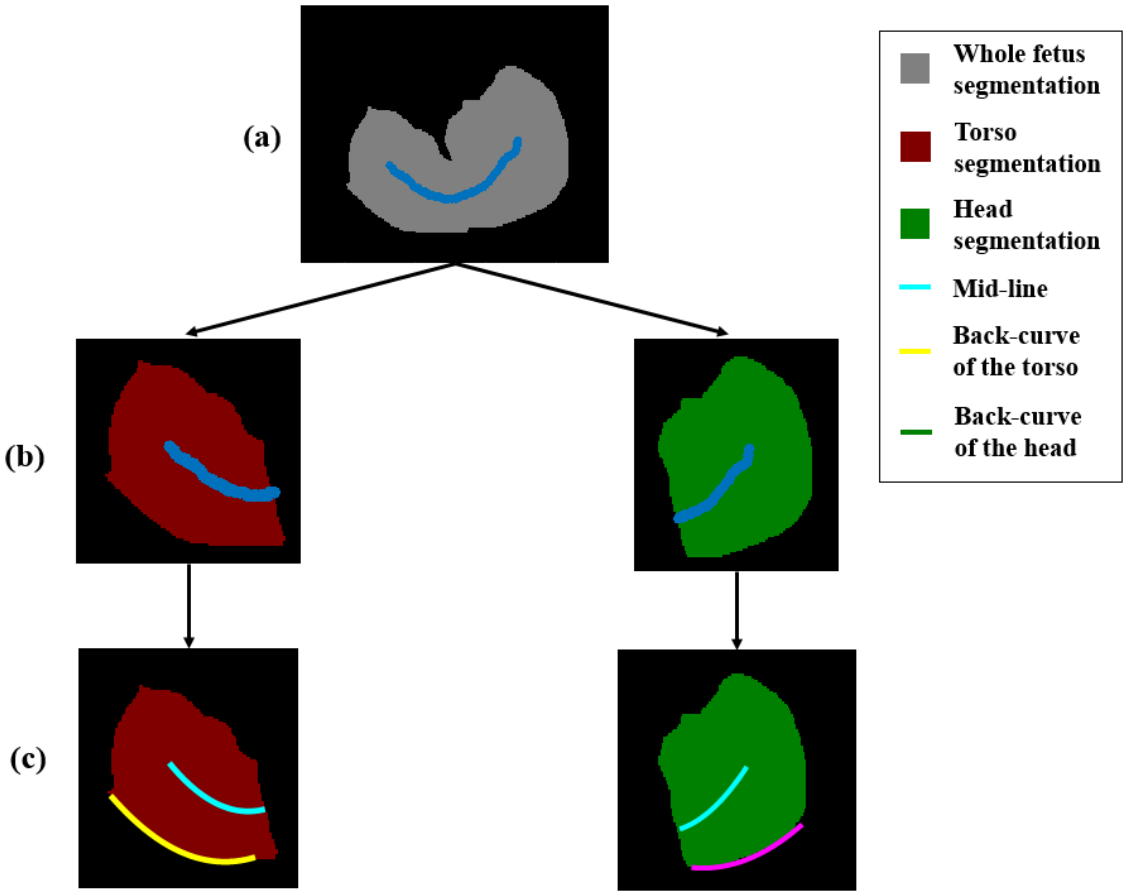


Figure 5.8: The process of getting the orientation for the head and the torso. (a) Based on the annotation from a 2D slice extracted from a 3D volume, the mid-line of the whole fetus is obtained using the skeletonization transform. (b) Separate the whole fetus into the head and the torso. (c) Use polynomial fitting to each skeletonization of the head and the torso to get the *mid-line curve*. Also, the boundary of the backside of each region, *back-curve* is obtained. Then calculate the average slope of mid-curve and back-curve to get the orientation of the head and the torso.

of the head and the torso. Unlike the abdomen, the orientation of the transthalamic plane/standardized head plane is slightly different from the head orientation, which the frontal face is facing, because HC is measured on the transthalamic plane/standardized head plane which is shown in Figure 5.9. Therefore, based on the training data, the differences between the head orientation and the transthalamic plane/standardized head plane orientation is averaged to extract a more accurate transthalamic plane/standardized head plane. The average difference between the two was found to be 17.9°.

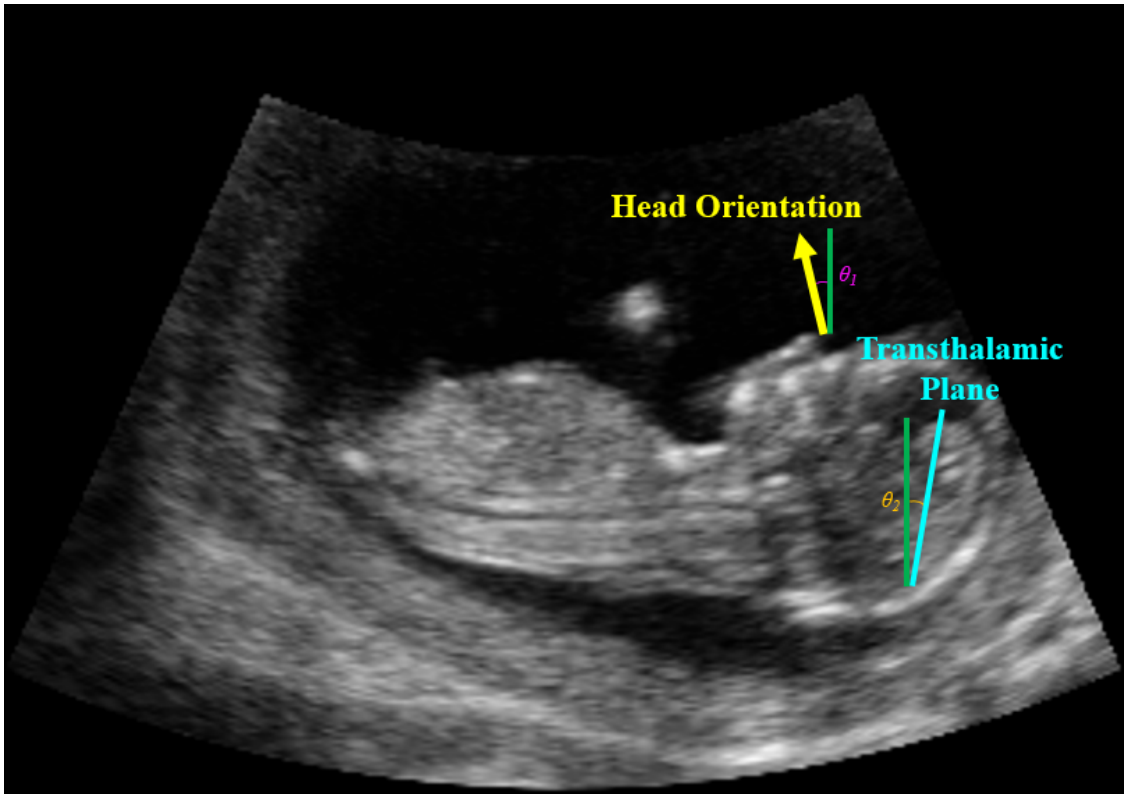


Figure 5.9: There is a difference in orientation between how the head is facing (θ_1) and the transthalamic plane (θ_2).

Limbs

The orientation of each of the left and right limbs (arms and legs) is based on the segmented results from the U-Net in the axial view.

However, since not all limbs can be viewed in a single slice, the segmented regions from each slice are stacked as shown in Figure 5.10 (a). Then by using the averaged class probability of each of the stacked regions, the region pixels that received more than 50% probability score are extracted as the final segmentation of the limbs (Figure 5.10 (b)). With this segmentation region, the mid-line is obtained using the skeletonization transform [182] (Figure 5.10 (c)). Finally, using polynomial curve fitting, the mid-line of the segmentation region and the orientation of left and right limbs are obtained as shown in Figure 5.10 (d).

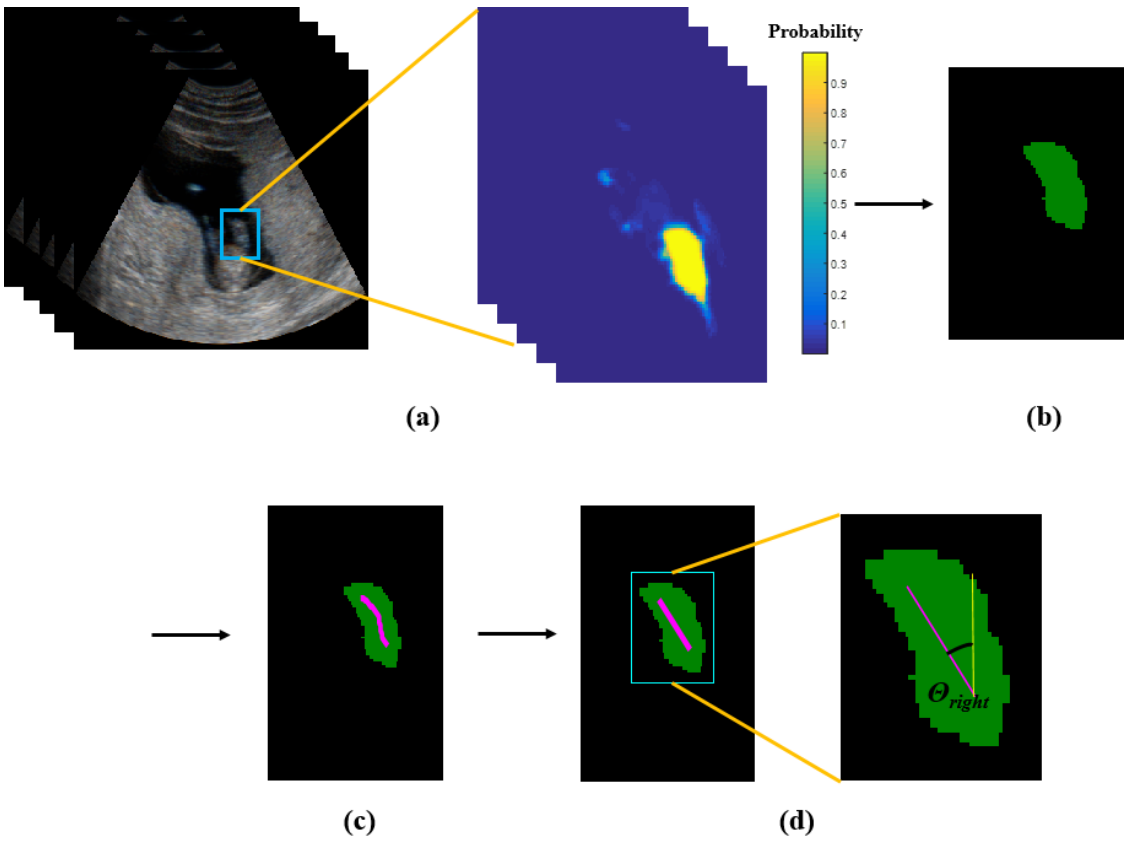


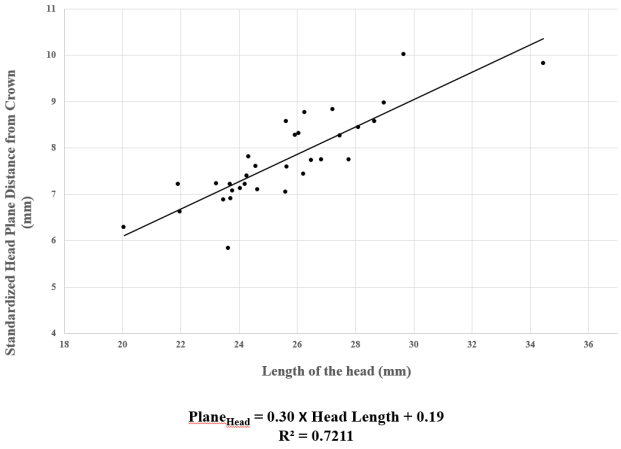
Figure 5.10: The process of determining the orientation for the right leg. (a) Based on the probability map for the segmentation of the right leg, the pixels with the averaged probability larger than 50% are extracted. (b) Based on the segmentation results, (c) use skeletonization transform [182] to get the mid-line of the right arm. (d) Linear polynomial fitting is then applied to obtain the orientation angle.

5.2.3 Estimation of Plane Location

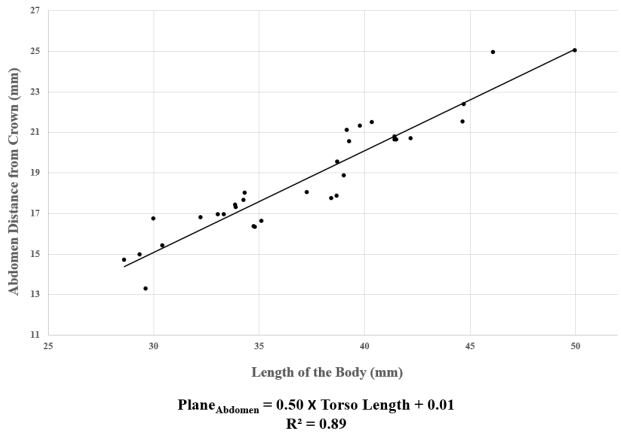
In Chapter 4, the plane for fetal biometry assessment was estimated by using the linear regression based on the head and torso length without considering the orientations of the head and the abdomen. However, as mentioned in Section 4.4, we found this had a low R^2 value, since the orientation is not considered in the estimation. However, in this chapter, the orientation of the head and the torso are considered. Using the method in Section 4.4.2, linear regression is needed to find the best plane for the head and abdomen as shown in Figure 5.11,

The location of the best planes for the head and the abdomen after the correction of the orientation are,

$$Plane_{Head} = 0.30 \times Head\ Length + 0.19 \tag{5.4}$$



(a)



(b)

Figure 5.11: The linear regression graph with the equations and R^2 value to approach (a) Head and (b) Abdomen.

$$Plane_{Abdomen} = 0.50 \times Torso\ Length + 0.01 \tag{5.5}$$

with R^2 values 0.72 and 0.89, respectively. Since we know both the location and the orientation of the head and the abdomen, we can then rotate the volume according to the orientation and extract the 2D slice according to the location found using the equation 5.4 and 5.5.

Plane Extraction

The 2D slices were extracted from a 3D volume using the existing Matlab code [2]. We first create the slice surface in the domain of the volume using the estimated location,

[2]The code is available in https://uk.mathworks.com/matlabcentral/fileexchange/32032-extract-slice-from-volume

and then orient this surface with respect to the axes using the rotation matrix, R as shown in equation 5.6,

$$R = \begin{bmatrix} cos\theta + u_x^2(1-cos\theta) & u_x u_y(1-cos\theta) - u_z sin\theta & u_x u_z(1-cos\theta) - u_y sin\theta \\ u_y u_x(1-cos\theta) + u_z sin\theta & cos\theta + u_y^2(1-cos\theta) & u_y u_z(1-cos\theta) - u_x sin\theta \\ u_z u_x(1-cos\theta) - u_y sin\theta & u_z u_y(1-cos\theta) - u_x sin\theta & cos\theta + u_z^2(1-cos\theta) \end{bmatrix} \tag{5.6}$$

where θ is angle/orientation, $u = (u_x, u_y, u_z)$ is unit vector where $u_x^2 + u_y^2 + u_z^2 = 1$. After the rotation, get the coordinates of each pixel on the surface to get the intensity values using the nearest neighbour interpolation to draw the slice plane within the volume.

5.2.4 Automatic Measurement

In the subsection we explain how biometry measurement is performed.

Crown Rump Length (CRL)

The CRL is the length between two points placed at the crown and the rump respectively. The point at the farthest point of each oriented segmented part is chosen. As shown in Figure 5.12, the rump is the farthest point of the segmented region of the lower part of the torso, which in this case is the region near at the back of the fetus. It is possible that the fetus can stretch out and causes the CRL length to be longer. However, in this thesis, the data has been taken based on the CRL criteria as mentioned in Appendix A which includes the fetus must be in a neutral position. Therefore, all of the fetal data used in this thesis does not contain the fetus that is hyperextended or flexed.

