Fetal Brain Ultrasound Analysis: Sub-Region Segmentation of the Cerebellum and Texture Variability in the Thalamus

Zhuyi Lu Mphil

A thesis submitted in fulfillment
of the requirements of the degree of
Master of Philosophy

THE UNIVERSITY OF
SYDNEY

School of Computer Science
Faculty of Engineering
University of Sydney

2025

Student Plagiarism: Compliance Statement

I certify that:

This thesis is my original work and has not been submitted for any other degree or purpose. Any parts that are not original have been properly acknowledged and cited to indicate their sources.

During the preparation of this thesis, ChatGPT was used for the purposes of text enhancement, including sentence structure and paraphrasing for better readability. Where any text was modified by generative AI, the author then reviewed the resulting content for any errors, inaccuracies or biases, and modified it as required. The author takes full responsibility for the submitted thesis and ensures the work is their own and has used generative AI within the parameters of use, see University of Sydney generative AI guide for researchers.

I understand that any failure to comply with the Student Plagiarism: Academic Board Policy on Academic Dishonesty and Plagiarism may result in the University commencing proceedings against me for potential student misconduct under the Thesis and Examination of Higher Degrees by Research Procedures 2020.

Zhuyi Lu Signature: 23 March 2025

Abstract

Ultrasound (US) imaging is routinely used during pregnancy because it is affordable, non-invasive and non-ionizing. Different scans are taken during different gestation stages for various purposes. It is scientifically proven that between 18 and 22 weeks of gestation, the fetal brain undergoes significant structural growth and development. This provides an opportunity to extract rich information from fetal brain US images taken during this period, including size, shape, and structural characteristics. These information can provide valuable insights into fetal brain growth. During gestation, all the fetal structures undergo rapid growth but, in particular, thalamus, the central relay of the brain, experiences the most dynamic changes compared to other brain structures. There have been attempts at studying the fetal brain growth by analyzing volume development or biometric measurements. However, existing studies on fetal US images solely focus on investigating the size(e.g., head circumferences, biparietal diameter) and shape characteristics (e.g., morphological changes) of brain structures to assess their clinical values. Nonetheless, the possible clinical value of radiomics, which includes texture features, in relation to fetal growth has been failed to examine. Texture-based analysis is important as it can provide additional information such as tissue microstructure or the differentiation between gray and white matters, providing a more comprehensive aspect of growth and this cannot be captured by simple size or shape measurement.

To address this gap, the first study in this thesis applied radiomics-based feature extraction methods, combining deep learning-based and conventional statistical texture extraction methods to extract the texture features of the thalamus from the fetal brain US images, then investigated the relationship with gestational age at birth. The results indicate that several texture feature characteristics show statistically significant variations across different gestational age groups at birth, which are 26.4-36 weeks, 36-39 weeks, 39-42.4 weeks. Additionally, texture feature characteristics demonstrate Pearson correlation coefficients of 0.2 - 0.3, including a weak to moderate correlation with the gestational age. Comparative analysis further highlights that deep features can capture more subtle texture details than statistical features. This study suggests the feasibility of using texture feature as a novel approach to investigate the fetal brain growth, which is beneficial to future investigation of fetal neurodevelopment or

gestational age estimation.

Accurate and robust region-of-interest (ROI) segmentation is a prerequisite for reliable texture analysis. However, due to the noisy nature of US images and lack of specific model for US image sub-region segmentation, accurate segmentation of sub-regions like cerebellum in the US fetal brain images remains challenging. To address this, this study introduces a novel Adaptive Multi-Fusion Attention (AMFA) Network, incorporating a novel adaptive attention gate and multi-fusion mechanism to enhance segmentation accuracy. Results show that the proposed AMFA network outperformed other state-of-art segmentation models in segmenting the cerebellum from the fetal brain US images. It also performed well in segmenting the breast tumor from the external public Breast Ultrasound Images(BUSI) dataset compared to other models. These results prove that the proposed segmentation model works well not only on sub-region segmentation but also larger and normal general regions.

In short, these two studies work together as a process where accurate sub-region segmentation can enable more reliable texture analysis, thus improving fetal brain analysis from US images.

Acknowledgements

I am grateful to my main supervisor, Prof. Jinman Kim, for his invaluable guidance and support throughout my whole research journey. His expertise in computer science and great experience in clinical research have been immensely helpful, providing me with clarity and direction in this research.

I also want to express my sincere gratitude to my secondary supervisor Prof. Ralph Nanan, for his valuable insights and professional expertise in the clinical field, which have greatly enriched my research perspective.

Moreover, I would like to thank Dr. Euijoon Ahn for his professional guidance in the computer science area as well as his insightful advice and constructive feedback.

Additionally, I would like extend my appreciation to Duan Ni for his support in the statistical area and guidance as for the quantitative aspects of my study. And I want to thank Yuan Yuan and Boyuan Tan for their help and support throughout my Mphil degree.

I am also grateful to the ultrasound research group, including clinicians and ultrasonographers, for their collaboration and support in providing domain knowledge and data-related assistance.

Finally, I want to thank the entire BDAV group for their collaborative spirit, valuable discussions, and continuous support throughout my whole research journey.

Contents

List of Figures

List of Tables

Authorship Attribution Statement

Chapter 4 of this thesis is under preparation for submission to Journal of Biomedical and Health Informatics (JBHI) as

> Z.Lu, E.Ahn, R.Nanan, NJ.Kennedy, A.Quinton, J.Kim, "AMFA: An Adaptive Multi-Fusion Attention Network for Robust Sub-Region Segmentation in Ultrasound Images"

I was the primary contributor to the design and the implementation of the study, including the development of the segmentation model, data preprocessing, and the overall experimental framework. Additionally, I was responsible for writing the manuscript.

In addition to statements above, in cases where I am not the corresponding author of a published item, permission to include the published material has been granted by the corresponding author.

Zhuyi Lu Signature: 23 March 2025

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

Professor Jinman Kim Signature: 23 March 2025

Chapter 1

Introduction

1.1 Introduction and Motivations

1.1.1 US Texture Analysis

Ultrasound (US) imaging is widely used during pregnancy for fetal monitoring and anomaly detection because it is non-invasive, safe and affordable. As a routine prenatal screening tool, it allows for a detailed examination of different fetal structures such as spine, heart, head and brain. For example, according to Kinman (2018), fetal echocardiogram is taken during 18-24 weeks gestational age to evaluate the baby's heart to check possible congenital heart disease (CHD) if needed. In one of the study from International Society of Ultrasound in Obstetrics and Gynecology (2022), it showed that biometric US scans are usually taken to calculate the the head circumference (HC) to assess the fetal growth. According to Marion et al. (2023), for fetal brain, cranial scans are performed to check possible hydrocephalus and Bekele et al. (2024) proved transcerebellar scans are taken during mid-trimester(18-22 weeks) to assess cerebellar development.