The *Head Circumference* (HC) can be calculated from the *Biparietal Diamter* (BPD), the shortest radius, and *Occipitofrontal Diameter* (OFD), the longest radius, using equation 5.5,

$$HC = \frac{\pi \times (BPD + OFD)}{2} \tag{5.7}$$

The *Abdomen Circumference* (AC) measurement is calculated from the *Antero-posterior Abdominal Diameter* (APAD), the shortest radius, and *Transverse Abdominal Diameter* (TAD), the longest radius, using equation 5.6,

$$AC = \frac{\pi \times (APAD + TAD)}{2} \tag{5.8}$$

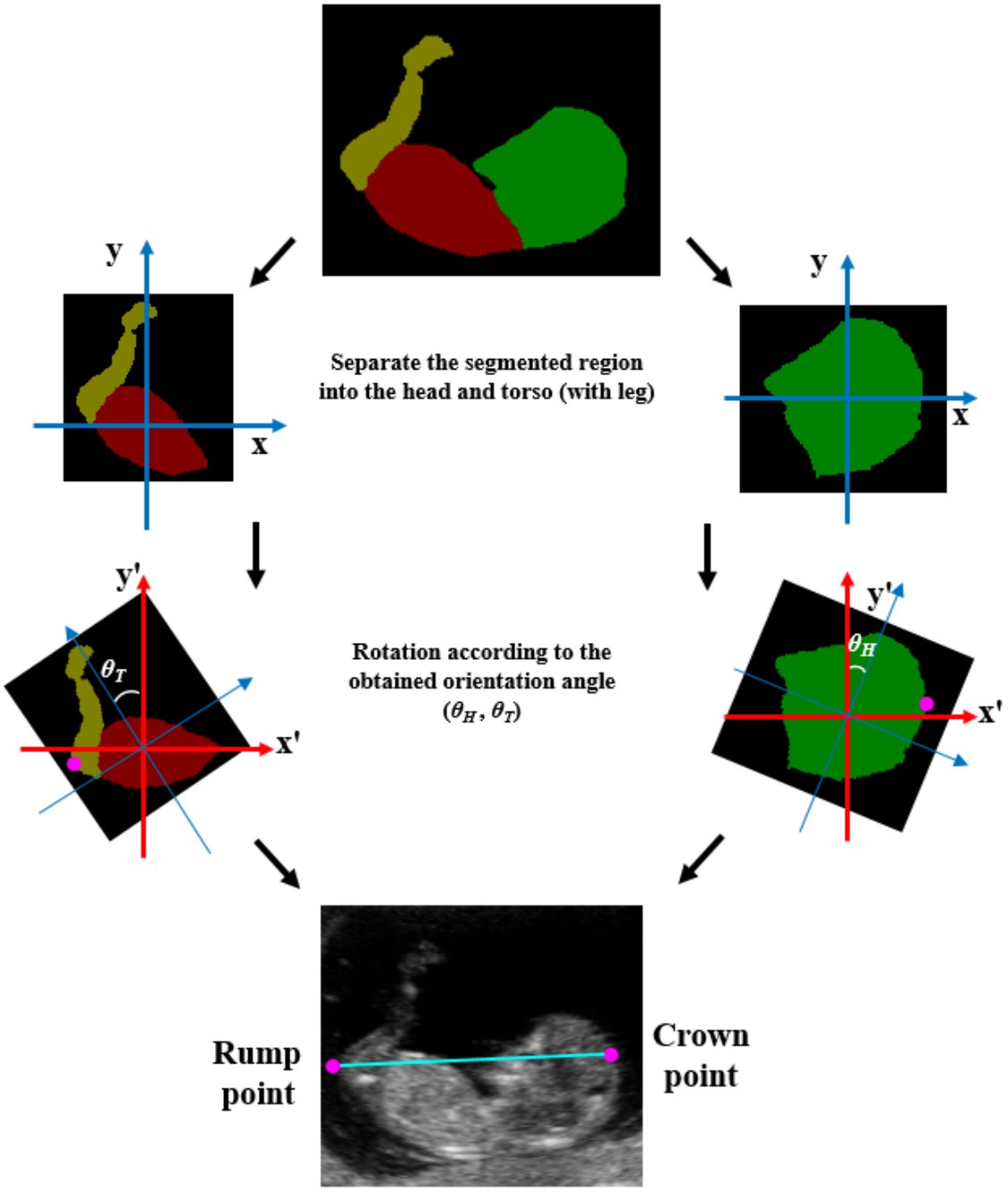
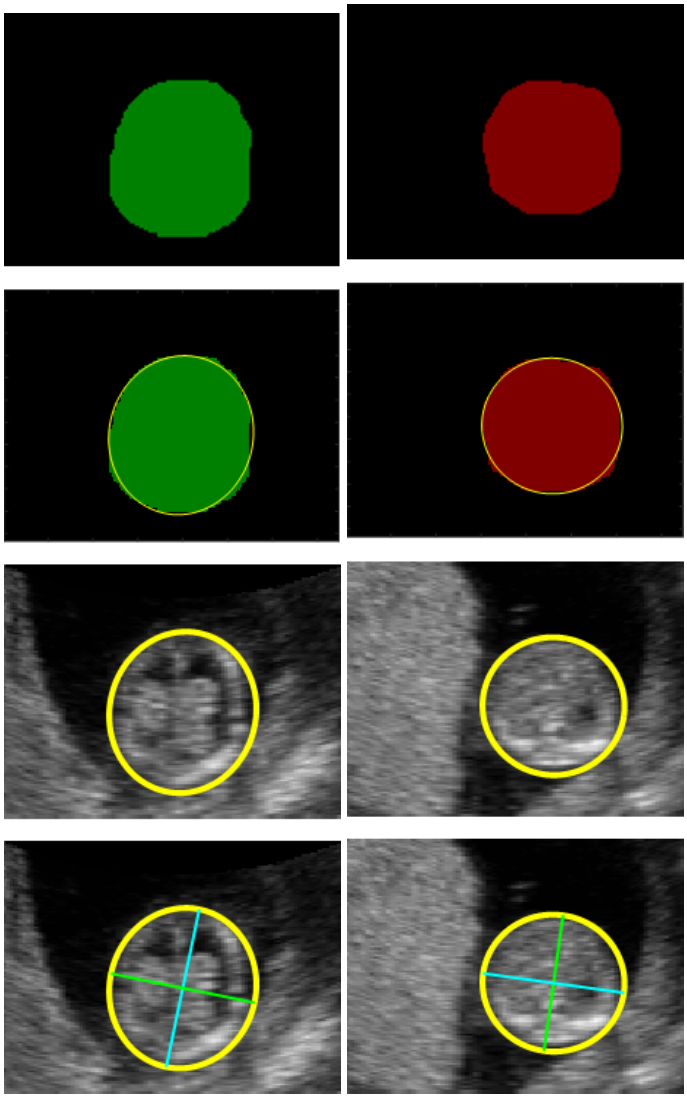


Figure 5.12: The overall process of getting the crown and rump points for the CRL measurement. Based on the segmented region, the head and torso (including the lower limb) are separated first. Then rotate each region according to the obtained orientation angles (θ_H,θ_T) from the centre point of the head and torso. Based on the rotated region, obtain the crown and rump coordinate by finding out the farthest point inside the segmented region. Finally, link the two points and do the CRL measurement.



(a) **(b)**

Figure 5.13: The examples of the automatic measurement of (a) the HC and (b) the AC. When the segmentation result is obtained, fit the ellipse to the segmented region. The short and long axis of the resulted ellipse is then obtained to calculate the circumferences.

For the HC and AC, an ellipse is fitted to each segmented region/given-points using the least-squares fitting method following [183] by having the minimum sum of the squares of the distances to the given points. After the ellipse is obtained, the longest and shortest outer-outer diameter are estimated as shown in the Figure 5.13.

GA is calculated from the CRL using a standard formula [93],

$$GA(Days) = 40.9041 + 3.21585 \times CRL^{0.5} + 0.348956 \times CRL \tag{5.9}$$

5.3 Experimental Setup

The data was separated into training and test sets. The volume and the images from the same fetus were placed in either the training or the test set, not in both. In total 65 3D fetal volumes were used; 44 as the training set and 21 as the test set. For the training set, an additional 8 data were included for the head because some volumes had a good head axial views, although they did not satisfy the CRL criteria. Also, to compare the localization result with the result from the previous chapter, the same data set was used, and only the head and torso were segmented. Image preparation for input to the network is explained in Section 5.3.1 and 5.3.2.

5.3.1 Sagittal Plane

Classification

For plane classification, sagittal slices that satisfied the CRL criteria were manually classified as *satisfied* and other slices as *unsatisfied*, as illustrated in Figure 5.14.

Segmentation

For whole fetus segmentation in the sagittal view, the image was manually annotated into 4 classes: head, torso, lower limb and background, as shown in the Figure 5.15. Note that the annotation has been done not only for the images that contain all the fetal parts, but also for an image if it contains a clear visualization of one of the fetal classes. This is because such an image still contains important fetal features but if they are annotated as background, it may confuse the CNN in training.

5.3.2 Axial Plane

Head and Abdomen Axial Plane

The head and abdomen slices were first cropped based on the segmentation of the whole fetus in the sagittal plane as described in Chapter 4. For training, the image was first cropped manually based on the manual segmentation of the whole fetus. Then each head and each abdomen was annotated manually as shown in Figure 5.16.

(a) (b)

Figure 5.14: Typical example of a)*satisfied* and b) *unsatisfied* sagittal planes. They were labelled based on whether the slice satisfied the CRL criteria or not.

Limbs

Compared to the head and the abdomen, the arms and legs have greater appearance variance due to variable pose. Therefore, instead of cropping the image as for the head and abdomen, the original and full size image is used. Both limbs and the torso are manually annotated as shown in Figure 5.17. Since the presence of limb parts (upper, lower limbs and hands or feet) are difficult to visualize in a single axial slice, we learn the spatial relationship between consecutive slices. Therefore, instead of having a single slice as input, 3 slices are input with the focus on the middle slice being evaluated.

Data Augmentation

Similar to Section 4.3.1, the data augmentation was performed by flipping left-right and the image rotation around the centre to a maximum $\pm$ 30 $^\circ$ randomly with uniform distribution. With these augmentation methods, the training data size is increased to 5 times larger the original set. For the limbs, however, only the image rotation is considered

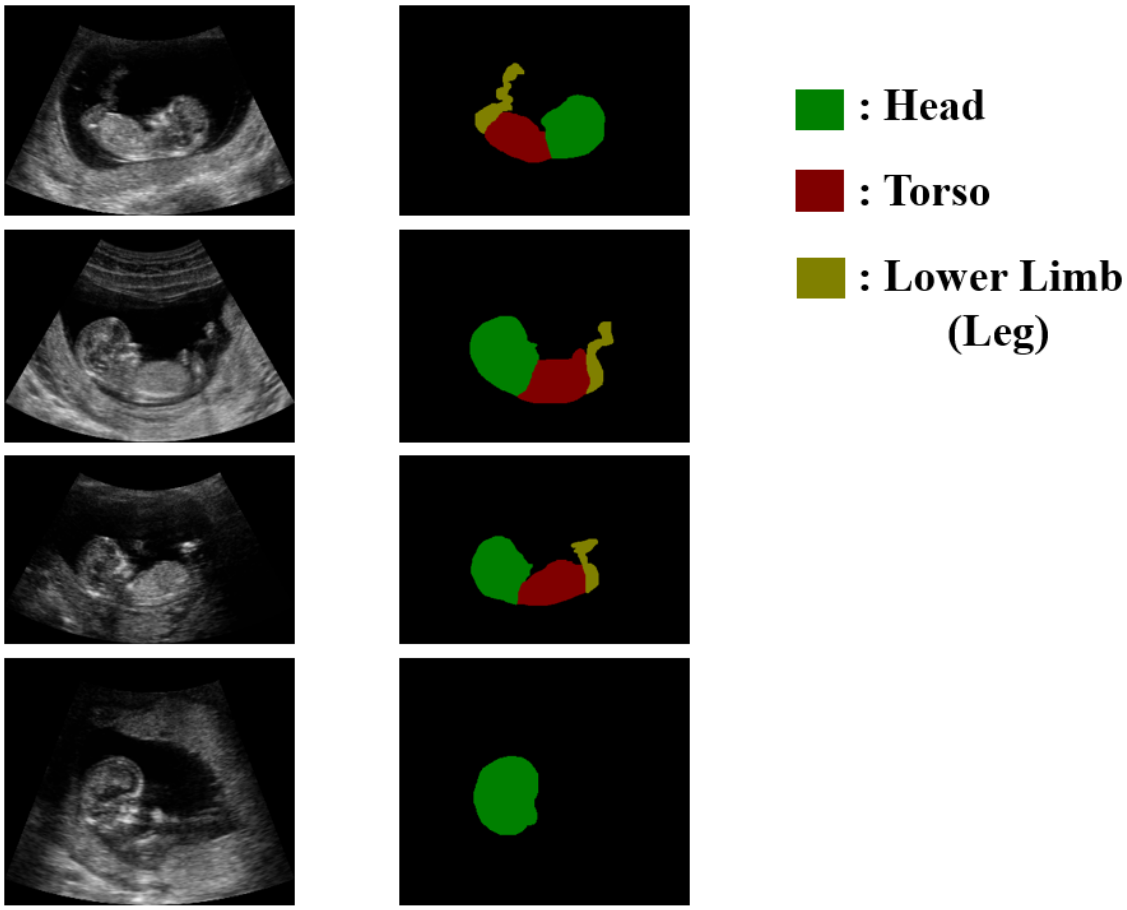


Figure 5.15: An example of a) raw image and b) its corresponding manually-derived labels. Note that if the image contains any fetal parts with good visualization, it has been annotated even though the slice itself may not satisfy the CRL criteria.

since the left and right are considered as different classes for limb segmentation, which yielded 4 times lager training data size for the limbs. As pre-processing, the intensity normalization of the training data (with mean and standard deviation) is performed.

5.4 Evaluation

5.4.1 Segmentation (Quantitative Measures)

Two comparative metrics were used.

Intersection over Union of bounding box (IoU_{box}) is defined as,

$$IoU_{box} = \frac{\text{Area of Automatic Box} \cap \text{Area of GT Box}}{\text{Area of Automatic Box} \cup \text{Area of GT Box}} \tag{5.10}$$

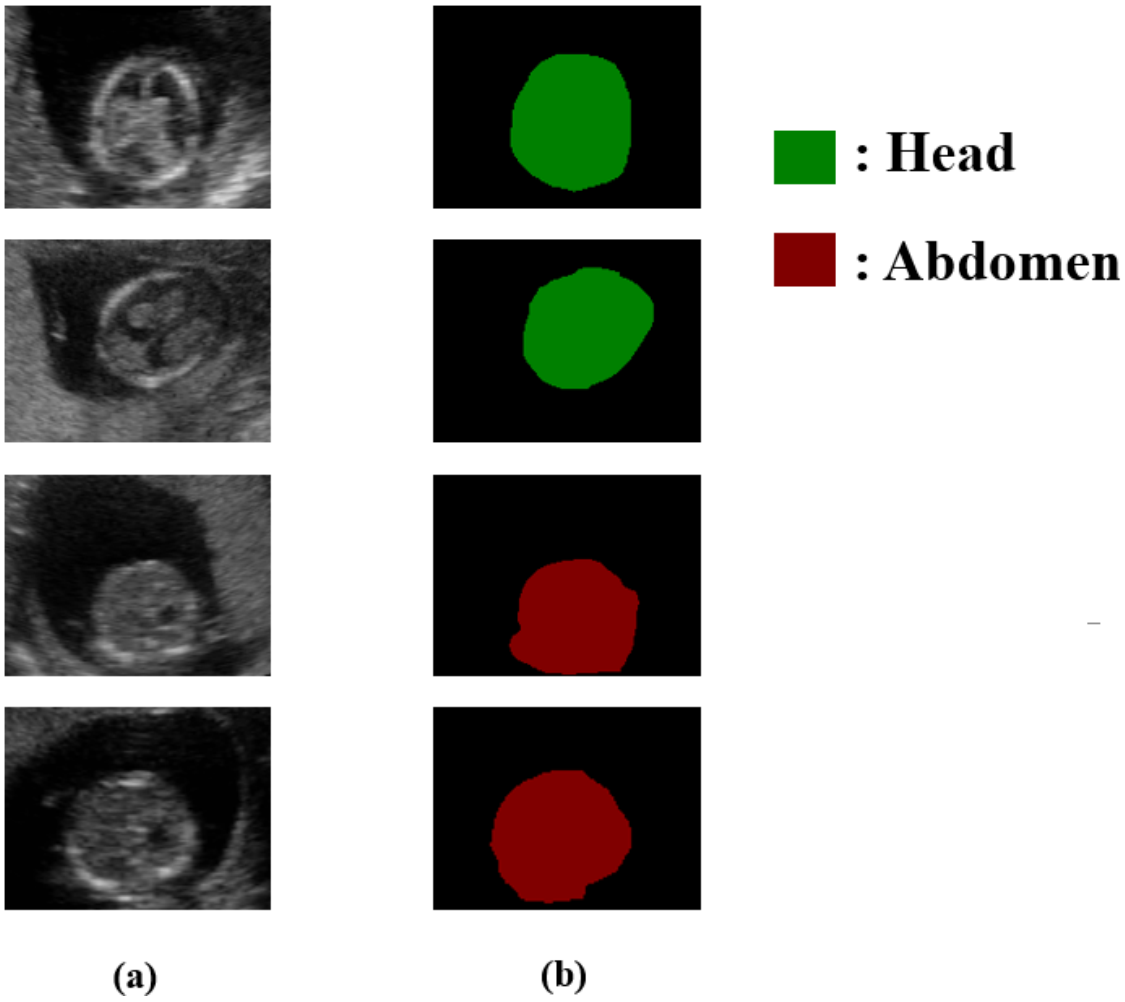


Figure 5.16: Typical examples of head and abdomen annotation a) Raw image, and b) corresponding manually annotated labels.

where GT is groundtruth.

Pixel accuracy and *Intersection over Union of segmentation* IoU_{seg} were used as the performance metric for the FCN. Pixel accuracy defines how many pixels are correctly classified compared to the ground truth. IoU_{seg} is defined as,

$$IoU_{seg} = \frac{\text{Number of Pixels of Auto} \cap \text{Number of Pixels of GT}}{\text{Number of Pixels of Auto} \cup \text{Number of Pixels of GT}} \tag{5.11}$$

IoU is very similar to Dice coefficient which represents how similar two segmentation masks are, which in this case, automatic and manual annotation. IoU has been used as evaluation metric for number of papers for bounding box and segmentation [165], [136], [140], [184], [40], [185].