A lot of the scans mentioned above are performed during mid-trimester, typically between 18-22 weeks of gestational age. In this stage, the fetus undergoes a rapid growth and significant developmental transformations, particularly in the fetal brain. In an

MRI-based study, Schumacher et al. (2020) found out that during this period of time, fetal brain shows triple volume in growth, forms plenty of new synaptic connections, and begins axonal myelination. Given these rapid neurological changes have been observed in MRI studies — which can visualize detailed brain anatomy and structural development, it is reasonable to expect that ultrasound (US), although more prone to noise and artifacts, still has the potential to capture important indicators of fetal brain growth and development. Many studies have increasingly focused on assessing fetal growth on US modality. For example, Sridar et al. (2020b) investigated on 1111 transcerebellar scans and made the conclusion that transverse thalamic diameter showed a significant linear increase with the growing gestational age, head circumference, and transcerebellar diameter. In another study, Lee et al. (2013) proposed the possibility of using transverse cerebellar diameter to predict fetal gestational age with asymmetric intrauterine growth retardation. These studies shows the potential of US brain imaging in investigating fetal development and gestational assessment.

Notably, existing studies that utilize US brain imaging to assess fetal growth and development primarily focused on measuring the size, shape or volume of brain structures. However, many studies implied that brain growth does not just show variations in size but also in brain structures and different tissue textures. As mentioned above, Schumacher et al. (2020) found that fetal brain shows triple volume in growth during 18-22 weeks of gestation. In another study, Scott et al. (2012) demonstrated that during 20 - 40 weeks of gestation, the fetal brain undergoes rapid growth with cortical folding and volume changes. These MRI-based studies have focused on analysing macroscopic structural changes such as brain volume growth or cortical development. However, there's a lack of studies using radiomics-based features such as structural or textual feature to investigate fetal brain growth in US images. This might be because the sub-regions of the fetal brain US image like thalamus (Figure 1.1) and cerebellum (Figure 1.2) are usually very small with diameter only in few pixels, those textural changes are extremely hard to notice by human eyes.

With the development of deep learning recently, textural features can now be extracted from the US images. Compared to normal statistical texture features, deep

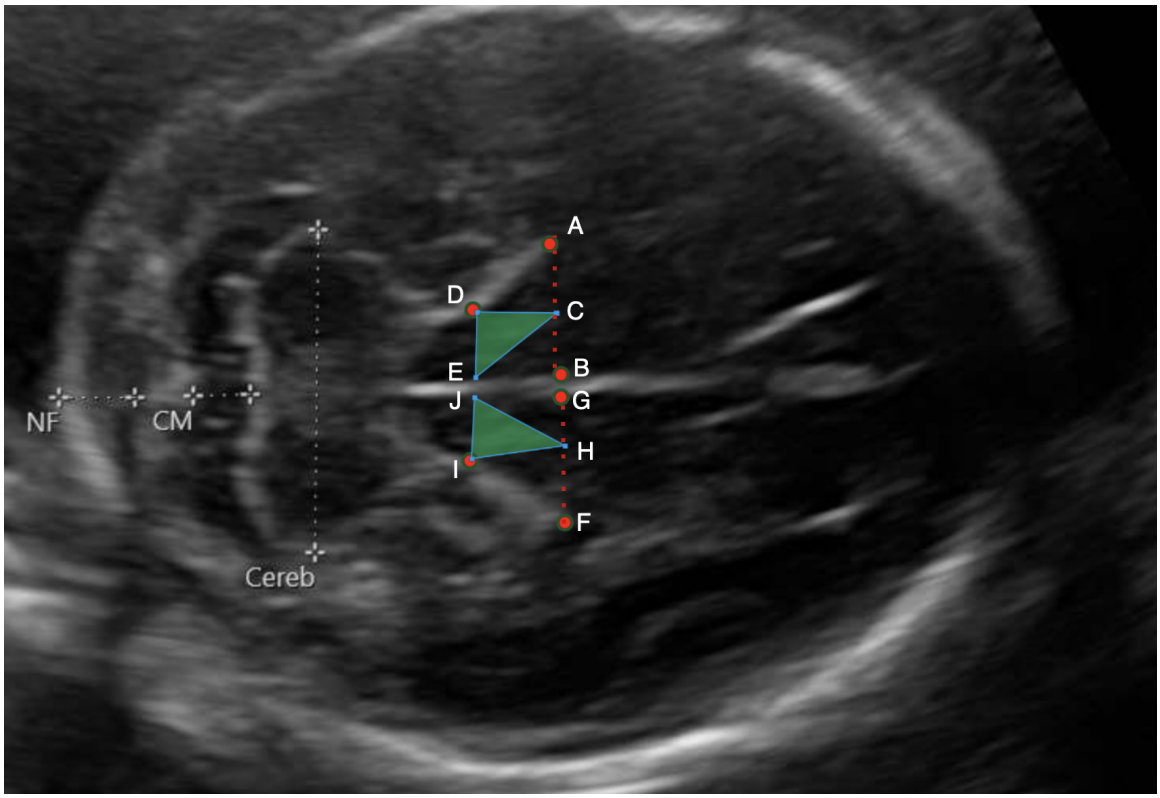


Figure 1.1 – Thalamus region example of the fetal brain US scan

features contains more complex information with higher-order spatial relationship and deeper structural patterns between pixels, enabling a more comprehensive image characteristics representations (Ming et al. (2018)). Many existing studies used deep learning-based methods to extract deep features for following investigation on multiple medical image modalities. Kriti et al. (2020) used pretrained VGG-19, SqueezeNet, etc to extract deep features from breast tumor US images and they showed good classification outcome in the following classification task. Cruz-Roa et al. (2015) used an enhanced VGG model to extract deep features from histopathology for medulloblastoma tumor classification. In another study, Amin et al. (2019) utilized AlexNet and GoogleNet to extract features then used for downstream brain tumor classification. Moreover, Lai and Deng (2018) trained a deep convolutional neural network to extract high-level features from raw medical images then for downstream classification. They conducted experiments on the HIS2828 Dataset for histopathological tumor classification and ISIC2017 Dataset for skin lesion classification. These studies prove

the feasibility and the advantages of using deep features for medical image analysis. Ming et al. (2018) compared the classification performance of GLCM-extracted texture features and deep features, showing that deep features provide more accurate classification results.

The potential of 18-22 weeks of gestational age fetal brain US images to contain rich developmental information, combined with the strengths of existing statistical and deep learning-based methods for texture feature extraction and the current lack of studies in this area, motivate us to use these extracted texture features from the fetal brain US images in relation to fetal growth. The first study of this thesis chose a sub-region: thalamus from the 18-22 weeks gestational age transcerebellar scan, extracting the textural features from it and use the quantified textural feature to investigate the possible texture variations across different gestational age groups. More details of methodology and experimental setup are discussed in Section 3.2 and Section 3.3.

1.1.2 US Imaging Segmentation

US images are noisy and prone to artifacts, which makes segmentation harder, especially for the sub-region segmentation, where the boundaries are less clear and more susceptible to interference. Currently, according to the literature review, there's no segmentation model that is optimized and evaluated for their performance in US fetal brain sub-region segmentation. Existing segmentation models like U-Net (Ronneberger et al. (2015)) - an encoder-decoder convolutional neural network for medical image segmentation and its enhanced models like U-Net++ (Zhou and Siddiquee (2018)) have been shown to provide accurate segmentation outcome in segmenting larger, more general regions such as fetal head. However, based on the experiments in this study, these methods struggle to provide reliable segmentation for sub-regions like cerebellum from the fetal brain US image, more details in Section 4.5. This motivates us to develop an enhanced model that is optimized for US sub-region segmentation.

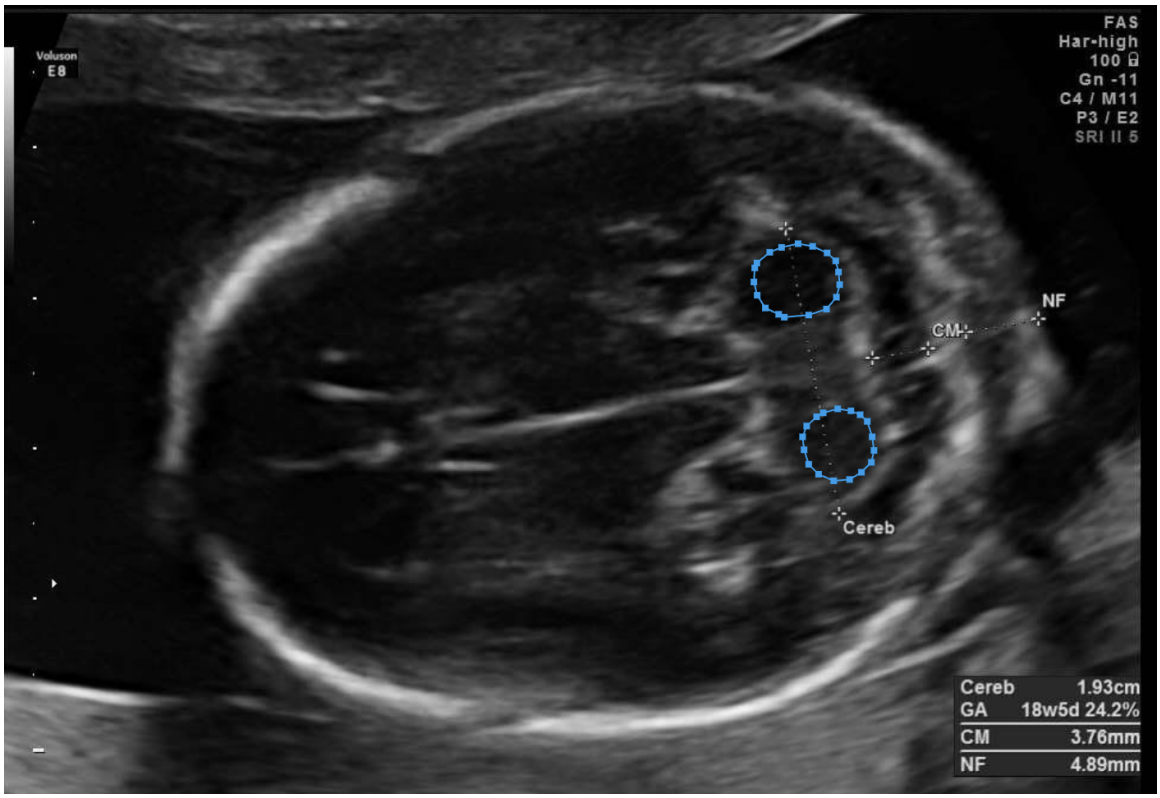


Figure 1.2 – Cerebellum region example of the fetal brain US scan

1.2 Research Aims and Contributions

The aims of the thesis are divided into the following two parts. (i) Leverage deep learning and statistical methods to extract thalamus texture features from 18-22 weeks of fetal brain US images and assess their possible variations across different gestational groups at birth. (ii) Develop an enhanced segmentation model specifically for US sub-region segmentation and prove its feasibility in the 18-22 weeks fetal brain US images.

For (i), a radiomics-based approach was applied, which refers to a class of work that extracts high-dimensional quantitative image features to capture subtle and meaningful patterns that not visually apparent. Deep learning and conventional statistical techniques were used to extract the texture features from the thalamus sub-region of the 18-22 weeks of gestational age then the extracted features were used to investigate their relationship and possible texture variations in three different gestational

age groups at birth. Here, gestational age groups at birth refer to the total gestational age of the fetus when delivery in stead of the gestational age when the US images were taken. These groups (26.4-36 weeks, 36-39 weeks, and 39-42.4 weeks) were defined based on the dataset distribution for a balanced sample size for each group. The results show that a notable number of texture (statistical and deep) feature characteristics show significant difference across the three gestational age groups, despite that all the scans were taken in the same gestational window (18-22 weeks gestational age). This emphasizes that early-stage thalamus texture features reflects differences in relation to gestational age at birth, suggesting their potential indicative value in fetal development.

For (ii), Adaptive Multi-Fusion Attention (AMFA) network was proposed and it was built specifically for US imaging sub-region segmentation. By incorporating adaptive attention gate and multi-fusion mechanism, the AMFA network is expected to obtain more accurate attention weights and capture fine-grained feature maps, thus addressing the limitation of current models' incapability of segmenting sub-regions. To assess its effectiveness, tests were conducted on the cerebellum on the 18-22 weeks fetal brain US images. The results demonstrate superior performance when segmenting the cerebellum from the 18-22 weeks fetal brain US images compared to other state-of-art models. To evaluate the model's generalizability, the AMFA network was further evaluated on the public Breast Ultrasound Image (BUSI) dataset. The results show that AFMA also performs well when segmenting breast tumors and this indicates its robustness and generalizability when segmenting larger and more general regions in the US images.

The two studies in this thesis are complementary to each other. The enhanced sub-region segmentation model shows the potential to be applied to other sub-region segmentation tasks such as the thalamus (Figure 1.1) region as mentioned in the first study of this thesis. Replacing the manual thalamus segmentation step in the texture analysis with the accurate automatic thalamus sub-region segmentation can makes the whole texture analysis pipeline automatic, thus providing a seamless and efficient workflow. Furthermore, in future work, radiomics-based texture analysis

will not be limited to thalamus but also applied to other brain sub-regions such as cerebellum and brainstem. By leveraging the accurate segmentation generated by the proposed sub-region segmentation model, this approach will lay a solid foundation for radiomics-based texture features across multiple brain sub-regions, enabling a more comprehensive investigation between fetal brain growth during gestation.