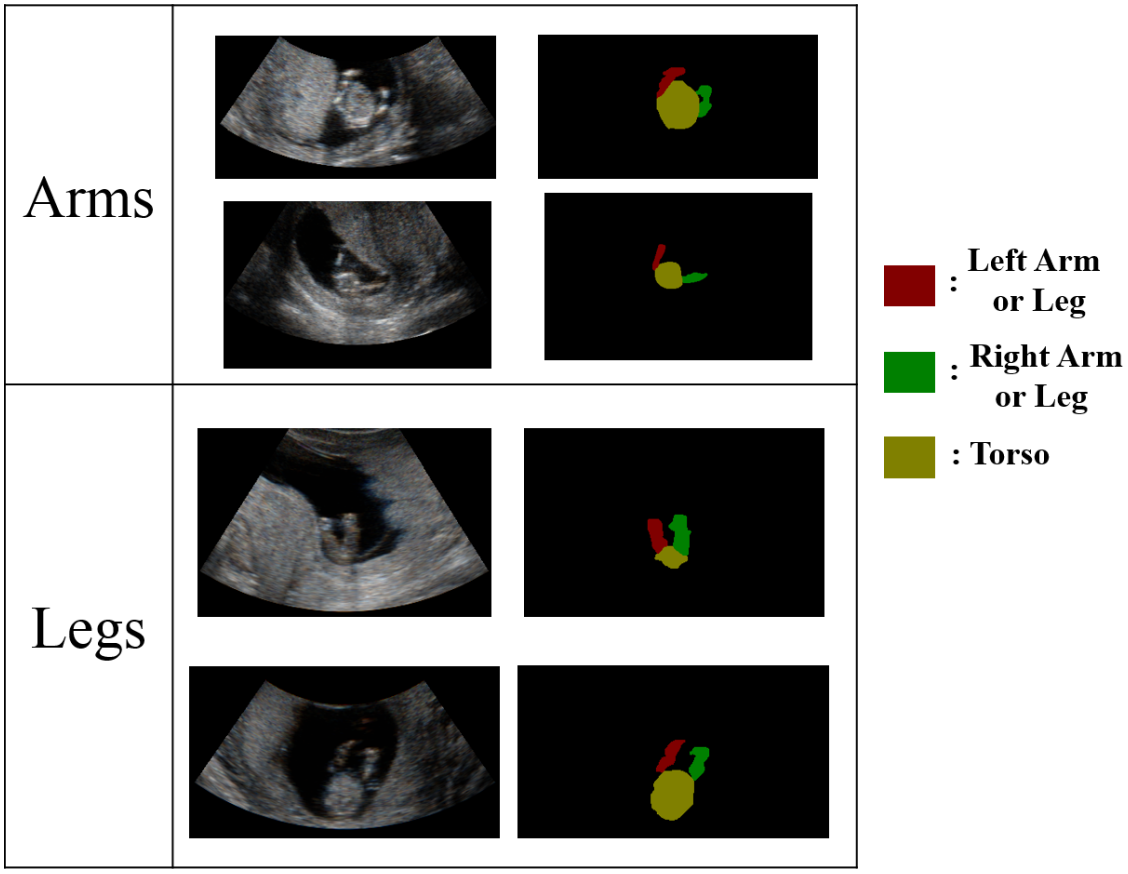


Figure 5.17: The example of axial slice of each left and right arms and legs, and torso with corresponding labels.

5.4.2 Visualization (Qualitative analysis)

As a qualitative evaluation, two clinicians were given the manually and automatically extracted plane for each fetal part from each data without any information about which images had been automatically or manually extracted. Then they were asked to decide which images might be suitable for anatomical assessment based on the ISUOG guidelines (Table 2.1) [5].

5.4.3 Extracted Plane and Orientation (Quantification)

To evaluate accuracy of estimation of the plane and the orientation, two metrics were used; (1) the distance between the manually and automatically extracted plane [186]; and (2) the angle difference between the manually and automatically chosen angle. For the evaluation for the limbs the detection rate and the plane orientation are used to

Total Pixel Acc (%)	BG Acc (%)	Torso Acc (%)	Head Acc (%)
96.1	**97.2**	**88.1**	**90.8**

Table 5.1: Total pixel accuracy and the accuracy for each class (head, torso and background (BG)), where Acc represents accuracy.

Mean IoU_{seg}	BG IoU_{seg}	Torso IoU_{seg}	Head IoU_{seg}
0.83 $\pm$ 0.12	**0.96**	**0.73**	**0.79**

Table 5.2: Mean IoU_{seg} and the standard deviation, and the IoU_{seg} for each classes.

evaluate the fetal limb assessment.

5.5 Results

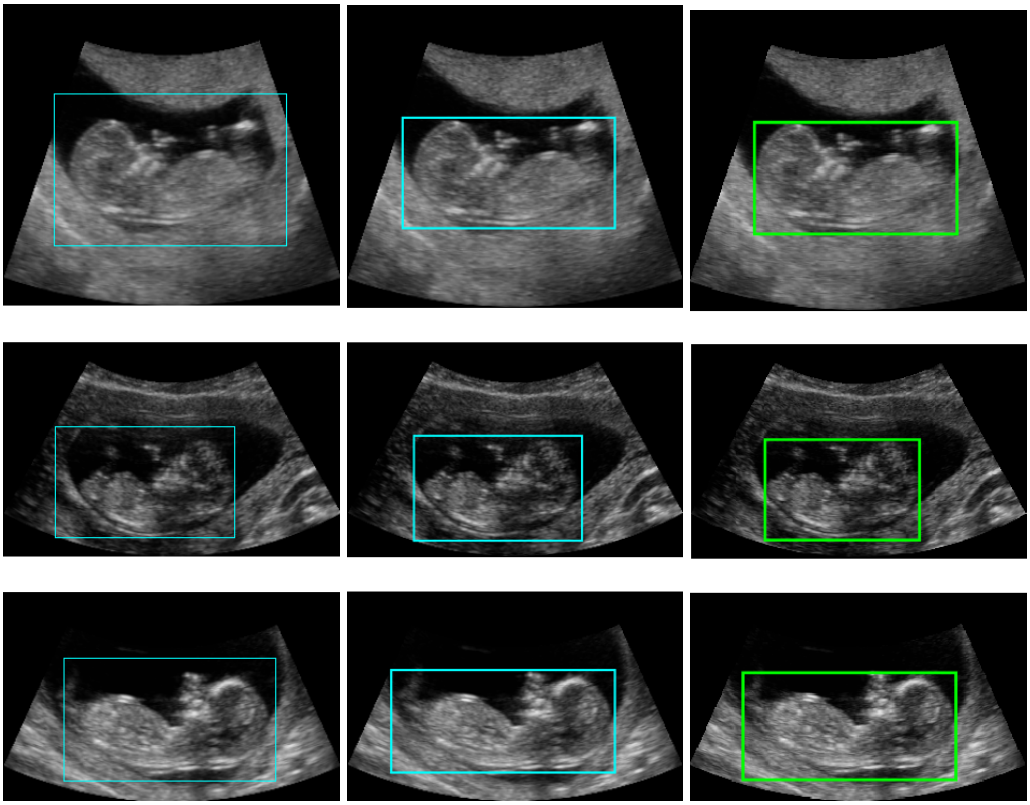
5.5.1 FCN and Random Forests

The performance of segmentation is shown in Table 5.1 and 5.2. Due to the large number of background pixels, the pixel accuracy and the IoU of the background class is much higher than for the fetal parts. The torso achieved the lowest accuracy and IoU. For comparison, IoU_{box}, results from Chapter 3 and FCN are 0.74 and 0.84, respectively as shown in Table 5.3. A few examples of the localization of the whole fetus are shown in Figure 5.18.

From Table 5.3 and Figure 5.18, we observe that the FCN achieved higher IoU than the methods following Chapter 3. As mentioned in Section 5.1, advantages of the FCN compared to the methods from previous chapters include that it can achieve a higher accuracy without the requirement of the handcrafted features. As shown in Figure 5.18, the FCN-derived bounding box surrounded the whole fetus more tightly than the method proposed in Chapter 3.

Method	IoU_box
RF and CNN	**0.74**
FCN	**0.84**

Table 5.3: The IoU_{box} from Chapter 3 and FCN that IoU_{box} from the FCN showed higher IoU than Chapter 3.



(a) (b) (c)

Figure 5.18: The examples of bounding box around the whole fetus in sagittal view (a) Chapter 3 (b) FCN (c) Groundtruth.

5.5.2 FCN for Fetal parts

Whole Fetus Segmentation

In this subsection, whole fetus segmentation has been performed for 4 classes (head, torso, lower limb and background). We compare three methods: FCN-32s with the original number of nodes, FCN-8s with reduced nodes, and a multi-task FCN-8s. Results are given in Table 5.4 and 5.5.

	Mean Accuracy (%) and Standard deviation	Torso (%)	Head (%)	Lower Limb (%)
FCN-32s (large)	**85.8±14.8**	**91.7**	**90.8**	**65.6**
FCN-8s (small)	**88.6±11.8**	**91.7**	**92.5**	**71.6**
Multitask	**89.4±11.7**	**91.9**	**94.2**	**72.9**

Table 5.4: Performance comparison of different FCN architectures based on the pixel accuracy.

	Mean IoU	Torso	Head	Lower Limb
FCN-32s (4096)	**0.58±0.26**	**0.59**	**0.62**	**0.15**
FCN-8s (256)	**0.74±0.22**	**0.79**	**0.81**	**0.41**
Multitask	**0.76±0.22**	**0.80**	**0.82**	**0.43**

Table 5.5: The performance comparison of different FCN architectures based on IoU.

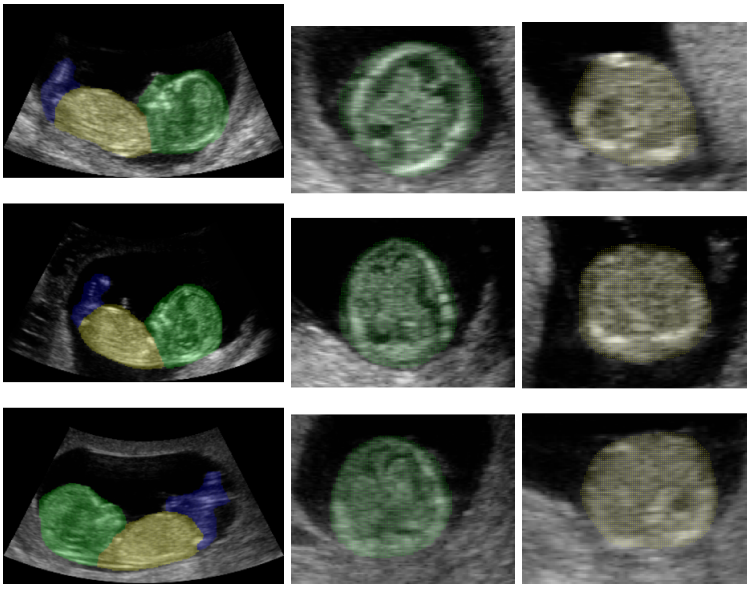


Figure 5.19: Examples of automatic segmentation of a) the whole fetus, b) the head and c) the abdomen. Green (Head), yellow (Torso), blue (limbs)

The first thing to note from the results shown in Table 5.4 and 5.5 is that the FCN-8s showed better results than the FCN-32s, although the FCN-8s used less nodes. The prediction from the FCN-32s is coarse because it was upsampled from the prediction with small size. Because of this, the boundary of the prediction appears coarse after upsampling. The FCN-8s architecture upsamples the prediction only eight times which yields a finer boundary. Because of this, both pixel accuracy and IoU are higher for the FCN-8s architecture. Another interesting finding to note from Table 5.4 and 5.5 is that the multi-task network achieved slightly better accuracy and IoU than FCN-8s. This indicates that the multi-task network can do both tasks at the same time with better performance compared to a single-task network. The multi-task network also achieved 98.9% classification accuracy. Examples of whole fetus segmentation are shown in Figure 5.19 (a).

Head and Abdomen

Two separate FCNs could be used for head and abdomen segmentation. However, it is also possible to have one FCN to simultaneously perform both segmentation tasks. To

	Head pixel acc (%)	Head IoU	Abdomen pixel acc (%)	Abdomen IoU
FCN (Head)	**95.4**	**0.90**	**-**	**-**
FCN (Abdomen)	**-**	**-**	**96.6**	**0.94**
FCN (Both)	**95.0**	**0.92**	**96.3**	**0.94**

Table 5.6: Segmentation results of the head and abdomen : comparing the performance between different FCNs.

	Mean Accuracy (%)	Left (%)	Right (%)	Torso (%)
FCN-8s	**80.1±15.4**	**67.9**	**66.9**	**87.5**
U-Net (single)	**84.7±13.6**	**71.1**	**75.2**	**93.5**
U-Net (Stacked)	**85.0±12.6**	**72.3**	**76.5**	**92.1**

Table 5.7: Performance of arm segmentation assessed by pixel accuracy: comparison of FCN-8s, a standard U-Net architecture with single channel and a same U-Net architectures with multiple channels.

evaluate the performance of a FCN for head and abdomen segmentation, three FCNs were implemented and compared. Results are summarized shown in Table 5.6.

As shown in Table 5.6, the segmentation results show negligible difference between the FCN for two individual segmentations and each FCN for either head or abdomen segmentation. This might be due to their different internal structures' image patterns, and the FCN can recognize these differences. Examples of head and abdomen segmentation are shown in Figure 5.19 (b) and (c).

Limbs

The performance of limb segmentation was compared for a FCN and a standard U-Net architecture and a U-Net architecture where the input with single channel and three channels that corresponding to relative slice and adjacent slices. Results are summarized for the arm in Table 5.7 and Table 5.8, and for the legs in Table 5.9 and Table 5.10.

	Mean IoU	Left	Right	Torso
FCN-8s	**0.57±0.20**	**0.32**	**0.32**	**0.67**
U-Net (single)	**0.70±0.18**	**0.51**	**0.50**	**0.82**
U-Net (Stacked)	**0.70±0.17**	**0.52**	**0.50**	**0.81**

Table 5.8: The performance of arm segmentation and the comparison based on mean IoU with standard deviations, and IoU for left, right arms and torso.

	Mean Accuracy (%)	Left (%)	Right (%)	Torso (%)
FCN-8s	**83.4±12.0**	**71.4**	**76.9**	**86.7**
U-Net (single)	**88.5±8.7**	**80.5**	**82.1**	**92.6**
U-Net (Stacked)	**88.3±8.0**	**81.8**	**82.6**	**90.0**

Table 5.9: Performance of leg segmentation assessed by pixel accuracy: comparison of FCN-8s, a standard U-Net architecture with single channel and a same U-Net architectures with multiple channels.

	Mean IoU	Left	Right	Torso
FCN-8s	**0.60±0.26**	**0.43**	**0.42**	**0.56**
U-Net (single)	**0.62±0.25**	**0.46**	**0.46**	**0.59**
U-Net (Stacked)	**0.63±0.25**	**0.45**	**0.46**	**0.63**

Table 5.10: The performance of leg segmentation and the comparison based on mean IoU with standard deviations, and IoU for left, right legs and torso.

As shown in Table 5.7-5.10, the U-Net gives better segmentation results compared to the FCN-8s. Due to the small size of a limb compared to both background and torso, the limb segmentation requires finer boundaries than the torso does. The U-Net has shown that using information from all the lower layers can improve the segmentation performance. Also, a 3-slice input was found to improve limb segmentation for the same architecture. Examples of the localization of the limbs and their corresponding sagittal view with and without correcting for the orientation are shown in Figure 5.20 and 5.21,

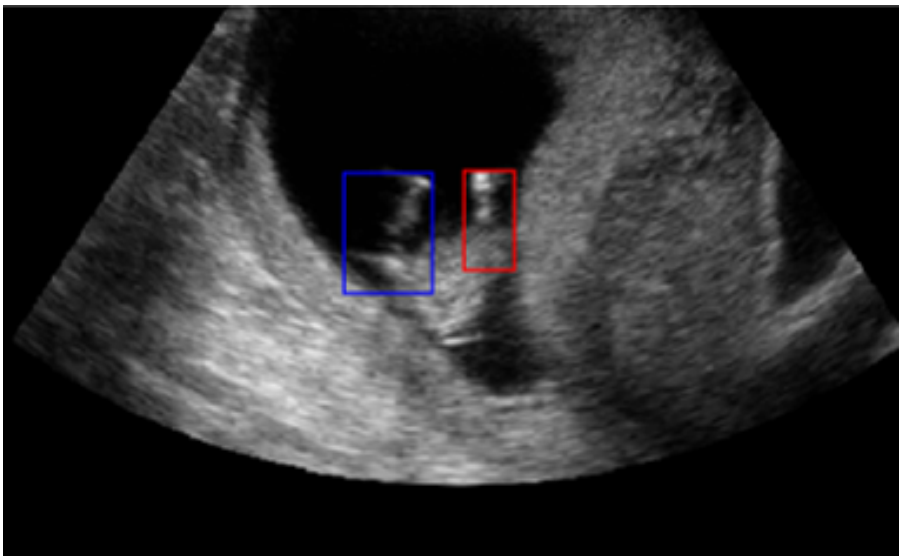
As shown in Figure 5.20 and 5.21, limb visualization after correcting for orientation is more focused on the limbs. In Figure 5.20, for example, the visualization of the left arm shows the upper and lower arm, and the hand.

5.5.3 Plane Extraction and Measurement

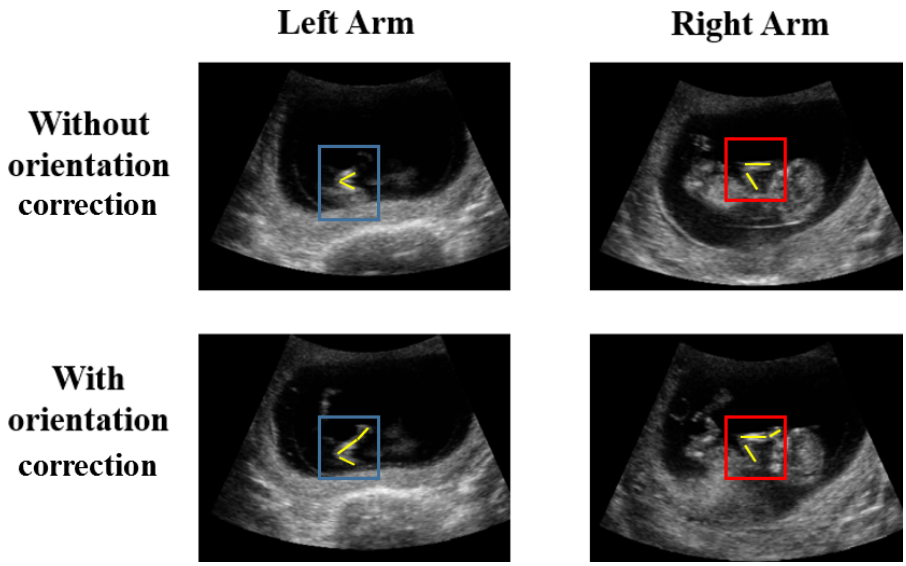
Head and abdomen plane extraction is quantified by computing (1) the averaged difference in the orientation and (2) the distance between the center point of the automatically and manually extracted planes. The results are shown in Table 5.11 and Table 5.12.

	Head	Abdomen
The averaged distance (mm)	**1.98 ± 1.19**	**2.51 ± 1.58**

Table 5.11: The averaged distance and the standard deviation between automatically and manually extracted planes for the head and the abdomen from 21 test data.



(a)

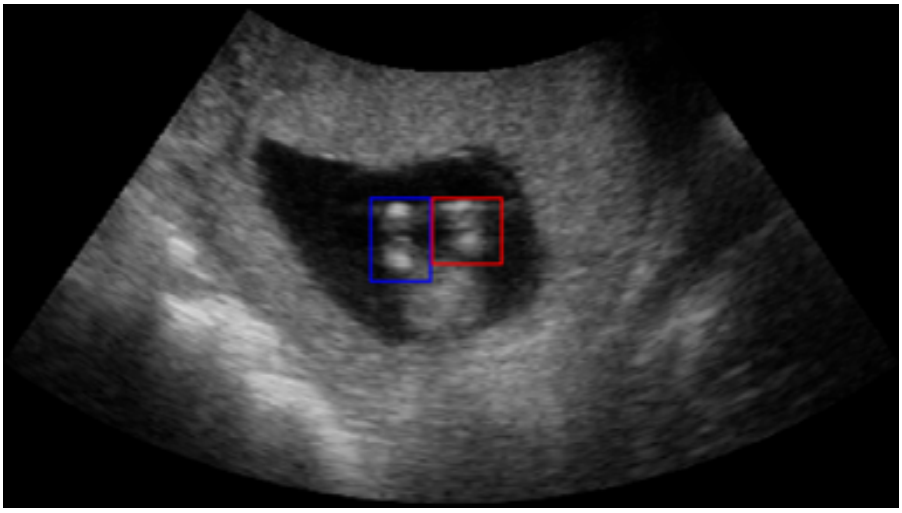


(b)

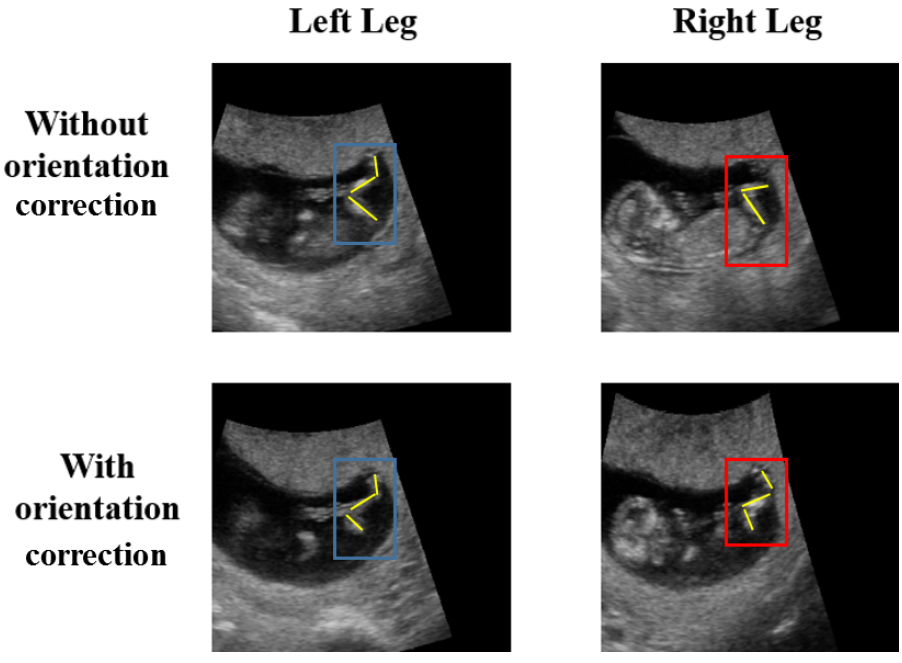
Figure 5.20: Fetal arm detection and its visualization in sagittal view where blue indicates the left arm and red box indicates the right arm, and yellow lines indicate each arm segment (upper, lower, hand). (a) The localization of the arms in the axial view based on the segmentation results using a U-Net. (b) Based on the segmentation, the orientation is changed following the methods explained in Section 5.2.3. The example shows that the image quality is improved after correcting the orientation.

	Head	**Abdomen**
The averaged orientation difference (°)	**9.45 ± 7.34**	**4.32 ± 2.42**

Table 5.12: The averaged difference and the standard deviation in the orientation angle between automatically and manually extracted planes for the head and the abdomen.



(a)



(b)

Figure 5.21: Fetal leg detection and its visualization in the sagittal view where blue indicates the left leg and red box indicates the right leg, and yellow lines indicate each leg segment (upper, lower, foot). (a) The localization of the legs in axial view based on the segmentation results using a U-Net. (b) Based on the segmentation the orientation is changed following the methods explained in Section 5.2.3. The example shows that the image quality is improved after correcting the orientation.

	HC (mm)	AC (mm)	CRL (mm)	GA (days)
Averaged Difference	**6.03 ± 3.62**	**3.34 ± 2.84**	**-1.65 ± 5.21**	**-0.91 ± 2.83**

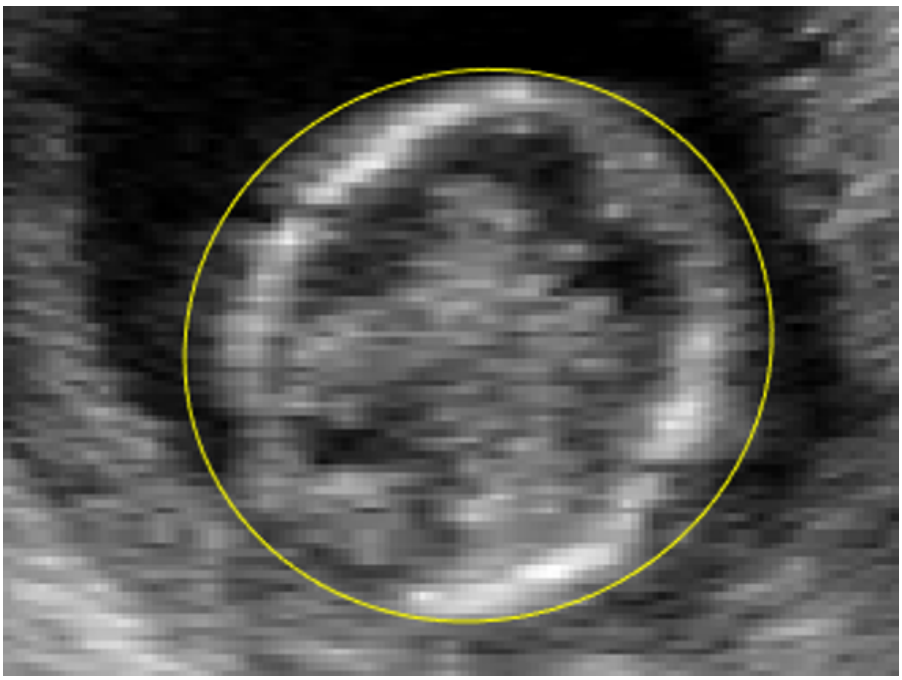
Table 5.13: The averaged difference and the standard deviation of the fetal biometry measurements between automatic and manual measurements.

As shown in Table 5.11, the averaged distances between the automatically and manually extracted plane for the head and abdomen are 1.98mm ±1.19 and 2.51mm ±1.58, respectively. As shown in Table 5.12, auto-manual orientation differences are 9.45 ± 7.34 and 4.32 ± 2.42 for the head and abdomen, respectively.

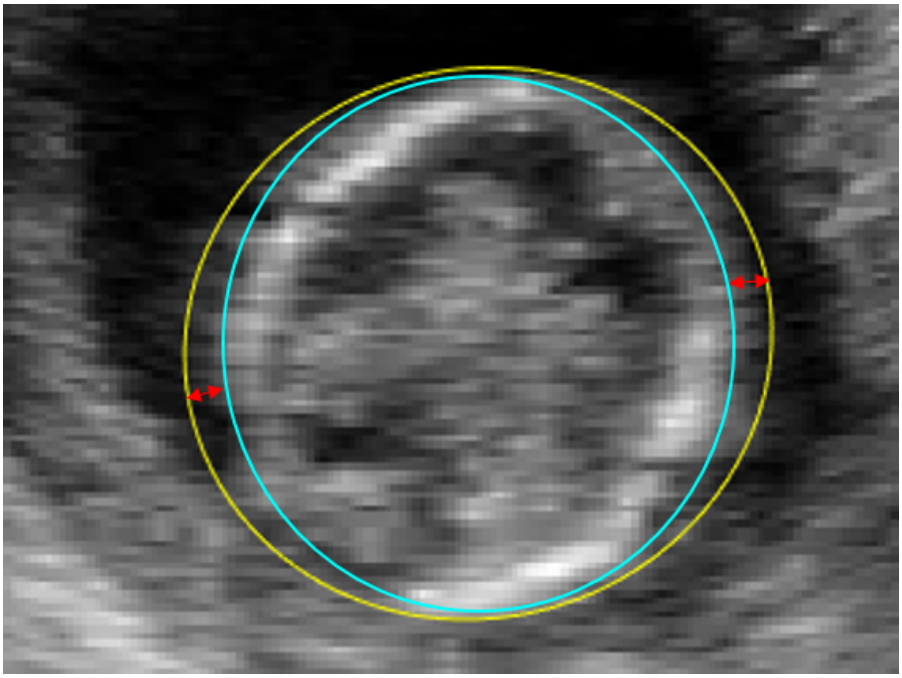
Automatic and manual measurements are compared in Table 5.13. As Table 5.13 shows, the differences in each measurement are small compared to the size of the corresponding fetal parts. For example, the averaged HC measurement for the data used in this work is 77.8 mm and the average difference is 6.03 mm. GA difference is only 1 day. The large auto-manual differences in the HC and AC are mainly caused by the segmentation results as shown in Figure 5.22. The boundary around the fetal head is unclear which may have led to over-segmentation. Therefore, the ellipse around the segmentation result becomes larger than the correct ellipse as shown in Figure 5.22 (b). Since the measurement is based on the fitted ellipse, the automatic measurement becomes larger than the manual measurement.

Automatic-Manual and Manual-Manual

To validate the automatic results, comparison between the automatic and manual$_1$ (A-M$_1$), automatic and manual$_2$ (A-M_2) and the inter-observer variability (M_2-M$_1$) has been made. The averaged distance and the averaged orientation difference are reported in Table 5.14 and 5.15.



(a)



(b)

Figure 5.22: The example of the measurement result for the head. (a) The ellipse is fitted based on the segmentation results, but due to the unclear boundary, the algorithm over-segments the head and this yields larger ellipse to be fitted compared to the correct ellipse (blue) as shown in (b).

Fetal parts	**Head**			**Abdomen**		
Comparison	**A-M_1**	**A-M_2**	**M_2-M_1**	**A-M_1**	**A-M_2**	**M_2-M_1**
Averaged Distance (mm)	**1.98 $\pm$ 1.19**	**2.43 $\pm$ 0.94**	**2.11 $\pm$ 1.06**	**2.51 $\pm$ 1.58**	**2.34 $\pm$ 1.60**	**1.92 $\pm$ 1.69**

Table 5.14: The averaged distance between automatic and *manual*$_1$ (A-M_1), automatic and *manual*$_2$ (A-M_2) and the inter-observer variability (M_2-M_1). The averaged distance between both A-M_1 and A-M_2 have shown similar results compared to M_2-M_1 in head- and abdomen-plane extractions

Fetal parts	**Head**			**Abdomen**		
Comparison	**A-M_1**	**A-M_2**	**M_2-M_1**	**A-M_1**	**A-M_2**	**M_2-M_1**
Orientation difference ($^\circ$)	**9.45 $\pm$ 7.34**	**20.30 $\pm$ 9.39**	**20.13 $\pm$ 12.65**	**4.32 $\pm$ 2.42**	**5.93 $\pm$ 4.78**	**7.12 $\pm$ 5.13**

Table 5.15: Orientation differences in A-M_1, A-M_2 and M_2-M_1. Note that the orientation differences are larger for the head which shows that the variability is large for the head-plane extraction.

	A-M_1	A-M_2	M_2-M_1
HC (mm)	**6.03 ± 3.62**	**5.04 ± 3.15**	**1.00 ± 3.43**
AC (mm)	**3.34 ± 2.84**	**4.57 ± 4.34**	**-1.40 ± 4.22**
CRL (mm)	**-1.65 ± 5.21**	**-0.69 ± 4.38**	**-0.93 ± 1.96**
GA (days)	**-0.91 ± 2.83**	**-0.41 ± 2.49**	**-0.51 ± 1.08**

Table 5.16: The averaged difference in the measurement in A-M_1, in A-M_2 and in M_2-M_1. Due to the over-segmentation results from a FCN, the automatic measurements for both HC and AC are larger than both manaual measurements. The CRL and GA, however, showed similar results.

	Detection rate
Left Arm	**0.65**
Right Arm	**0.50**
Left Leg	**0.55**
Right Leg	**0.35**

Table 5.17: Fetal limb detection rate. (n=21)

As shown in Table 5.14, the averaged distances from A-M_1 and A-M_2 are similar to those of M_2-M_1, which shows that the automatic plane extraction result is within an acceptable range. It is worthy to note that the orientation differences vary alot for the head as shown in Table 5.15. This means that there is a large variability in plane extraction, especially for the head where the inter-observer difference is 20.13° whereas the abdomen showed less difference (7.12°).

The difference in the measurement is also compared in Table 5.15. As shown in Table 5.16, the automatic measurements for the HC and AC are larger than both manual measurements. This supports the previous analyses based on Table 5.13 and Figure 5.22. Since the FCN over-segments the fetal parts, the automatic measurements are larger than the manual measurements whereas the inter-observer difference is much smaller. There is not much difference for the CRL measurements and the GA that is within 1 day of the manual prediction.

Limbs

Fetal limb detection results are shown in Table 5.17. Note that the limb detection rate is low as shown in Table 5.17. Therefore, for the orientation validation, we excluded the data for which automatic limb detection failed. The orientation differences are shown in Table 5.18.

	Orientation Difference(°)
Left Arm	**12.6 ± 16.1**
Right Arm	**13.7 ± 12.9**
Left Leg	**8.7 ± 9.7**
Right Leg	**9.5 ± 6.3**

Table 5.18: The averaged orientation difference of the limbs. Although the automatically extracted plane is similar to the manually extracted plane, there is a large variance in the orientation.

	$Head_{Auto}$	$Head_{Manual}$	$Abdomen_{Auto}$	$Abdomen_{Manual}$
Qualitative Assessment (n=21)	**19**	**20**	**15**	**19**

Table 5.19: The qualitative analysis of head and abdomen plane detection. The clinician has verified that the automatically extracted plane of the head has shown good visualization for the anatomical assessment since both automatically and manually extracted plane obtained similar results. The automatically extracted plane of the abdomen also showed good visualization.

Overall, the detection rate of the limbs is low. In this thesis, the location of the limbs is simply approximated based on the location of the head and torso where the arms are assumed to be located in the upper torso and the legs are located at the end of the torso. However, the low detection rate suggests that there needs to be further prior knowledge applied for detection. See chapter 6 (Conclusions) for a further discussion.

5.5.4 Qualitative Assessment

Head and Abdomen

As mentoned in Section 5.4.2, a qualitative assessments of head and abdomen plane detection have been performed and the results are shown in Table 5.19.

The qualitative analysis of the head and abdomen has shown good results, especially for the head, as shown in Table 5.19. A clinician has verified that the automatically extracted plane of the head has shown good visualization for the anatomical assessment since both automatically (19/21) and manually extracted plane (20/21) obtained similar results. The automatically extracted plane of the abdomen also showed good visualization. (15/21) planes have obtained good visualization of the abdomen. The reason might be the only characteristic of the abdomen is the stomach bubble, but due to its small size, it might be more difficult to be visualized, whereas the choroid plexus in the head is large which is relatively easier to visualize.