1.3 Outline of the Thesis

The remainder of the thesis is structured as follows. Chapter 1 introduces the background of US texture analysis and US image segmentation. Research Aims and Contributions are included in this chapter as well. Then, chapter 2 provides a comprehensive literature review for both studies, covering the recent progress of deep learning-based segmentation models for medical images. A summary of segmentation tasks in fetal brain US images are provided as well. This chapter also discusses the recent progress of texture analysis on medical images for a better understanding of its background. Clinical related work, including 2D US scans on fetal brain and progress on the fetal brain US image analysis are also included.

Chapter 3 starts with an introduction and background of the texture analysis study, followed by introducing employed methods, which include dataset descriptions, a pipeline overview and utilized methods in this study. Experimental setup such as implementation details and evaluation metrics are then presented in the following part. Finally, results and discussion for this study are provided. Chapter 4 follows a similar structure to Chapter 3 but includes a more detailed literature review about state-of-art deep learning-based segmentation models. Although segmentation usually precedes before texture analysis in medical imaging workflow, in this thesis the texture analysis is presented first, as the motivation for segmentation arose directly from challenges identified in the texture analysis study.

Chapter 2

Related Work

2.1 Medical Image Segmentation

2.1.1 Recent Progress of Deep Learning-based Segmentation Model for Medical Images

Table 2.1 summarizes the recent progress of deep learning-based segmentation models for medical images. Deep learning-based segmentation models starts with Fully Convolutional Network (FCN), proposed by Long et al. (2014). Compared to conventional classification model such as VGG and ResNet, FCN replaces fully connected layers with 1 x 1 convolutions to preserve spatial structure. In addition, skip connections are incorporated to the model to allow multi-scale feature fusion, thus improving segmentation performance. Furthermore, FCN employed transposed convolution layers to upsample the feature maps to the original input resolution. The creation of FCN established the foundation for the future deep learning-based segmentation models.

With the emergence of FCN, Ronneberger et al. (2015) proposed U-Net based on the FCN network, specifically designed for biomedical image segmentation. In the U-Net architecture, it retains the idea from FCN of using convolutional and up-convolutional layers to perform segmentation. However, in Ronneberger et al. (2015), it also proposed a new symmetric encoder-decoder architecture to the U-Net architecture for

capturing hierarchical features maps while retaining the feature map resolution. In Ronneberger et al. (2015), U-Net is proved to be effective in conducting segmentation tasks on microscopy images. On the basis of this, researchers start to utilize U-Net on various segmentation tasks on multiple medical image modalities and U-Net shows satisfactory segmentation outcomes as well. For example, Mehta et al. (2019) successfully used 3D U-Net for brain tumor segmentation on MRI images. Zhao et al. (2018) used a patch-based 3D U-Net to segment the lung nodule from CT images.

Then, variants of U-Net such as U-Net++ (Zhou and Siddiquee (2018)) and Attention U-Net (Oktay et al. (2018)) came up, showing more accurate segmentation outcomes by addressing specific limitations of the original U-Net architecture. For example, U-Net++ used nested and dense skip connections to solve the limitation of semantic gap between the encoder and decoder. Attention U-Net used attention gate to mitigate the problem where focused regions and irrelevant regions are given the same importance. Even though models like U-Net++ and Attention U-Net were initially being experimented on CT datasets, plenty of research have also proved their feasible application to the US modality. Eduardo et al. compared the breast tumor segmentation results by using U-Net and U-Net++ respectively and proved that U-Net++ does bring benefits to the segmentation performance. Besides, Chen et al. proposed an enhanced version of Attention U-Net on the breast lesions segmentation in US images.

However, despite the advancements in the variants of U-Net, they still have the following similar limitations. First, convolutional neural network (CNN)-based models like U-Net all highly rely on local convolutions, which indicates their struggle in capturing global spatial relationship in the image (Singh et al. (2025)). However, medical images such as US images often contain long-range dependencies, which makes current models hard to accurately segment the region out. Moreover, medical images are known with significant heterogeneity across different datasets, which makes the ability to capture global features important. Even though variants like Attention U-Net can mitigate this problem to some extent by enhancing feature selection, they still struggles to effectively capture broader and long spatial dependencies, limiting

the segmentation performance. Second, the encoder-decoder architecture also poses a problem as it causes spatial loss. Although skip connections help recover some high-resolution features, they are still insufficient in preserving fine-grained details, especially when the images are complex and vague with noise.

To solve these limitations, Chen et al. proposed TransUNet, which integrates Transformer with the conventional U-Net architecture. Originally introduced in Natural Language Processing (NLP), Transformers leverage self-attention mechanism to process 1D sequential data by capturing relationship between words and this helps model handle long-range dependencies globally. With the help of self-attention from Transformer, TransUNet can effectively capture and process long-range dependencies and global information in the image and the U-Net part can still focusing on the localized spatial information. This combined characteristics of the model can provide more accurate segmentation outcomes. Chen et al. have evaluated TransUNet in segmenting the cardiac from ACDC MRI datasets and multiple organs (liver, kidney, aorta, stomach, etc) from the Synapse CT dataset, demonstrating its superior performance across multiple segmentation tasks.

However, as Transformer process the entire image as a sequence of tokens, it leads to a computational complexity that grows quadratically in regard to the image resolution, thus making it computational expensive as well. In this case, Cao et al. (2021) proposed Swin-Unet, utilizing Swin Transformer instead of ViT-Transformer to apply the shifted window attention to process local regions first then handle global information to reduce computational complexity. Experiments show that Swin-Unet not only outperforms TransUNet but also reduces complexity to linear scale.

After the success of TransUNet and Swin-Unet, researchers continue to explore more efficient architectures for medical image segmentation. Mamba (Gu and Dao (2024)) is an innovative State Space Model (SSM) that initially designed for natural language processing (NLP) to replace self-attention in Transformers to improve long-sequence processing for more efficient sequence modeling. However, due to Mamba's initial plan of handling 1D sequential data flow, it is unable to handle 2D visual data.