	Manual	$Auto_1$
Left Arm (n=21)	**17**	**7**
Right Arm (n=21)	**13**	**5**
Left Leg (n=21)	**21**	**7**
Right Leg (n=21)	**16**	**5**

Table 5.20: The qualitative assessment for the limbs. Both arms and legs have shown low qualitative analysis results due to the low detection rate as shown in Table 5.17.

	$Auto_1$	$Auto_2$
Left Arm	**0.33**	**0.47**
Right Arm	**0.24**	**0.45**
Left Leg	**0.33**	**0.70**
Right Leg	**0.24**	**0.56**

Table 5.21: Comparison between automatic results with/without non-detected plane for the limbs ($Auto_1$ and $Auto_2$). The qualitative analysis results have improved especially for the left leg.

Limbs

A qualitative assessment of the limb plane detection is summarized in Table 5.20.

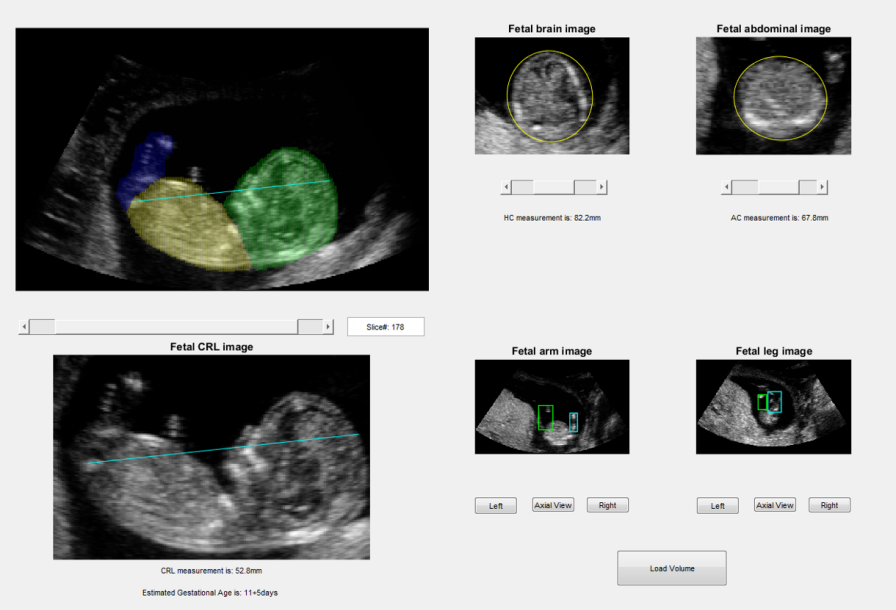
First, note that manually extracted planes are quite good (17/21, 13/21, 21/21 and 16/21). The automatic results are not as good but reflect the challenging nature of the plane detection task.

Since the limbs are not detected in some data, we also analysed how much the qualitative results would improve if we analyse the data where the limb detection was successful. This is to analyse how much the detection rate affects the visualization. Table 5.21 summarizes comparison in the qualitative assessment between automatic results from all test data ($Auto_1$) and from the test data excluding non-detected data ($Auto_2$).

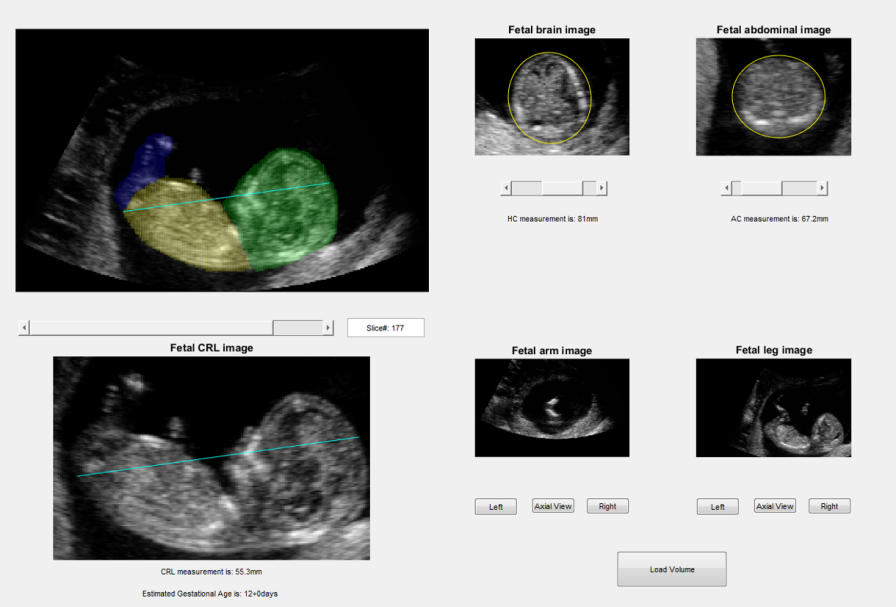
The qualitative assessment results are better (column 3, Table 5.21). The qualitative assessment for the left leg, for example, is improved from 0.33 to 0.70. This indicates that in the case when the limbs can be detected, visualization is better.

5.5.5 Integrated System

The developed methods were integrated into a software application with an interactive user interface. Refer to Figure 5.23.



(a)



(b)

Figure 5.23: Integrated software interface. a) The initial result b) The user can change the slices and the system outputs the corresponding measurement in real time.

As shown in Figure 5.23 (b), instead of showing only one slice for each fetal parts, the system outputs 10 slices. This allows the user to check the other slices with the corresponding measurements for the better assessments.

5.6 Summary

This chapter has described original methods for automatic first trimester fetal biometry assessment based on FCN-based neural networks which is, in our knowledge, the first attempt to do automatic biometry from a single 3D volume. As the first task, the localization of the whole fetus is performed using a FCN and a comparison is presented with methods proposed in Chapter 4 (handcrafted features). The FCN-based algorithm was found to have a higher localization accuracy of the whole fetus (0.84 versus 0.74). A FCN multi-task network was implemented to extract a sagittal plane and perform segmentation. The accuracy of the method was 89.4% (versus 85.8%) and IoU is 0.76 (versus 0.58). This was the first attempt to apply a FCN multi-task network on a 3D first trimester fetal volume to partition the whole fetus into the head, torso and the lower limb.

Based on the localization and the segmentation result, we were able to focus attention on the fetal parts in the axial view and able to correct the plane orientation for better visualization of the fetal parts. With the image focused on the fetal parts, we achieved 95.0% and 96.3% in pixel accuracy and 0.92 and 0.94 in IoU of the head and abdomen, respectively. In this chapter, the limbs are also included which is the most challenging part of the research and has not been studied before. We achieved promising results - 85.0% and 88.3% in pixel accuracy, and IoU of 0.70 and 0.63 for the arms and the legs, respectively.

With the whole fetus segmentation results, we were able to extract the biometric plane for the head and abdomen with orientation-correction. As shown in Figure 5.3, the orientation is the important factor for the better visualization of the fetal parts in the axial view. We were able to automatically change the plane orientation for each head and torso based on the segmentation results, which is another original contribution of the thesis. With the orientation-corrected plane, we achieved an averaged orientation difference between estimation and groundtruth of $9.45^{\circ}(\pm 7.34)$ and $4.32^{\circ}(\pm 2.42)$ which was found

to be in the acceptable range compared to the inter-observer variance (20.13°(±12.65) and 7.12°(±5.13) for the head and abdomen, respectively). By having the corrected-orientation, plane extraction achieved a similar performance with respect to a manual reference (1.98 mm(±1.19) versus 2.11 mm(±1.06) and 2.51 mm(±1.58) versus 1.92 mm(±1.69) for the head and abdomen, respectively). As mentioned before, one of the primary goal of providing computer-aided analysis is to reduce the variance, and as the result shows, we achieved less than or similar to inter-observer variance. This shows the potential of inter-observer-variance reduction using automatic biometry.

We also achieved good plane detection for the head and abdomen with respect to the manually extracted plane (19/21 versus 20/21 and 15/21 versus 19/21 for the head and abdomen respectively). Automatic measurements for the HC and AC were found to be larger than the manual measurements (6.03 mm(±3.62) and 3.34 mm(±2.84)) due to the over-segmentation. The CRL has shown promising results. The GA prediction from the CRL had an estimated 1 day difference between automatic and manual measurement.

Due to the low detection rate and qualitative results, we cannot draw strong conclusions about limb assessment. However, we found that if the limb can be detected well enough, the limbs' plane images could be found.

6
Conclusions and Future Work

This chapter summarizes the contributions of this thesis and describes possible directions of future studies.

6.1 Conclusions

In this thesis, we applied machine learning method to do automatic fetal assessments in the first trimester and showed the feasibility. We first proposed to develop two consecutive random forests to localize the whole fetus within the manually selected sagittal plane (chapter 3) and then used a CNN to extract the head and the abdomen plane (chapter 4). In these two chapters, we showed the potential of using machine learning approach to do the automated fetal assessments. We also showed that deep learning methods can achieve similar or better results compared to machine learning methods that require the handcrafted features. However, although the results looked promising, the final results still require several consecutive steps that can yield more errors in the later steps. Especially, the first step, which is the edge detection by structured random forests, is not based on the fetus itself. Therefore, we used a FCN to achieve better performance with less number of steps and no handcrafted features. In addition, we were also able to do more automated fetal assessments such as the automated measurements, correction of the plane orientation to the standardized plane orientation and the fetal limb assessments.

6.1.1 Contributions

In this work, we developed a FCN-based approach to automatic biometry assessment for the first trimester fetus from a 3D ultrasound volume. To our knowledge, this is the first study to do this. Specifically, our first task localizes the whole fetus. This is the first study to exploit deep learning for automatic localization of the whole fetus from a 3D volume. The developed method has shown a very good localization with a good performance in both good sagittal plane extraction and the semantic segmentation of the fetus compared to a method using handcrafted features. Especially, the performance in the semantic segmentation of the whole fetus was achieved surprisingly well that the network was able to distinguish the head and the torso without any human interactions like the handcrafted features. This might prove that the network can distinguish the different structures even though they are attached to each other, and have the potential to be used for more complex segmentation such as the internal organs since they are more likely to be adjacent to each other. We also proved that whole fetus localization improved the performance of a FCN relative to processing a full 3D volume.

Secondly, based on the segmentation/partitioning results, we developed a method for plane orientation-correction to each of the head and the body for visualization of fetal parts in the axial view. This is another unique work that has not been studied before. Having both segmentation and corrected orientation, we were able to perform automatic plane extraction and measurement. Automated measurement compared favorably with the manual case.

We also performed automatic limb assessment which has not been studied before. Surprisingly, although the result was not as good as other fetal anatomies, a FCN was able to distinguish the limbs from the torso and the non-fetal tissue. Unlike the head and the abdomen which have unique pattern that can be distinguished from others, the limbs have relatively less pattern. It is possible that if any region that is attached to the torso, the network annotate them as the limbs. This is encouraging result because we achieved this without any particular prior knowledge, and this will be the potential future work to study.

As mentioned before, this is the fist work of automatic first trimester fetal assessment from a 3D ultrasound volume. We have shown that our methods are capable of performing

automatic biometry assessment and has potential to reduce inter-observer variability in measurement.

6.2 Limitations and Future Work

There are several possible extensions of this thesis in both clinical and technical directions.

6.2.1 Dataset

The ultrasound dataset used in this thesis is a 3D first trimester fetal ultrasound volume dataset based on the CRL criteria, as explained in Appendix A. However, this 3D scan is currently not considered routine scan. Because of this, the size of the data used in this research is small. Although we increased the data size by extracting 2D slices from a 3D volume and did data augmentation, the data size is still relatively small compared to the data that can be obtained from the 2D routine scan. To be more generalized, the larger data size is still required. Also, to improve the performance, acquiring more data is necessary since the data size is an important factor in deep learning. Following the guidelines/criteria mentioned in Appendix A, acquiring more 3D first trimester fetal ultrasound volume (at least 100 volumes) allow us to extract larger dataset of the 2D slices and it is even possible to design 3D network architectures to use the data as 3D instead of 2D. Also, it would be possible to use the existing 2D data that includes the classes that we worked on for this thesis (the head, the abdomen, the whole fetus and the limbs). Although the fetal anatomies in the 2D slices might be visually different in terms of the size and their more detailed internal structures, the general pattern of the fetal anatomies will be similar. For example, the ring-shaped fetal skull surrounding the head and the stomach bubble in the abdomen can be visualized in both 2D and 3D scan. Therefore, we can use the 2D data as the training set or we can use transfer learning [173] from the pre-trained network based on the 2D slices and fine-tune the parameters of our network. This would be interesting to do in future work.

6.2.2 Clinical

Although a prototype has been built, it has not undergone clinical evaluation. Having more than two clinicians to compare with their manual assessments is one possible way of evaluation. Another requirements to make the user interface more feasible, the system might need to check the quality as well. Since the quality of the image can be varied due to the ultrasound's dependence on the sonographer's skill and the image artefacts, the quality of the image needs to be checked before the assessments. These automatic algorithms can be done by training the network with the good-quality images as the classification task whether the image has good quality or not. If we adopt the quality-check or scoring algorithms into the system, we can either filter them out or notify the sonographer to reperform the scan before the actual assessments.

6.2.3 Biometry plane extraction and measurements

In this thesis, regression is used to estimate the location of the biometry plane. However, this approach can be highly affected by the data size. Other than acquiring more datasets, locating the biometry plane might be improved by using regression or a classifier based on the deep learning to the fetal parts in the axial view. One possible way is to use multitask network. As shown in Chapter 5, a multitask network can do a multiple tasks instead of having multiple single networks. We can use multitask network to do two tasks: classify whether the image contains fetal anatomies that we want to do the assessments, and if it is, do the segmentation and the measurements. Using the regression can give wrong images because it approximates the location without knowing whether that is correct image or not. However, since a multitask network can distinguish the fetal anatomies from the others, we might be able to get good performance in the measurements due to the better quality.

Plane orientation estimation might also be improved by considering the coronal view. Due to the fetal movement, it is highly possible that the fetus might not lie straight. Because of this, the coronal plane orientation can be different from standardized plane orientation as shown in Figure 6.1 (c).

As shown in Figure 6.1, when we did not correct any plane orientation (Figure 6.1 (a)), the quality of the head is not good enough to be used for the assessments. However,

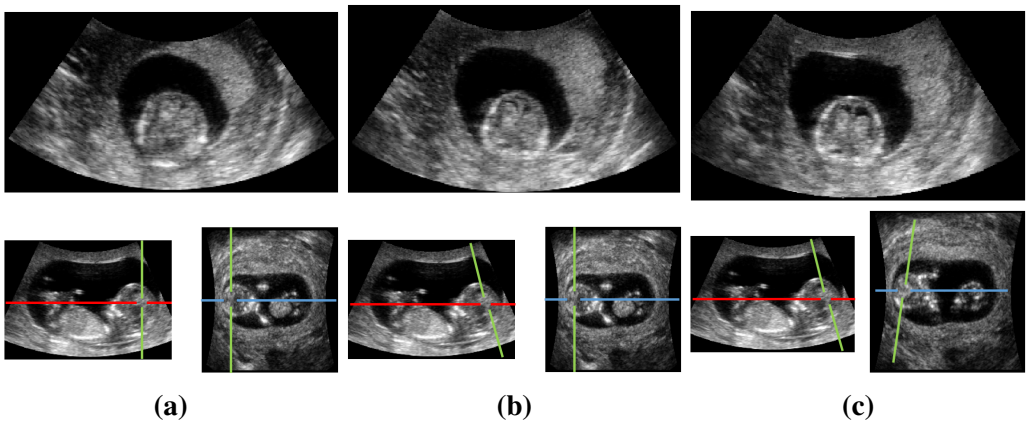


Figure 6.1: The examples of how the head is visualized when we consider a) no plane orientation, b) only axial plane orientation and c) axial- and coronal plane orientation. When we do not consider any plane orientation (a), the quality of the head is good enough to be used for the assessments. However, as we consider the plane orientations (b and c), the quality of the head is getting better.

as we correct the plane orientations (Figure 6.1 (b) and (c)), the quality of the head is getting better that the visualization of the choroid plexus becomes more clear. Since both the anatomical assessments and automated measurements depend on the quality of the image, the coronal plane orientation must be corrected to the standard plane orientation. One possible way is to do the segmentation of the whole fetus into the head and torso in the coronal view, and then estimate the orientation using the mid-line of the segmented region as we did in the sagittal view in chapter 5. Another possible way is to do the 3D segmentation using the 3D network, but this requires manual 3D annotation which is very difficult to obtain. Therefore, it might be better to use one 2D slice from each view and estimate the orientation.

The automated measurement accuracy might also be increased with improved automated segmentation. As shown in Figure 5.21, the larger ellipse is obtained due to the oversegmentation. To overcome this, the segmentation with finer boundaries or post-process the segmentation region will give more accurate ellipse. Using networks, like U-net, might improve the performance due to their ability to output finer boundaries or using other techniques such as edge-detection to post-process the segmentation region might also yield finer boundaries. For example, instead of using the segmentation result itself, use it to crop the image to focus more on the fetal anatomy and then use phase-based

method similar to [164] or use structured random forests [121] to extract the boundary of the fetal anatomy. Based on the extracted boundary, fit the ellipse, and then do the automatic measurements. Another possible way is to use the graphical model similar to [187] as post-processing to refine the boundary of the segmented region.

6.2.4 Biometry assessment for the limbs

Automatic fetal limb assessment formed the most challenging part of the research. The limbs are difficult to visualize and their size is relatively small compared to the fetus and non-fetal tissue. Due to this, the data size of the limbs are small. One simple way to increase the data size is to get more volumes, but it might not increase as well as other fetal anatomies due to their difficulties in the visualization especially in the 3D unless the data is focused on the limbs. Therefore, another possible way is to include the 2D data of the limbs as part of the training set, since they might have more detailed pattern.

Technically, it might be possible to improve the performance using a 3D network and/or incorporates prior knowledge such as the spatial relationship between the localized fetal parts. However, the only criteria for the limbs in clinical practice is their presence. Unlike other fetal anatomies such as the head which should visualize the butterfly-shaped choroid plexus, the limbs do not have any particular patterns except their attachment to the torso. More clinical knowledge might be needed to be studied.

Appendices

A Appendix

A.1 Data acquisition

Data was acquired by the clinicians with a Philips V7-3 (7.3 MHz) volume-curved transducer with a spatial resolution of 0.33mm $\times$ 0.33mm $\times$ 0.33mm. The acquisition protocol followed the CRL criteria (good sagittal plane for whole fetus):

1. The fetus should be horizontal (at 90° to the angle of insonation) – posterior spine position;
2. The image should fill at least 30% of the monitoring screen;
3. The crown and rump should both be clearly visible (the whole fetus must be present); and,
4. The fetus should be in a neutral position (not hyperextended or flexed).

A typical example of a visually good and bad data acquisition are shown in Figure A.1.

Figure A.1, for example, shows the fetus in the mid-sagittal section in the horizontal and neutral position. Also, the image has a good magnification of the fetus. Figure A.1(b), however, shows the fetus is too small compared to the size of the whole image and the fetus is not facing upward. Only data satisfying the acquisition protocol above was used in this research.

CORRECT:

The section is mid-sagittal

Horizontal position

Neutral position

Good magnification

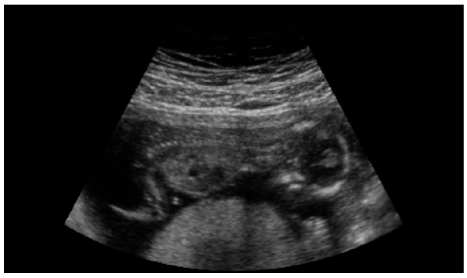
(a)

INCORRECT:

The section is not mid-sagittal

Magnification is poor

You cannot see the bottom part of the fetus



INCORRECT:

The fetus is in anterior spine position.

We need to acquire volumes of fetuses in posterior spine position with visible amniotic fluid behind the back

(b)

Figure A.1: The data acquisition criteria a) Visually good data acquisition. The image is taken in the mid-sagittal section and the fetus is in horizonatal and neutral position. Also, the fetus covers more than 30% of the image. b) Visually bad data acquisition. The image is taken with incorrect acquisition that the magnification is poor and the fetus is in anterior spine position which the fetus is not facing upward.

A.2 Groundtruth

The information regarding the groundtruth has been given by the experts/clinicians.

A.2.1 Bounding Box

The bounding box is manually drawn around the whole fetus from the 2D slices extracted from the 3D volumes as shown in Figure A.2. The criteria for the groundtruth is to contain the whole fetus without any missing fetal parts. The manually drawn boxes are

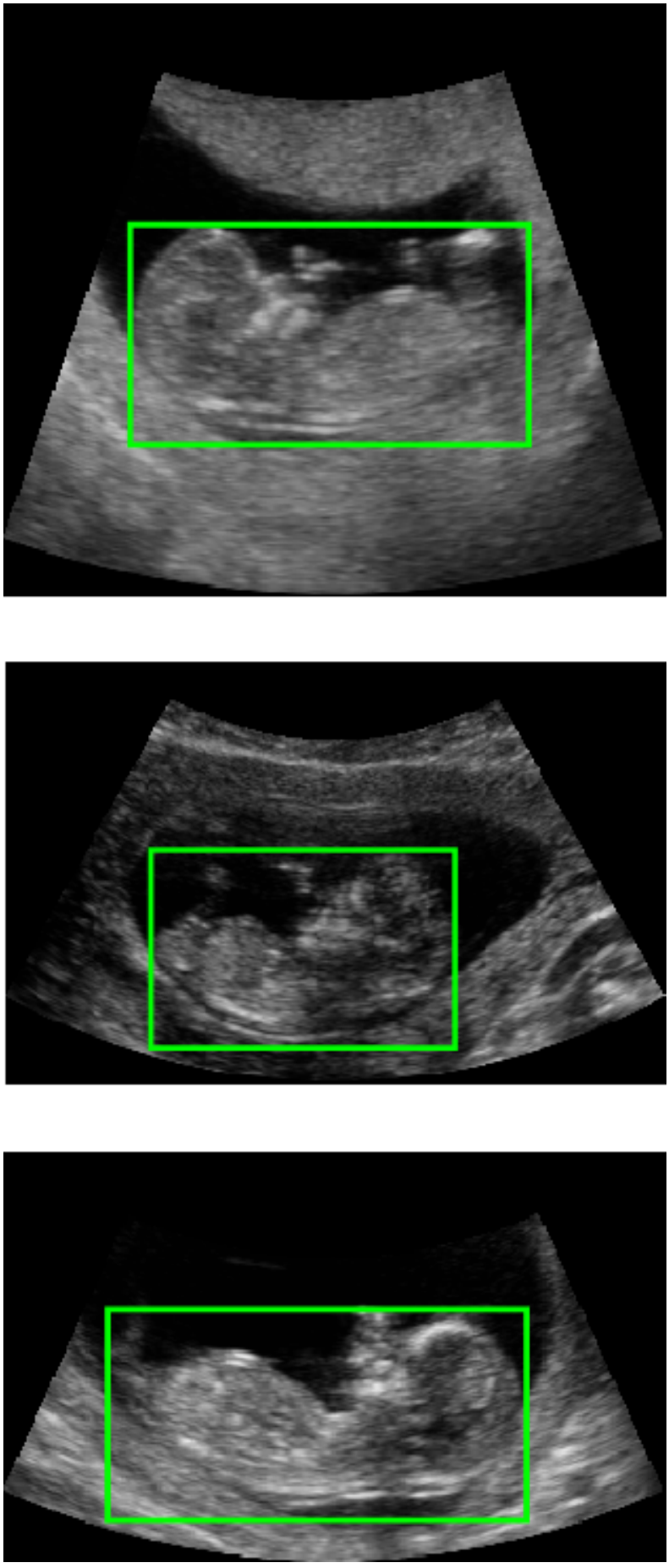


Figure A.2: The example of the bounding box manually drawn around the whole fetus.

then checked by the clinicians to ensure that there are no missing fetal parts inside the bounding-box.

A.2.2 Classification

The 2D slices are extracted from the 3D volume and has been classified by the clinicians according to ISUOG guidelines [5] as shown in Table 2.1 in Section 2.2.1.

A.2.3 Segmentation

The segmentation labelling for the fetal parts has been done by ourselves and has been checked by the clinicians.

Bibliography

[1] S. Pathak and C. Lees, "Ultrasound structural fetal anomaly screening: an update," *Archives of Disease in Childhood - Fetal and Neonatal Edition*, vol. 94, no. 5, pp. F384–F390, 2009.

[2] W. H. Organization, W. H. Organization, *et al.*, "Report on the regional consultation towards the development of a strategy for optimizing fetal growth and development," *WHO Regional Office for the Eastern Mediterranean: Cairo*, 2005.

[3] L. Ammon Avalos, C. Galindo, and D.-K. Li, "A systematic review to calculate background miscarriage rates using life table analysis," *Birth Defects Research Part A: Clinical and Molecular Teratology*, vol. 94, no. 6, pp. 417–423, 2012.

[4] D.-K. Li, R. Odouli, S. Wi, T. Janevic, I. Golditch, T. D. Bracken, R. Senior, R. Rankin, and R. Iriye, "A population-based prospective cohort study of personal exposure to magnetic fields during pregnancy and the risk of miscarriage," *Epidemiology*, vol. 13, no. 1, pp. 9–20, 2002.

[5] "Isuog practice guidelines: performance of first-trimester fetal ultrasound scan," *Ultrasound in Obstetrics & Gynecology*, vol. 41, no. 1, pp. 102–113, 2013.

[6] Y. S., C. S. M., S. I., and V. D. V., "3d and 4d ultrasound in fetal cardiac scanning: a new look at the fetal heart," *Ultrasound in Obstetrics & Gynecology*, vol. 29, no. 1, pp. 81–95.

[7] C. C., H. J. L., R. M. M., and P. D. J., "Population-based study of antenatal detection of congenital heart disease by ultrasound examination," *Ultrasound in Obstetrics & Gynecology*, vol. 29, no. 6, pp. 619–624.

[8] R. F. Brem, M. J. Lenihan, J. Lieberman, and J. Torrente, "Screening breast ultrasound: Past, present, and future," *American Journal of Roentgenology*, vol. 204, pp. 234–240, Feb. 2015.

[9] K. J. Taylor, C. Merritt, C. Piccoli, R. Schmidt, G. Rouse, B. Fornage, E. Rubin, D. Georgian-Smith, F. Winsberg, B. Goldberg, and E. Mendelson, "Ultrasound as a complement to mammography and breast examination to characterize breast masses," *Ultrasound in Medicine & Biology*, vol. 28, no. 1, pp. 19 – 26, 2002.

[10] K. M. Hiltawsky, M. Krüger, C. Starke, L. Heuser, H. Ermert, and A. Jensen, "Freehand ultrasound elastography of breast lesions: clinical results," *Ultrasound in Medicine & Biology*, vol. 27, no. 11, pp. 1461 – 1469, 2001.

[11] L. J. Salomon, Z. Alfirevic, V. Berghella, C. Bilardo, E. Hernandez-Andrade, S. L. Johnsen, K. Kalache, K.-Y. Leung, G. Malinger, H. Munoz, F. Prefumo, A. Toi, W. Lee, and on behalf of the ISUOG Clinical Standards Committee, "Practice guidelines for performance of the routine mid-trimester fetal ultrasound scan," *Ultrasound in Obstetrics & Gynecology*, vol. 37, no. 1, pp. 116–126, 2011.

[12] A. Ebrashy, A. El Kateb, M. Momtaz, A. El Sheikhah, M. M. Aboulghar, M. Ibrahim, and M. Saad, "13–14-week fetal anatomy scan: a 5-year prospective study," *Ultrasound in Obstetrics and Gynecology*, vol. 35, no. 3, pp. 292–296, 2010.

[13] M. M. J. Letteboer, P. W. A. Willems, M. A. Viergever, and W. J. Niessen, "Brain shift estimation in image-guided neurosurgery using 3-d ultrasound," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 2, pp. 268–276, Feb.

[14] J. A. Noble, N. Navab, and H. Becher, "Ultrasonic image analysis and image-guided interventions," *Interface Focus*, vol. 1, pp. 673–685, May 2011.

[15] J. E. Aldrich, "Basic physics of ultrasound imaging," *Critical Care Medicine*, vol. 35, no. 5, 2007.

[16] J. W. Hunt, M. Arditi, and F. S. Foster, "Ultrasound transducers for pulse-echo medical imaging," *IEEE Transactions on Biomedical Engineering*, vol. BME-30, no. 8, pp. 453–481, Aug.

[17] Y.-S. Sohn, M.-J. Kim, J.-Y. Kwon, Y.-H. Kim, and Y.-W. Park, "The usefulness of fetal mri for prenatal diagnosis," *Yonsei Medical Journal*, vol. 48, pp. 671–677, Jan. 2007.

[18] D. Pugash, P. C. Brugger, D. Bettelheim, and D. Prayer, "Prenatal ultrasound and fetal mri: The comparative value of each modality in prenatal diagnosis," *European Journal of Radiology*, vol. 68, pp. 214–226, Nov. 2008.

[19] N. Girard, K. Chaumoitre, F. Chapon, S. Pineau, M. Barberet, and H. Brunel, "Fetal magnetic resonance imaging of acquired and developmental brain anomalies," vol. 33, pp. 234–250, 2009.

[20] M. Hamisa, N. Dabees, W. M. Ataalla, and D. H. Ziada, "Magnetic resonance imaging versus ultrasound examination in detection of prenatal fetal brain anomalies," *The Egyptian Journal of Radiology and Nuclear Medicine*, vol. 44, pp. 665–672, Sept. 2013.

[21] R. L. d. S. Ximens, J. Szejnfled, A. R. d. S. Ximenes, and V. Zanderigo, "A critical review of benefits and limitations of magnetic resonance imaging as a complementary method in the diagnosis of fetal malformations," *Radiologia Brasileira*, vol. 41, pp. 313 – 318, 10 2008.

[22] E. Kanal, A. J. Barkovich, C. Bell, J. P. Borgstede, W. G. Bradley, J. W. Froelich, T. Gilk, J. R. Gimbel, J. Gosbee, E. Kuhni-Kaminski, J. W. Lester, J. Nyenhuis, Y. Parag, D. J. Schaefer, E. A. Sebek-Scoumis, J. Weinreb, L. A. Zaremba, P. Wilcox, L. Lucey, and N. Sass, "Acr guidance document for safe mr practices: 2007," *American Journal of Roentgenology*, vol. 188, pp. 1447–1474, June 2007.

[23] P. D., M. G., B. P. C., C. C., D. C. L., D. K. B., F. G. L., G. P., G. L. F., G. G. M., L.-N. S., L. W., M. A.-E., M. M., N. J., P. L., P. D., R. P., S. L. J., S. M., T.-T. I. E.,

T. B., T. D., W. M., X. R., and R.-F. N., "Isuog practice guidelines: performance of fetal magnetic resonance imaging," *Ultrasound in Obstetrics & Gynecology*, vol. 49, no. 5, pp. 671–680.

[24] N. T. R., P. D. H., H. A., R. M., S. M. S., and J. G., "Sources and impact of artifacts on clinical three-dimensional ultrasound imaging," *Ultrasound in Obstetrics & Gynecology*, vol. 16, no. 4, pp. 374–383.

[25] A. Hindi, C. Peterson, R. G. Barr, C. Richard, and G. Barr, "Reports in medical imaging 2013:6 29–48 reports in medical imaging artifacts in diagnostic ultrasound."

[26] P. B. Bertrand, R. A. Levine, E. M. Isselbacher, and P. M. Vandervoort, "Fact or artifact in two-dimensional echocardiography: Avoiding misdiagnosis and missed diagnosis," *Journal of the American Society of Echocardiography : official publication of the American Society of Echocardiography*, vol. 29, pp. 381–391, Mar. 2016.

[27] M. K. Feldman, S. Katyal, and M. S. Blackwood, "Us artifacts," *RadioGraphics*, vol. 29, no. 4, pp. 1179–1189, 2009. PMID: 19605664.

[28] F. G. Sommer, R. Filly, and M. J. Minton, "Acoustic shadowing due to refractive and reflective effects," *American Journal of Roentgenology*, vol. 132, no. 6, pp. 973–979, 1979.

[29] H. Ahn, E. Hernández-Andrade, R. Romero, M. Ptwardhan, L. F. Goncalves, A. Aurioles-Garibay, M. Garcia, S. S. Hassan, and L. Yeo, "Mirror artifacts in obstetric ultrasound: case presentation of a ghost twin during the second-trimester ultrasound scan," *Fetal diagnosis and therapy*, vol. 34, no. 4, pp. 248–252, 2013.

[30] A. C. of Radiology *et al.*, "Acr-spr-sru practice guideline for performing and interpreting diagnostic ultrasound examinations," *American College of Radiology website*, 2011.

[31] I. Sarris, C. Ioannou, P. Chamberlain, E. Ohuma, F. Roseman, L. Hoch, D. Altman, and A. Papageorghiou, "Intra-and interobserver variability in fetal ultrasound measurements," *Ultrasound in Obstetrics & Gynecology*, vol. 39, no. 3, pp. 266–273, 2012.

[32] S. Perni, F. Chervenak, R. Kalish, S. Magherini-Rothe, M. Predanic, J. Streltzoff, and D. Skupski, "Intraobserver and interobserver reproducibility of fetal biometry," *Ultrasound in obstetrics & gynecology*, vol. 24, no. 6, pp. 654–658, 2004.

[33] N. Dudley and E. Chapman, "The importance of quality management in fetal measurement," *Ultrasound in obstetrics & gynecology*, vol. 19, no. 2, pp. 190–196, 2002.

[34] R. L. Deter, R. B. Harrist, F. P. Hadlock, and R. J. Carpenter, "Fetal head and abdominal circumferences: I. evaluation of measurement errors," *Journal of Clinical Ultrasound*, vol. 10, no. 8, pp. 357–363, 1982.

[35] T. Chang, S. C. Robson, J. A. Spencer, and S. Gallivan, "Ultrasonic fetal weight estimation: Analysis of inter-and intra-observer variability," *Journal of clinical ultrasound*, vol. 21, no. 8, pp. 515–519, 1993.

[36] H. Steiner, A. Staudach, D. Spitzer, and H. Schaffer, "Diagnostic techniques: Three-dimensional ultrasound in obstetrics and gynaecology: technique, possibilities and limitations," *Human Reproduction*, vol. 9, pp. 1773–1778, Sept. 1994.

[37] A. Fenster and D. B. Downey, "3-d ultrasound imaging: a review," *IEEE Engineering in Medicine and Biology Magazine*, vol. 15, no. 6, pp. 41–51, Nov/.

[38] R. W. Prager, U. Z. Ijaz, A. Gee, and G. M. Treece, "Three-dimensional ultrasound imaging," *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, vol. 224, no. 2, pp. 193–223, 2010.