To address this limitation, Liu et al. introduced VMamba, an extension of Mamba

Segmentation Model	Paper	Imaging Modality
FCN	Long et al. (2014)	Natural Images
U-Net	Ronneberger et al. (2015)	Microscopy Images
3D U-Net	Mehta et al. (2019)	MRI
Patch-based 3D U-Net	Zhao et al. (2018)	CT
U-Net++	Eduardo et al. (2025); Zhou and Siddiquee (2018)	CT, US
Attention U-Net	Oktay et al. (2018)	CT
AAU-Net	Zhang et al. (2022)	US
TransUNet	Chen et al. (2021)	CT, MRI
Swin-Unet	Cao et al. (2021)	CT, MRI
Mamba U-Net	Wang et al. (2024)	CT, MRI

Table 2.1 – Summary of deep learning based segmentation models for medical images

for visual tasks. 2D Selective Scan (SS2D) is the core innovation of VMamba, which allows the model to capture long-range spatial dependencies within 2D images. Visual State Space (VSS) block is another key component in VMamba. It replaces the original convolutional layers in deep learning models so the model can effectively capture long-range dependencies while maintaining computational efficiency. According to Liu et al., the proposed VSS block is an effective alternative to not only convolutions but also traditional self-attention mechanisms.

Inspired by the success of VMamba, Wang et al. applied the VSS blocks into U-Net architecture and proposed Mamba U-Net, which proved to provide more accurate segmentation result on two medical image dataset: ACDC MRI Cardiac and Synapse CT Abdomen segmentation dataset.

2.1.2 Segmentation Tasks in Fetal Brain US Images

There are quite a few studies focusing on the segmentation tasks in fetal brain US images. Those segmentation tasks can be classified into the following two categories:

Segmentation Region	Paper	Model ArchitectureUsed Model	Image Modality
Fetal Head	Ye et al. (2019)	Enhanced U-Net	2D US
Fetal Head	Liu et al. (2020)	SAFNet	2D US
Cerebellum	Maraci et al. (2021)	FCN	2D US
Cerebellum	Singh et al. (2021)	ResU-Net-c	2D US
Thalamus	Sridar et al. (2021)	Distance regularized level set	2D US

Table 2.2 – Summary of deep learning based segmentation tasks for fetal brain US images

(i) fetal head segmentation and (ii) inner brain structures segmentation.

Fetal brain segmentation refers to those segmentation tasks that segments the whole fetal brain region from the US images while brain inner structures segmentation focuses on segmenting those single brain structures such as cerebellum and thalamus. Contour-based segmentation is one of the common methods in fetal brain segmentation. It isolates out the brain region from surrounding tissues, which benefits the future tasks such as fetal size analysis, brain growth monitoring, etc. With the advancement of deep learning models and techniques, coutour-based segmentation starts to show more reliable segmentation performance. Table 2.2 lists a selection of representative segmentation tasks on fetal brain US images. Ye et al. and Liu et al. utilized an enhanced U-Net model and SAFNet respectively to segment the brain region from the fetal brain US images.

For inner brain structures segmentation, Maraci et al. used FCN to segment the cerebellum from the transcerebellar fetal brain US images and Singh et al. proposed ResU-Net-c to segment the guitar-shape cerebellum from the fetal brain US images. Besides from cerebellum, researchers also try to segment the thalamus. For example, Sridar et al. utilized shape prior constrained regularized level sets to segment the thalamus from the 2D fetal brain US images.

2.2 Texture Analysis

2.2.1 Recent Progress of Texture Analysis on Medical Images

In the last decades, plenty of studies focusing on the texture characteristics from the medical images. In general, texture analysis studies need to have the tools for extracting the texture features from the region of interests, then the extracted features will be used for future texture analysis such as classification, segmentation, etc. This section covers the recent progress of texture analysis as well as the texture extraction tools for medical imaging modalities such as MRI, CT, and US.

Statistical texture extraction methods are among those earliest and most widely used ones. Statistical texture features can also be classified into first-level statistical methods, second-level statistical methods and high-order statistical methods. First-order statistical methods are mainly used for extracting the simplest texture feature such as mean, variance, skewness from an image. Gray-Level Co-occurrence Matrix (GLCM, Haralick et al. (1973)) and Gray-Level Run-Length Matrix (GLRLM, Galloway (1975)) are second-order statistical methods. Gray-Level Size Zone Matrix (GLSZM, Thibault et al. (2013)), Neighborhood Gray-Tone Difference Matrix (NGTDM, Amadasun and King (1989)) and Gray-Level Dependence Matrix (GLDM, Weszka et al. (1976)) belong to higher-order statistical methods. More details are listed and mentioned in Table 2.3 and Section 3.2.4.

Classification	Model
Fist-order statistical methods	Mean, Variance, Skewness, Kurtosis, Energy, Entropy
Second-order statistical methods	GLCM (Haralick et al. (1973)), GLRLM (Galloway (1975))
Higher-order statistical methods	GLSZM (Thibault et al. (2013)), NGTDM (Amadasun and King (1989)), GLDM (Weszka et al. (1976))

Table 2.3 – Classification of statistical texture extraction tools

Plenty of studies have used these statistical texture methods to extract texture features considering the usefulness of these texture feature characteristics. Table 2.4 summarizes a selected set of representative work. Zhang et al. used GLCM and run-length(RL) matrix to extract the texture features from the MRI of Multiple Sclerosis (MS) patients for normal brain tissue and MS lesions classification. Almeida and Santos used extracted GLCM texture features for skin tumor detection in MRI images. By extracting some GLRLM texture features, Daginawala et al. demonstrated the feasibility of using these texture features to distinguish between fibrosis groups. Brusasco et al. shows satisfactory accuracy of utilizing second-order gray-scale texture features on the pleural US images to distinguish between acute respiratory distress syndrome (ARDS) and cardiogenic pulmonary edema (CPE).

Recently, with the development of deep learning, deep convolutional neural network (CNN) techniques have been gradually applied to texture analysis field. CNN models like VGGNet (Simonyan and Zisserman (2015)), ResNet (He et al. (2015)), DenseNet (Huang et al. (2018)), etc have been developed and pretrained on the ImageNet (Russakovsky et al. (2015)), providing them resources to learn multi-level feature characteristics among diverse datasets, thus equipping them capabilities to work on different medical image tasks. These pre-trained models can be fine-tuned or simply used for feature extraction tasks.

Deep features extracted by pretrained deep-learning models are significantly different from those statistical features extracted by models like GLCM. Statistical texture features such as contrast and entropy are well-defined and interpretable features that human can understand. However, deep features are automatically learned abstract and complex patterns that humans cannot recognize and cannot be explicitly defined. Studies show that deep features outperform conventional statistical methods. In Dehbozorgi et al. (2025), comparison were made between different feature extraction tools on multiple image analysis tasks. On the hematoxylin and eosin staining (H&E-Stained) dataset, deep features extracted by pretrained models such as VGG16 and ResNet50 outperforms statistical features by around 6% in distinguishing between different tissue types or pathological conditions. Besides, for the abnormal chest X-

Texture Extraction Tool	Tasks	Paper	Modality
GLCM, RL	MS Classification	Zhang et al. (2008)	MRI
GLCM	Skin Tumor Detection	Almeida and Santos (2020)	MRI
GLRLM	Fibrosis Grouping Classification	Daginawala et al. (2016)	CT
GLCM, GLDM, etc	ARDS and CPE Classification	Brusasco et al. (2020)	US
Statistical, Deep Features, etc	H&E-Stained Classification	Dehbozorgi et al. (2025)	Histopathology
Statistical, Deep Features, etc	Chest Abnormality Classification	Dehbozorgi et al. (2025)	X-ray
Deep Features	Medulloblastoma Tumor Classification	Cruz-Roa et al. (2015)	Histopathology
Deep Features	Brain Tumor Classification	Amin et al. (2019)	MRI, CT
Deep Features	Brain Tumor Classification	Chato and Latifi (2017)	MRI
Deep Features	Tumor and Skin Lesion Classification	Lai and Deng (2018)	Histopathology, Dermoscopy
Deep Features	Brest Tumor Classification	Kriti et al. (2020)	US

Table 2.4 – Summary of studies of texture analysis on medical images

ray classification, deep feature methods also outperforms statistical texture features.

Cruz-Roa et al. utilized an enhanced version of VGG16 for medulloblastoma tumor classification task and showed good classification performance. Amin et al. proposed a new score level fusion technique that combines pretrained AlexNet and GoogleNet for brain tumor classification, demonstrating superior accuracy compared to individual models across multiple datasets. Chato and Latifi used VGG16, VGG19, and AlexNet to extract the texture features, then predict the survival time from brain tumor.

2.3 Fetal Brain US Image Analysis

This section provides the background and information for 2D fetal brain US images as well as the progress on their image analysis.

2.3.1 2D US Scans on Fetal Brain

2D US scans are taken on the basis of the following three views: (i) Axial view, (ii) Coronal view, and (iii) Sagittal view. Each view provides unique anatomical information and specific planes are selected for different clinical purposes (Torres et al. (2022)).

For axial views, transventricular(TV), transthalamic(TT) and transcerebalar(TC) are three commonly used planes. In the TV plane, clinicians usually focus on landmarks such as cavum septi pellucifi(CSP), laternal ventricles(LV), and choroid plexus(CP). While TT plane crosses through thalami and CSP so biometric landmarks like biparietal diameter(BPD) and head circumference(HC) can be measured in this plane. For the TC plane, transcerebellar diameter(TCD) can be measured.

Coronal view is obtained perpendicular to the axial plane, which can provide a front-facing illustration of the brain. It also contains transthalamic(TT) and transcerebalar(TC) planes but in a vertial angle to those two from the axial views. Besides, transfrontal(TF) and transcaudate(TC_a) are another two planes in this view. In general, midline structures like CSP, third ventricle, etc can be assessed in this view.

Lastly, sagittal view is taken to divide the fetal brain into left and right hemispheres, corpus callosum and brainstem are usually assessed in this view. Table 2.5 outlines the imaging planes for each view and the assessed anatomies for each plane.

2.3.2 Progress on the Fetal Brain US Image Analysis

Besides from the segmentation tasks on fetal brain US images listed in Section 2.1.2, there's some other analysis on fetal brain US images and most of them focus on

View	Planes	Assessed Anatomy
Axial View	Transventricular (TV), Transthalamic (TT), Transcerebellar (TC)	CSP, LV, CP, BPD, HC, TCD
Coronal View	Transfrontal (TF), Transcaudate (TC_a), Transthalamic (TT), Transcerebellar (TC)	CSP, Third Ventricle, Corpus Callosum, Thalami
Sagittal View	Midsagittal Plane	Corpus Callosum, Brainstem, Cerebellum

Table 2.5 – Summary of 2D fetal brain US views and their clinical relevance

Assessed Region	Task	Analysis Type	Paper
HC, BPD	Gestational Age Prediction	Size/Length	Burgos-Artizzu et al. (2021)
BPD, FL	Gestational Age Prediction	Size/Length	Mul et al. (1996)
TCD, CC	Relationship Analysis	Size/Length	Bookstein et al. (2024)
Thalamus Diameter	Automatic Measurement	Size/Length	Sridar et al. (2017)

Table 2.6 – Summary of studies on fetal brain US image analysis

the size and shape of the anatomy in the US images (Table 2.6). Burgos-Artizzu et al. used head circumference(HC) and biparietal diameter(BPD) measurement to predict the gestational age. In Mul et al. (1996), researchers measure BPD and femur length(FL) from the fetal US images and it showed good accuracy in predicting gestational age. Besides, Bookstein et al. investigated the relationship and agreement between corpus callsom and transverse cereballar diameter in fetal imaging. They reached a conclusion that TCD measurements show high consistency between US and MRI. Moreover, Sridar et al. developed an automatic way to measure the thalamic diameter in 2D fetal brain US images.

Chapter 3

Fetal Ultrasound Image Analysis: Texture Feature Variability in Thalamus

3.1 Introduction

US imaging is extensively utilized globally during pregnancy due to its non-invasive, non-ionizing, and cost-effective charateristics. US scans are taken in various angles of the mother's abdomen to assess fetal health and growth. They are usually taken in different gestational stages due to different assessment purposes. For example, a transvaginal scan is taken during 6 to 10 weeks of gestational age (GA) to confirm baby's due date and the number of fetus (Queensland Health (2020)). During 11 to 14 weeks of GA, a transdominal scan is commonly used to assess the risk of chromosomal abnormalties (A. P. Souka and A. Pilalis and A. Antsaklis (2004)). During 18 to 22 weeks of GA, the fetal growth rate accelerates significantly. In this case, a few more US scans are usually taken during this stage. For example, fetal echocardiogram is taken to evaluate fetus heart growth (Kinman (2018)), anatomy scan is used to measure the head circumference (HC), thus assessing fetal growth (International Society of Ultrasound in Obstetrics and Gynecology (2022)). Besides, in Marion et al. (2023),

cranial scans are taken to check possible hydrocephalus in fetal brain and in Bekele et al. (2024), transcerebellar scans are taken to assess cerebellar development.

During 18-22 weeks of gestation, fetus shows rapid growth (Schumacher et al. (2020)). Compared to other brain structures such as cerebellum, thalamus undergoes more rapid and dynamic changes during fetal development. In Vathy and Ilona (2002), experiments show that during 20 and 28 weeks of gestation, thalamus shows non-linear but dynamic changes. Besides, in Sridar et al. (2020a), the thalamus diameter (TD) is proved to grow in relation to the gestational age (GA). Additionally, Villiger et al. (1982) demonstrates that the thalamus contains the highest concentration of serotonin and opioid receptors in the brain. This density keep increasing until adulthood. This emphasizes the high dynamic growth in the thalamus region during gestation and implies the possible richer information may contain in the US thalamus region compared to other brain structures.