[39] A. Fenster, G. Parraga, and J. Bax, "Three-dimensional ultrasound scanning," *Interface Focus*, vol. 1, pp. 503–519, May 2011.

[40] Q. Huang and Z. Zeng, "A review on real-time 3d ultrasound imaging technology," *BioMed Research International*, vol. 2017, p. 6027029, Mar. 2017.

[41] J. F. Brinkley, S. K. Muramatsu, W. D. McCallum, and R. L. Popp, "In vitro evaluation of an ultrasonic three-dimensional imaging and volume system," *Ultrason Imaging*, vol. 4, pp. 126–139, Apr. 1982.

[42] M.-X. Xie, X.-F. Wang, T. O. Cheng, Q. Lu, L. Yuan, and X. Liu, "Real-time 3-dimensional echocardiography: A review of the development of the technology and its clinical application," *Progress in Cardiovascular Diseases*, vol. 48, pp. 209–225, Nov. 2005.

[43] A. Fenster, D. B. Downey, and H. N. Cardinal, "Three-dimensional ultrasound imaging," *Physics in Medicine & Biology*, vol. 46, no. 5, p. R67, 2001.

[44] R. Nelson T, H. Pretorius D, M. Sklansky, and S. Hagen-Ansert, "Three-dimensional echocardiographic evaluation of fetal heart anatomy and function: acquisition, analysis, and display.," *Journal of Ultrasound in Medicine*, vol. 15, pp. 1–9, Jan. 1996.

[45] W. Lee, "3d fetal ultrasonography," *Clinical obstetrics and gynecology*, vol. 46, no. 4, pp. 850–867, 2003.

[46] P. A. Yushkevich, J. Piven, H. C. Hazlett, R. G. Smith, S. Ho, J. C. Gee, and G. Gerig, "User-guided 3d active contour segmentation of anatomical structures: significantly improved efficiency and reliability," *Neuroimage*, vol. 31, no. 3, pp. 1116–1128, 2006.

[47] I. Wolf, M. Vetter, I. Wegner, T. Böttger, M. Nolden, M. Schöbinger, M. Hastenteufel, T. Kunert, and H.-P. Meinzer, "The medical imaging interaction toolkit," *Medical image analysis*, vol. 9, no. 6, pp. 594–604, 2005.

[48] H.-C. Kuo, F.-M. Chang, C.-H. Wu, B.-L. Yao, and C.-H. Liu, "The primary application of three-dimensional ultrasonography in obstetrics," *American Journal of Obstetrics & Gynecology*, vol. 166, no. 3, pp. 880–886, 1992.

[49] N. Zosmer, D. Jurkovic, E. Jauniaux, K. Gruboeck, C. Lees, and S. Campbell, "Selection and identification of standard cardiac views from three-dimensional volume scans of the fetal thorax.," *Journal of ultrasound in medicine*, vol. 15, no. 1, pp. 25–32, 1996.

[50] D. H. Pretorius, M. House, T. R. Nelson, and K. A. Hollenbach, "Evaluation of normal and abnormal lips in fetuses: comparison between three-and two-dimensional sonography.," *AJR. American journal of roentgenology*, vol. 165, no. 5, pp. 1233–1237, 1995.

[51] K. Baba, T. Okai, S. Kozuma, and Y. Taketani, "Fetal abnormalities: Evaluation with real-time-processible three-dimensional us–preliminary report," *Radiology*, vol. 211, pp. 441–446, May 1999.

[52] P. Acar, L. Battle, Y. Dulac, M. Peyre, H. Dubourdieu, S. Hascoet, M. Groussolles, and C. Vayssière, "Real-time three-dimensional foetal echocardiography using a new transabdominal xmatrix array transducer," *Archives of Cardiovascular Diseases*, vol. 107, no. 1, pp. 4 – 9, 2014.

[53] R. Dyson, D. Pretorius, N. Budorick, D. Johnson, M. Sklansky, C. Cantrell, S. Lai, and T. Nelson, "Three-dimensional ultrasound in the evaluation of fetal anomalies," *Ultrasound in obstetrics & gynecology*, vol. 16, no. 4, pp. 321–328, 2000.

[54] G. Michailidis, P. Papageorgiou, R. Morris, and D. Economides, "The use of three-dimensional ultrasound for fetal gender determination in the first trimester," *The British journal of radiology*, vol. 76, no. 907, pp. 448–451, 2003.

[55] J. Pedrosa, D. Barbosa, N. Almeida, O. Bernard, J. Bosch, and J. D'hooge, "Cardiac chamber volumetric assessment using 3d ultrasound - a review," 2016.

[56] T. A. Wallny, I. Theuerkauf, R. L. Schild, L. Perlick, and D. S. Bertelsbeck, "The three-dimensional ultrasound evaluation of the rotator cuff – an experimental study," *European Journal of Ultrasound*, vol. 11, pp. 135–141, May 2000.

[57] Z. Chen, Y. Chen, and Q. Huang, "Development of a wireless and near real-time 3d ultrasound strain imaging system," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 10, no. 2, pp. 394–403, Apri.

[58] C. z. Jin, K. H. Nam, and D. G. Paeng, "The spatio-temporal variation of rat carotid artery bifurcation by ultrasound imaging," in *2014 IEEE International Ultrasonics Symposium*, pp. 1900–1903, 3-6.

[59] S. Chaoyang, T. Carlos, W. Xianliang, I. Seiichi, K. Kimihiro, Y. Kiyohito, A. Fumihito, and F. Toshio, "Real-time in vitro intravascular reconstruction and navigation for endovascular aortic stent grafting," *Int J Med Robotics Comput Assist Surg*, vol. 12, pp. 648–657, Dec. 2016.

[60] D. Mingyue, C. H. Neale, and F. Aaron, "Automatic needle segmentation in three-dimensional ultrasound images using two orthogonal two-dimensional image projections," *Med. Phys.*, vol. 30, pp. 222–234, Feb. 2003.

[61] P. Yan, "Su-f-t-41: 3d mtp-trus for prostate implant," *Med. Phys.*, vol. 43, pp. 3470–3471, June 2016.

[62] D. B. Downey, A. Fenster, and J. C. Williams, "Clinical utility of three-dimensional us," *RadioGraphics*, vol. 20, pp. 559–571, Mar. 2000.

[63] T. R. Nelson and T. T. Elvins, "Visualization of 3d ultrasound data," *IEEE Computer Graphics and Applications*, vol. 13, no. 6, pp. 50–57, Nov.

[64] S. Saltvedt, H. Almström, M. Kublickas, L. Valentin, and C. Grunewald, "Detection of malformations in chromosomally normal fetuses by routine ultrasound at 12 or 18 weeks of gestation—a randomised controlled trial in 39 572 pregnancies," *BJOG: An International Journal of Obstetrics & Gynaecology*, vol. 113, no. 6, pp. 664–674, 2006.

[65] A. C. Rossi and F. Prefumo, "Accuracy of ultrasonography at 11-14 weeks of gestation for detection of fetal structural anomalies: A systematic review," *Obstetrics & Gynecology*, vol. 122, no. 6, 2013.

[66] M. L. Brizot, R. J. Snijders, N. A. Bersinger, P. Kuhn, and K. H. Nicolaides, "Maternal serum pregnancy-associated plasma protein a and fetal nuchal translucency thickness for the prediction of fetal trisomies in early pregnancy.," *Obstetrics and gynecology*, vol. 84, no. 6, pp. 918–922, 1994.

[67] S. Kevin, E. Spencer Christine, P. Maureen, D. Carolynne, and H. Nicolaides Kypros, "Screening for chromosomal abnormalities in the first trimester using ultrasound and maternal serum biochemistry in a one-stop clinic: a review of three years prospective experience," *BJOG: An International Journal of Obstetrics & Gynaecology*, vol. 110, pp. 281–286, Mar. 2003.

[68] H. C. Fan, Y. J. Blumenfeld, U. Chitkara, L. Hudgins, and S. R. Quake, "Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing dna from maternal blood," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, pp. 16266–16271, July 2008.

[69] A. N. Imudia, S. Kumar, M. P. Diamond, A. H. DeCherney, and D. R. Armant, "Transcervical retrieval of fetal cells in the practice of modern medicine: A review of the current literature and future direction," *Fertility and sterility*, vol. 93, pp. 1725–1730, Jan. 2010.

[70] K. Sundberg, J. Bang, S. Smidt-Jensen, V. Brocks, C. Lundsteen, J. Parner, N. Keiding, and J. Philip, "Randomised study of risk of fetal loss related to early amniocentesis versus chorionic villus sampling," *The Lancet*, vol. 350, pp. 697–703, Sept. 1997.

[71] J. W. Seeds, "Diagnostic mid trimester amniocentesis: How safe?," *American Journal of Obstetrics and Gynecology*, vol. 191, pp. 607–615, Aug. 2004.

[72] K. H. Nicolaides, "A model for a new pyramid of prenatal care based on the 11 to 13 weeks' assessment," *Prenatal Diagnosis*, vol. 31, no. 1, pp. 3–6, 2011.

[73] "Aium practice guideline for the performance of obstetric ultrasound examinations," *Journal of Ultrasound in Medicine*, vol. 32, no. 6, pp. 1083–1101, 2013.

[74] A. Khalil, M. Rodgers, A. Baschat, A. Bhide, E. Gratacos, K. Hecher, M. D. Kilby, L. Lewi, K. H. Nicolaides, D. Oepkes, N. Raine-Fenning, K. Reed, L. J. Salomon, A. Sotiriadis, B. Thilaganathan, and Y. Ville, "Isuog practice guidelines: role of ultrasound in twin pregnancy," *Ultrasound in Obstetrics & Gynecology*, vol. 47, no. 2, pp. 247–263, 2016.

[75] L. A. Bartlett, C. J. Berg, H. B. Shulman, S. B. Zane, C. A. Green, S. Whitehead, and H. K. Atrash, "Risk factors for legal induced abortion–related mortality in the united states," *Obstetrics & gynecology*, vol. 103, no. 4, pp. 729–737, 2004.

[76] D. Iliescu, M. Cara, S. Tudorache, P. Antsaklis, I. Ceausu, D. Paulescu, L. Novac, N. Cernea, and A. Antsaklis, "Challenges in sonographic detection of fetal major structural abnormalities at the first trimester anomaly scan," vol. 9, pp. 239–259, 01 2015.

[77] S. Zane, A. A. Creanga, C. J. Berg, K. Pazol, D. B. Suchdev, D. J. Jamieson, and W. M. Callaghan, "Abortion-related mortality in the united states: 1998-2010.," *Obstetrics and gynecology*, vol. 126, no. 2, pp. 258–265, 2015.

[78] B. Kott and T. J. Dubinsky, "Cost-effectiveness model for first-trimester versus second-trimester ultrasound screening for down syndrome," *Journal of the American College of Radiology*, vol. 1, no. 6, pp. 415 – 421, 2004.

[79] A. M. Vintzileos, C. V. Ananth, A. J. Fisher, J. C. Smulian, D. Day-Salvatore, and T. Beazoglou, "An economic evaluation of first-trimester genetic sonography for prenatal detection of down syndrome," *Obstetrics & Gynecology*, vol. 91, no. 4, pp. 535–539, 1998.

[80] R. Snijders, P. Noble, N. Sebire, A. Souka, and K. Nicolaides, "Uk multicentre project on assessment of risk of trisomy 21 by maternal age and fetal nuchal-translucency thickness at 10–14 weeks of gestation," *The Lancet*, vol. 352, no. 9125, pp. 343 – 346, 1998.

[81] F. D. Malone, J. A. Canick, R. H. Ball, D. A. Nyberg, C. H. Comstock, R. Bukowski, R. L. Berkowitz, S. J. Gross, L. Dugoff, S. D. Craigo, I. E. Timor-Tritsch, S. R. Carr, H. M. Wolfe, K. Dukes, D. W. Bianchi, A. R. Rudnicka, A. K. Hackshaw, G. Lambert-Messerlian, N. J. Wald, and M. E. D'Alton, "First-trimester or second-trimester screening, or both, for down's syndrome," *New England Journal of Medicine*, vol. 353, no. 19, pp. 2001–2011, 2005. PMID: 16282175.

[82] R. Wapner, E. Thom, J. L. Simpson, E. Pergament, R. Silver, K. Filkins, L. Platt, M. Mahoney, A. Johnson, W. A. Hogge, R. D. Wilson, P. Mohide, D. Hershey, D. Krantz, J. Zachary, R. Snijders, N. Greene, R. Sabbagha, S. MacGregor, L. Hill, A. Gagnon, T. Hallahan, and L. Jackson, "First-trimester screening for trisomies 21 and 18," *New England Journal of Medicine*, vol. 349, no. 15, pp. 1405–1413, 2003. PMID: 14534333.

[83] K. H. Nicolaides, "Screening for fetal aneuploidies at 11 to 13 weeks," *Prenatal Diagnosis*, vol. 31, no. 1, pp. 7–15, 2011.

[84] R. Lachmann, R. Chaoui, J. Moratalla, G. Picciarelli, and K. H. Nicolaides, "Posterior brain in fetuses with open spina bifida at 11 to 13 weeks," *Prenatal Diagnosis*, vol. 31, no. 1, pp. 103–106, 2011.

[85] J.-P. Bernard, H. S. Cuckle, J. J. Stirnemann, L. J. Salomon, and Y. Ville, "Screening for fetal spina bifida by ultrasound examination in the first trimester of pregnancy using fetal biparietal diameter," *American Journal of Obstetrics and Gynecology*, vol. 207, no. 4, pp. 306.e1 – 306.e5, 2012.

[86] A. Khalil, E. Pajkrt, and L. S. Chitty, "Early prenatal diagnosis of skeletal anomalies," *Prenatal Diagnosis*, vol. 31, no. 1, pp. 115–124, 2011.

[87] A. Syngelaki, T. Chelemen, T. Dagklis, L. Allan, and K. H. Nicolaides, "Challenges in the diagnosis of fetal non-chromosomal abnormalities at 11–13 weeks," *Prenatal Diagnosis*, vol. 31, no. 1, pp. 90–102, 2011.

[88] W. Sepulveda, A. E. Wong, P. Martinez-Ten, and J. Perez-Pedregosa, "Retronasal triangle: a sonographic landmark for the screening of cleft palate in the first trimester," *Ultrasound in Obstetrics and Gynecology*, vol. 35, no. 1, pp. 7–13, 2010.

[89] J. N. Karim, N. W. Roberts, L. J. Salomon, and A. T. Papageorghiou, "Systematic review of first-trimester ultrasound screening for detection of fetal structural anomalies and factors that affect screening performance," *Ultrasound in Obstetrics & Gynecology*, vol. 50, no. 4, pp. 429–441, 2017. UOG-2016-0159.R1.

[90] K. O. Kagan, J. Sonek, P. Wagner, and M. Hoopmann, "Principles of first trimester screening in the age of non-invasive prenatal diagnosis: screening for other major defects and pregnancy complications," *Archives of Gynecology and Obstetrics*, vol. 296, pp. 635–643, Oct 2017.

[91] B. O. Verburg, E. A. P. Steegers, M. De Ridder, R. J. M. Snijders, E. Smith, A. Hofman, H. A. Moll, V. W. V. Jaddoe, and J. C. M. Witteman, "New charts for ultrasound dating of pregnancy and assessment of fetal growth: longitudinal data from a population-based cohort study," *Ultrasound in Obstetrics and Gynecology*, vol. 31, no. 4, pp. 388–396, 2008.

[92] G. E. Chalouhi, J. P. Bernard, G. Benoist, B. Nasr, Y. Ville, and L. J. Salomon, "A comparison of first trimester measurements for prediction of delivery date," *The Journal of Maternal-Fetal & Neonatal Medicine*, vol. 24, no. 1, pp. 51–57, 2011.

[93] A. T. Papageorghiou, S. H. Kennedy, L. J. Salomon, E. O. Ohuma, L. Cheikh Ismail, F. C. Barros, A. Lambert, M. Carvalho, Y. A. Jaffer, E. Bertino, M. G. Gravett, D. G. Altman, M. Purwar, J. A. Noble, R. Pang, C. G. Victora, Z. A. Bhutta, J. Villar, for the International Fetal, and N. G. C. for the 21st Century (INTERGROWTH-21st), "International standards for early fetal size and pregnancy dating based on ultrasound measurement of crown–rump length in the first trimester of pregnancy," *Ultrasound in Obstetrics & Gynecology*, vol. 44, no. 6, pp. 641–648, 2014.

[94] L. J. Salomon, C. Pizzi, A. Gasparrini, J.-P. Bernard, and Y. Ville, "Prediction of the date of delivery based on first trimester ultrasound measurements: an independent method from estimated date of conception," *The journal of maternal-fetal & neonatal medicine*, vol. 23, no. 1, pp. 1–9, 2010.

[95] R. Napolitano, J. Dhami, E. Ohuma, C. Ioannou, A. Conde-Agudelo, S. Kennedy, J. Villar, and A. Papageorghiou, "Pregnancy dating by fetal crown–rump length: a systematic review of charts," *BJOG: An International Journal of Obstetrics & Gynaecology*, vol. 121, no. 5, pp. 556–565, 2014.

[96] A. T. Papageorghiou, B. Kemp, W. Stones, E. O. Ohuma, S. H. Kennedy, M. Purwar, L. J. Salomon, D. G. Altman, J. A. Noble, E. Bertino, M. G. Gravett, R. Pang, L. Cheikh Ismail, F. C. Barros, A. Lambert, Y. A. Jaffer, C. G. Victora, Z. A. Bhutta, J. Villar, for the International Fetal, and N. G. C. for the 21st Century (INTERGROWTH-21st), "Ultrasound-based gestational-age estimation in late pregnancy," *Ultrasound in Obstetrics & Gynecology*, vol. 48, no. 6, pp. 719–726, 2016.