Assessing size and shape characteristics has been the primary focus of existing studies on fetal brain US images. For example, Sridar et al. (2020a) proposed a method to measure the thalamus diameter (TD) while Singh et al. (2021) developed an automatic approach for cerebellum segmentation from fetal US images. However, numerous studies suggest that brain structures undergo not only size changes but also structural or textural changes. According to Scott et al. (2012), during 20 - 40 weeks of gestation, the fetal brain undergoes rapid growth, accompanied by cortical folding and volume changes. These findings highlight the importance of going beyond conventional size-based analysis to investigate texture variations in fetal brain development during gestation.

However, due to the small size of sub-regions like thalamus, the uniform appearance and high noise levels in US image, it is usually extremely hard to notice these texture variations by human eyes. Fortunately, with the help of conventional texture extraction tools and pre-trained deep learning techniques, unrecognizable texture features can be effectively extracted from the targeted region, thus conducting further texture investigation.

This study investigates variations in thalamus texture features across different ges-

tational age groups at birth (26.4-36 weeks, 36-39 weeks, 39-42.4 weeks). Given the promising potential of texture feature analysis in studying fetal brain development, thalamus is chosen in this study to conduct texture analysis due to its high dynamic changes during gestation as mentioned above. This study aims to explore potential developmental differences in US images by analyzing the thalamus texture feature variations across different gestational age groups at birth. Notably, all images were taken within the same gestational window (18-22 weeks of gestation).

For texture feature extraction, conventional statistical methods such as GLCM (Haralick et al. (1973)) and GLDM (Weszka et al. (1976)) were used. Moreover, to extract deep features, pre-trained deep learning model: ResNet18 (He et al. (2016)) was used as well. Combining both statistical and deep texture features can provide a broader and more comprehensive perspective of texture characteristics and enables a comparative analysis between the two approaches. To evaluate the significance of texture differences across gestational age groups at birth, ANOVA test with p-values were employed to determine whether individual texture feature characteristic holds statistically significant variations. To further support the observed variations across groups, Pearson correlation was used to study the relationship between each texture feature and gestational age, reinforcing the significance of these variations are associated with gestational age at birth as well.

Results show that numerous deep texture features exhibit significant variations across different gestational age groups, with the Pearson correlation coefficient around 0.2 to 0.3, indicating weak to moderate correlation between the individual texture characteristic and gestational age at birth. For conventional statistical texture feature, there's still a few texture features demonstrates significant variations, even though a less extent. This proves that 18 to 22 weeks GA thalamus textures show significant difference across different gestational age groups at birth. Furthermore, deep texture features can capture more fine-grained texture details than statistical texture features, providing a more comprehensive representation of developmental differences.

This study paves the way for using texture features instead of simple size or shape analysis from US image sub-regions to better understand fetal brain development.

By identifying the texture feature differences in multiple gestational age groups at birth, this study establishes a connection between texture characteristics and fetal brain growth. This would benefit future neurodevelopmental studies such as early estimation of brain maturation or refining gestational age assessment, etc.

3.2 Methods

This section covers all the methods used in this study. Section 3.2.1 introduces the dataset used in this study. Section 3.2.2 goes through the pipeline of this whole study and Section 3.2.3 introduces the preprocessing steps in this study. Section 3.2.4 covers texture feature extraction methods used in this study, including statistical texture extraction methods and deep learning-based extraction techniques.

3.2.1 Datasets

All the data used in this study are provided by the Nepean Hospital, Sydney, Australia. The Human Research Ethics Committee (HREC) of the Nepean Blue Mountains Local Health District (NBMLHD) has waived the informed consent fo the use of de-identified images in this retrospective study (HREC Study Reference No: Study 07/002). All US images were captured on the transcerebellar plane during 18-22 weeks of gestation while the gestational age data was recorded at birth. The dataset contains 2045 US images, evenly distributed across different gestational ages at birth. This ensures the data is not biased towards one specific gestational age range, enabling a more comprehensive analysis of texture variations across multiple gestational age groups at birth.

The targeted thalamus region is demonstrated in Figure 1.1. Since the scan is transcerebellar, there is no clear boundary for the thalamus. In this case, it is necessary to define a region that is a sub-set of the thalamus to clear any ambiguity of having other tissues in the region. The two triangles in Figure 1.1 represent the thalamus

region used in this study, they are defined based on anatomical landmarks and geometric relationship in the image. For example, for $\triangle$ *CDE*, it is defined based on the following rules. *Point A* is the tail vertex of the guitar shape and *Point B* is the foot of the perpendicular dropped from *Point A*, and *Point C* is the midpoint of *AB*. *Point D* is the turning point of half of the guitar shape and *point E* is the foot of the perpendicular dropped from *point D*. Based on these, point C, D and E forms the $\triangle$ *CDE*. $\triangle$ *HJI* follows the same rules as $\triangle$ *CDE*.

The definition of the thalamus region mentioned above was verified and confirmed by experienced ultrasonographers.

3.2.2 Overview of the Pipeline

Figure 3.1 outlines the process flow for this study. Based on the US images from the dataset, the thalamus was first manually segmented. Next, the segmented region of interests (ROI) went through ROI enhancement for improved visibility. Then, texture features were extracted from the segmented thalamus region, using both statistical tools (GLCM, GLDM, etc) and deep learning-based tools (pretrained ResNet18) to capture comprehensive textural details. The extracted texture features were then analyzed to evaluate their relationship with gestational age at birth. More detailedly, ANOVA test with p-values were conducted to identify possible significant texture variations across different gestational age groups and Pearson correlation coefficient was calculated to investigate the possible relationship between each texture feature and gestational age at birth.

3.2.3 Preprocessing of the Images

Thalamus Segmentation

All thalamus regions were segmented manually to ensure accuracy and consistency, preventing any segmentation difference from affecting the future texture analysis

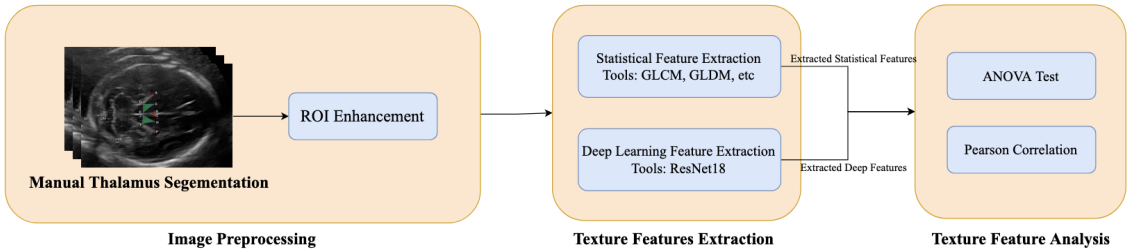


Figure 3.1 – Texture-based preterm birth prediction pipeline

tasks. More segmentation details are mentioned in Section 3.2.1. Although no formal inter-rater consistency evaluation was conducted, a random selection of 100 manually segmented images were confirmed by experienced ultrasonographers to ensure the consistency and precision of the segmented data.