[97] P. Loughna, L. Chitty, T. Evans, and T. Chudleigh, "Fetal size and dating: Charts recommended for clinical obstetric practice," *Ultrasound*, vol. 17, pp. 160–166, Aug. 2009.

[98] I. Zalud, S. Good, G. Carneiro, B. Georgescu, K. Aoki, L. Green, F. Shahrestani, and R. Okumura, "Fetal biometry: a comparison between experienced sonographers and automated measurements," *The Journal of Maternal-Fetal & Neonatal Medicine*, vol. 22, no. 1, pp. 43–50, 2009.

[99] J. Espinoza, S. Good, E. Russell, and W. Lee, "Does the use of automated fetal biometry improve clinical work flow efficiency?," *Journal of Ultrasound in Medicine*, vol. 32, no. 5, pp. 847–850, 2013.

[100] M. A. Maraci, R. Napolitano, A. Papageorghiou, and J. A. Noble, "Fisher vector encoding for detecting objects of interest in ultrasound videos," in *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, pp. 651–654, 16-1.

[101] B. Rahmatullah, I. Sarris, A. Papageorghiou, and J. A. Noble, "Quality control of fetal ultrasound images: Detection of abdomen anatomical landmarks using adaboost," in *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 6–9, Marc.

[102] A. Namburete, B. Rahmatullah, and J. Noble, "Nakagami-based adaboost learning framework for detection of anatomical landmarks in 2d detal neurosonograms," vol. 2013, pp. 1–16, 01 2013.

[103] V. Lempitsky, M. Verhoek, J. A. Noble, and A. Blake, *Random Forest Classification for Automatic Delineation of Myocardium in Real-Time 3D Echocardiography*, pp. 447–456. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009.

[104] M. Yaqub, R. Napolitano, C. Ioannou, A. T. Papageorghiou, and J. A. Noble, "Automatic detection of local fetal brain structures in ultrasound images," in *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)*, pp. 1555–1558, 2-5.

[105] M. Yaqub, M. K. Javaid, C. Cooper, and J. A. Noble, "Investigation of the role of feature selection and weighted voting in random forests for 3-d volumetric segmentation," *IEEE Transactions on Medical Imaging*, vol. 33, no. 2, pp. 258–271, Feb.

[106] J. Park, M. Sofka, S. Lee, D. Kim, and S. K. Zhou, "Automatic nuchal translucency measurement from ultrasonography," *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, vol. 16, no. Pt 3, p. 243—250, 2013.

[107] A. I. Namburete, R. V. Stebbing, B. Kemp, M. Yaqub, A. T. Papageorghiou, and J. A. Noble, "Learning-based prediction of gestational age from ultrasound images of the fetal brain," *Medical Image Analysis*, vol. 21, no. 1, pp. 72 – 86, 2015.

[108] G. Carneiro, B. Georgescu, S. Good, and D. Comaniciu, "Detection and measurement of fetal anatomies from ultrasound images using a constrained probabilistic boosting tree," *IEEE Transactions on Medical Imaging*, vol. 27, no. 9, pp. 1342–1355, Sept.

[109] H.-C. Chen, P.-Y. Tsai, H.-H. Huang, H.-H. Shih, Y.-Y. Wang, C.-H. Chang, and Y.-N. Sun, "Registration-based segmentation of three-dimensional ultrasound images for quantitative measurement of fetal craniofacial structure," *Ultrasound in Medicine & Biology*, vol. 38, no. 5, pp. 811 – 823, 2012.

[110] L. Zhang, S. Chen, C. T. Chin, T. Wang, and S. Li, "Intelligent scanning: Automated standard plane selection and biometric measurement of early gestational sac in routine ultrasound examination," *Medical Physics*, vol. 39, no. 8, pp. 5015–5027, 2012.

[111] M. Ahmed and J. A. Noble, "An eye-tracking inspired method for standardised plane extraction from fetal abdominal ultrasound volumes," in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pp. 1084–1087, 13-1.

[112] S. Nie, J. Yu, P. Chen, Y. Wang, and J. Q. Zhang, "Automatic detection of standard sagittal plane in the first trimester of pregnancy using 3-d ultrasound data," *Ultrasound in Medicine & Biology*, vol. 43, no. 1, pp. 286 – 300, 2017.

[113] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, pp. 273–297, Sep 1995.

[114] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119 – 139, 1997.

[115] A. Criminisi, J. Shotton, E. Konukoglu, *et al.*, "Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning," *Foundations and Trends® in Computer Graphics and Vision*, vol. 7, no. 2–3, pp. 81–227, 2012.

[116] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[117] M. A. Maraci, R. Napolitano, A. Papageorghiou, and J. A. Noble, *Object Classification in an Ultrasound Video Using LP-SIFT Features*, pp. 71–81. Cham: Springer International Publishing, 2014.

[118] F. Perronnin, J. Sánchez, and T. Mensink, *Improving the Fisher Kernel for Large-Scale Image Classification*, pp. 143–156. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010.

[119] Y. Deng, Y. Wang, and P. Chen, "Automated detection of fetal nuchal translucency based on hierarchical structural model," in *2010 IEEE 23rd International Symposium on Computer-Based Medical Systems (CBMS)*, pp. 78–84, 12-1.

[120] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 1, pp. I–511–I–518 vol.1, 2001.

[121] P. Dollár and C. L. Zitnick, "Fast edge detection using structured forests," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 8, pp. 1558–1570, 2015.

[122] H. Ryou, M. Yaqub, A. Cavallaro, F. Roseman, A. Papageorghiou, and J. A. Noble, "Automated 3d ultrasound biometry planes extraction for first trimester fetal assessment," in *International Workshop on Machine Learning in Medical Imaging*, pp. 196–204, Springer, 2016.

[123] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biological Cybernetics*, vol. 36, pp. 193–202, Apr 1980.

[124] S. C. B. Lo, S. L. A. Lou, J.-S. Lin, M. T. Freedman, M. V. Chien, and S. K. Mun, "Artificial convolution neural network techniques and applications for lung nodule

detection," *IEEE Transactions on Medical Imaging*, vol. 14, no. 4, pp. 711–718, Dec.

[125] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov.

[126] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, pp. 1097–1105, 2012.

[127] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, *et al.*, "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.

[128] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 2, pp. 1150–1157 vol.2, 1999.

[129] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, pp. 886–893 vol. 1, 25-2.

[130] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain tumor segmentation using convolutional neural networks in mri images," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1240–1251, May.

[131] A. Gupta, M. S. Ayhan, and A. S. Maida, "Natural image bases to represent neuroimaging data," in *Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28*, ICML'13, pp. III–987–III–994, JMLR.org, 2013.

[132] Y. Gao, M. A. Maraci, and J. A. Noble, "Describing ultrasound video content using deep convolutional neural networks," in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pp. 787–790, 13-1.

[133] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional neural networks for medical image analysis: Full training or fine tuning?," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1299–1312, May.

[134] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[135] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 06 2016.

[136] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431–3440, 2015.

[137] H. Greenspan, B. van Ginneken, and R. M. Summers, "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1153–1159, May.

[138] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annual review of biomedical engineering*, vol. 19, pp. 221–248, Mar. 2017.

[139] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, no. Supplement C, pp. 60 – 88, 2017.

[140] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Jan. 2015.

[141] H. Chen, Q. Dou, D. Ni, J.-Z. Cheng, J. Qin, S. Li, and P.-A. Heng, *Automatic Fetal Ultrasound Standard Plane Detection Using Knowledge Transferred Recurrent Neural Networks*, pp. 507–514. Cham: Springer International Publishing, 2015.

[142] A. Kumar, P. Sridar, A. Quinton, R. K. Kumar, D. Feng, R. Nanan, and J. Kim, "Plane identification in fetal ultrasound images using saliency maps and convolutional neural networks," in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pp. 791–794, 13-1.

[143] C. F. Baumgartner, K. Kamnitsas, J. Matthew, S. Smith, B. Kainz, and D. Rueckert, *Real-Time Standard Scan Plane Detection and Localisation in Fetal Ultrasound Using Fully Convolutional Neural Networks*, pp. 203–211. Cham: Springer International Publishing, 2016.

[144] J. Jang, J.-Y. Kwon, B. Kim, S. Min Lee, Y. Park, and J. Seo, "Cnn-based estimation of abdominal circumference from ultrasound images," 02 2017.

[145] Y. Gao and J. Alison Noble, *Detection and Characterization of the Fetal Heartbeat in Free-hand Ultrasound Sweeps with Weakly-supervised Two-streams Convolutional Networks*, pp. 305–313. Cham: Springer International Publishing, 2017.

[146] M. Yaqub, B. Kelly, A. T. Papageorghiou, and J. A. Noble, "A deep learning solution for automatic fetal neurosonographic diagnostic plane verification using clinical standard constraints," *Ultrasound in Medicine & Biology*, vol. 43, no. 12, pp. 2925 – 2933, 2017.

[147] V. Sundaresan, C. P. Bridge, C. Ioannou, and J. A. Noble, "Automated characterization of the fetal heart in ultrasound images using fully convolutional neural networks," in *Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on*, pp. 671–674, IEEE, 2017.

[148] L. Wu, Y. Xin, S. Li, T. Wang, P.-A. Heng, and D. Ni, "Cascaded fully convolutional networks for automatic prenatal ultrasound image segmentation," in *Biomedical*

Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on, pp. 663–666, IEEE, 2017.

[149] A. Schmidt-Richberg, T. Brosch, N. Schadewaldt, T. Klinder, A. Cavallaro, I. Salim, D. Roundhill, A. Papageorghiou, and C. Lorenz, "Abdomen segmentation in 3d fetal ultrasound using cnn-powered deformable models," in *Fetal, Infant and Ophthalmic Medical Image Analysis*, pp. 52–61, Springer, 2017.

[150] C. Raynaud, C. Ciofolo-Veit, T. Lefèvre, R. Ardon, A. Cavallaro, I. Salim, A. Papageorghiou, and L. Rouet, *Multi-organ Detection in 3D Fetal Ultrasound with Machine Learning*, pp. 62–72. Cham: Springer International Publishing, 2017.

[151] z. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3d u-net: Learning dense volumetric segmentation from sparse annotation," 06 2016.

[152] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," 10 2016.

[153] X. Yang, L. Yu, S. Li, X. Wang, N. Wang, J. Qin, D. Ni, and P.-A. Heng, *Towards Automatic Semantic Segmentation in Volumetric Ultrasound*, pp. 711–719. Cham: Springer International Publishing, 2017.

[154] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 6, pp. 679–698, Nov.

[155] M. Nikolic, E. Tuba, and M. Tuba, "Edge detection in medical ultrasound images using adjusted canny edge detection algorithm," in *2016 24th Telecommunications Forum (TELFOR)*, pp. 1–4, 22-2.

[156] S. Tsantis, S. Spiliopoulos, A. Skouroliakou, D. Karnabatidis, J. D. Hazle, and G. C. Kagadis, "Multiresolution edge detection using enhanced fuzzy c-means clustering for ultrasound image speckle reduction," *Med. Phys.*, vol. 41, p. 072903, July 2014.

[157] A. Zaim, "An edge-based approach for segmentation of prostate ultrasouind images using phase symmetry," in *2008 3rd International Symposium on Communications, Control and Signal Processing*, pp. 10–13, 12-1.

[158] A. Hafiane, P. Vieyres, and A. Delbos, "Phase-based probabilistic active contour for nerve detection in ultrasound images for regional anesthesia," *Computers in Biology and Medicine*, vol. 52, pp. 88–95, Sept. 2014.

[159] J. C. Bezdek, J. Keller, R. Krisnapuram, and N. Pal, *Fuzzy models and algorithms for pattern recognition and image processing*, vol. 4. Springer Science & Business Media, 1999.

[160] P. Kovesi, "Image features from phase congruency," *Videre: Journal of computer vision research*, vol. 1, no. 3, pp. 1–26, 1999.

[161] C. Xu and J. L. Prince, "Snakes, shapes, and gradient vector flow," *IEEE Transactions on image processing*, vol. 7, no. 3, pp. 359–369, 1998.

[162] H. Ling, S. K. Zhou, Y. Zheng, B. Georgescu, M. Suehling, and D. Comaniciu, "Hierarchical, learning-based automatic liver segmentation," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, 23-2.

[163] H. Chang, Z. Chen, Q. Huang, J. Shi, and X. Li, "Graph-based learning for segmentation of 3d ultrasound images," *Neurocomputing*, vol. 151, pp. 632–644, Mar. 2015.

[164] D. Ni, Y. Yang, S. Li, J. Qin, S. Ouyang, T. Wang, and P. A. Heng, "Learning based automatic head detection and measurement from fetal ultrasound images via prior knowledge and imaging parameters," in *2013 IEEE 10th International Symposium on Biomedical Imaging*, pp. 772–775, 7-11.

[165] C. L. Zitnick and P. Dollár, "Edge boxes: Locating object proposals from edges," in *European Conference on Computer Vision*, pp. 391–405, Springer, 2014.

[166] J. J. Lim, C. L. Zitnick, and P. Dollár, "Sketch tokens: A learned mid-level representation for contour and object detection," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3158–3165, 23-2.

[167] B. Rahmatullah, A. T. Papageorghiou, and J. A. Noble, "Integration of local and global features for anatomical object detection in ultrasound," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 402–409, Springer, 2012.

[168] M. Everingham, L. V. Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," vol. 88, pp. 303–338, 2010.

[169] D. Scherer, A. Müller, and S. Behnke, "Evaluation of pooling operations in convolutional architectures for object recognition," in *Artificial Neural Networks - ICANN 2010* (K. Diamantaras, W. Duch, and L. S. Iliadis, eds.), (Berlin, Heidelberg), pp. 92–101, Springer Berlin Heidelberg, 2010.

[170] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ICML'10, (USA), pp. 807–814, Omnipress, 2010.

[171] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014.

[172] A. Vedaldi and K. Lenc, "Matconvnet: convolutional neural networks for matlab," 10 2015.

[173] S. Hoo-Chang, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning," *IEEE transactions on medical imaging*, vol. 35, pp. 1285–1298, Feb. 2016.

[174] L. N. Smith, "A disciplined approach to neural network hyper-parameters: Part 1–learning rate, batch size, momentum, and weight decay," *arXiv preprint arXiv:1803.09820*, 2018.

[175] F. Milletari, S.-A. Ahmadi, C. Kroll, A. Plate, V. Rozanski, J. Maiostre, J. Levin, O. Dietrich, B. Ertl-Wagner, K. Bötzel, and N. Navab, "Hough-cnn: Deep learning for segmentation of deep brain regions in mri and ultrasound," *Computer Vision and Image Understanding*, vol. 164, pp. 92–102, Nov. 2017.

[176] T. DeVries and G. W. Taylor, "Dataset augmentation in feature space," *arXiv preprint arXiv:1702.05538*, 2017.

[177] L. Perez and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," *arXiv preprint arXiv:1712.04621*, 2017.

[178] S. C. Wong, A. Gatt, V. Stamatescu, and M. D. McDonnell, "Understanding data augmentation for classification: when to warp?," *arXiv preprint arXiv:1609.08764*, 2016.

[179] A. Ali, S. M. Shamsuddin, and A. L. Ralescu, "Classification with class imbalance problem: a review," *Int J Adv Soft Comput Appl*, vol. 7, no. 3, pp. 176–204, 2015.

[180] A. Singh and A. Purohit, "A survey on methods for solving data imbalance problem for classification," *Work*, vol. 127, no. 15, 2015.

[181] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International Conference on Machine Learning*, pp. 448–456, 2015.

[182] A. Telea and J. J. van Wijk, "An augmented fast marching method for computing skeletons and centerlines," in *Proceedings of the Symposium on Data Visualisation 2002*, VISSYM '02, (Aire-la-Ville, Switzerland, Switzerland), pp. 251–ff, Eurographics Association, 2002.

[183] A. Fitzgibbon, M. Pilu, and R. B. Fisher, "Direct least square fitting of ellipses," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 5, pp. 476–480, May.

[184] H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation," in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, pp. 1520–1528, Dec. 2015.

[185] C. F. Baumgartner, K. Kamnitsas, J. Matthew, T. P. Fletcher, S. Smith, L. M. Koch, B. Kainz, and D. Rueckert, "Sononet: Real-time detection and localisation of fetal standard scan planes in freehand ultrasound," *IEEE Transactions on Medical Imaging*, vol. 36, pp. 2204–2215, Nov. 2017.

[186] M. Yaqub, A. Kopuri, S. Rueda, P. B. Sullivan, K. McCormick, and J. A. Noble, *A Constrained Regression Forests Solution to 3D Fetal Ultrasound Plane Localization for Longitudinal Analysis of Brain Growth and Maturation*, pp. 109–116. Cham: Springer International Publishing, 2014.

[187] P. Christ, F. Ettlinger, F. Grün, M. Ezzeldin A. Elshaera, J. Lipkova, S. Schlecht, F. Ahmaddy, S. Tatavarty, M. Bickel, P. Bilic, M. Rempfler, F. Hofmann, M. D'Anastasi, S.-A. Ahmadi, G. Kaissis, J. Holch, W. Sommer, R. Braren, V. Heinemann, and B. Menze, "Automatic liver and tumor segmentation of ct and mri volumes using cascaded fully convolutional neural networks," 02 2017.