ROI Enhancement on the Segmented Thalamus

After manually segmenting the thalamus region, image enhancement techniques were applied to segmented thalamus region to improve visibility, and this is helpful for future feature extraction. Two enhancement methods were used in this study. Contrast Limited Adaptive Histogram Equalization (CLAHE) were applied first to improve local contrast and enhance fine-grained details of the texture. Then, gamma correlation was utilized to adjust image brightness and highlight subtle details, which can make texture patterns more visible in the images.

These enhancement methods are beneficial for radiomics-based texture analysis, as they can improve the robustness of the extracted features and maximize the discriminative potential.

3.2.4 Texture Feature Extraction

Statistical methods and deep learning-based methods were used to extract the texture features from the sub-region: thalamus from the US images.

Methods	Texture Descriptors
GLCM	Contrast
	Correlation
	Energy
	Homogeneity
	Entropy
	Dissimilarity
	Cluster Prominence
	Cluster Shade
	...

Table 3.1 – Common descriptors of GLCM

Statistical Methods

In this study, existing statistical texture extraction methods were used for the basic texture feature extraction, including Gray-Level Co-occurrence Matrix (GLCM, Haralick et al. (1973)) , Gray-Level Dependence Matrix (GLDM, Weszka et al. (1976)), Gray-Level Size Zone Matrix (GLSZM, Thibault et al. (2013)), Gray-Level Run Length Matrix (GLRLM, Galloway (1975)) and Neighborhood Gray Tone Difference Matrix (NGTDM, Amadasun and King (1989)). Each of the methods focuses on different aspects of the texture features.

GLCM captures and focuses on the co-occurrence of pixel pairs at a given distance and direction, and this helps GLCM capture features like contrast, correlation, homogeneity, etc. However, GLDM focuses on capturing gray-level dependence between a pixel and surrounding pixels, which means GLDM is more sensitive to local texture variations. Instead of measuring co-occurrence of gray levels like GLCM, GLDM captures how many of the neighboring pixels are within the determined intensity difference. Gray-level non-uniformity is one of the common GLDM features that measures variation in dependence sizes. Table 3.1 and Table 3.2 list some common descriptors for GLCM and GLDM respectively.

Methods	Texture Descriptors
GLDM	Dependence Non-Uniformity
	Dependence Non-Uniformity Normalized
	Dependence Variance
	Gray Level Non-Uniformity
	Gray Level Variance
	High Gray Level Emphasis
	Large Dependence High Gray Level
	Gray Level Non-Uniformity Normalized
	...

Table 3.2 – Common descriptors of GLDM

Not like GLCM or GLDM, GLSZM focuses on size of region or zone that shows the same gray-level intensity while GLRLM calculates the length of consecutive pixels that share the same gray level. These characteristics help finding large homogeneous areas and identifying narrow, elongated structures within the image respectively. And finally, NGTDM compares the gray-level differences between neighboring regions to effectively capture and detect fine-grained patterns.

These five statistical methods ensures that the extracted texture features are comprehensive enough to capture various aspects of the thalamus' statistical texture characteristics. Specifically, they extract features by analyzing different relationships: pixel-level intensity distributions, co-occurrence patterns between neighboring pixels, dependencies between pixels, regional area-based relationships, and elongated patterns in pixels. Combining these methods enables a complete and robust texture representation of the thalamus, acting as a concrete for more accurate future predictions.

Methods	Texture Descriptors
GLSZM	Small Zone Emphasis (SZE)
	Large Zone Emphasis (LZE)
	Gray-Level Non-Uniformity (GLN)
	Zone Size Non-Uniformity (ZSN)
	Zone Percentage (ZP)
	Gray-Level Variance (GLV)
	Zone Variance (ZV)
	Zone Entropy (ZE)
	...

Table 3.3 – Common descriptors of GLSZM

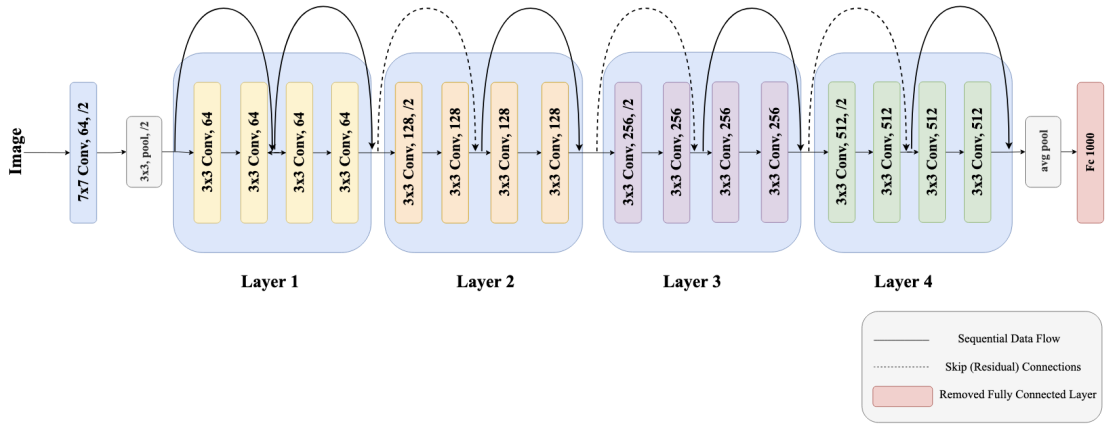


Figure 3.2 – Architecture of utilized ResNet18

Deep Learning-based Methods

To extract deep features from the thalamus, pre-trained deep learning-based methods were utilized in this study. More specifically, in this study, ResNet18, a Residual Neural Network (ResNet) was employed for deep feature extraction. Figure 3.2 demonstrates the overall architecture of ResNet18 used in this study. Proposed by He et al. (2016), ResNet18 is a deep convolutional neural network with 18 layers and has been widely used in medical image analysis. More details are mentioned in Chapter 2.

ResNet18 was chosen due to its lightweight structure compared to other pretrained

Methods	Texture Descriptors
GLRLM	Short Run Emphasis (SRE)
	Long Run Emphasis (LRE)
	Gray-Level Non-Uniformity (GLN)
	Run Length Non-Uniformity (RLN)
	Run Percentage (RP)
	Low Gray-Level Run Emphasis (LGRE)
	High Gray-Level Run Emphasis (HGRE)
	Short Run Low Gray-Level Emphasis (SRLGLE)
	Short Run High Gray-Level Emphasis (SRHGLE)
	Long Run Low Gray-Level Emphasis (LRLGLE)
	Long Run High Gray-Level Emphasis (LRHGLE)
	...

Table 3.4 – Common descriptors of GLRLM

Methods	Texture Descriptors
NGTDM	Coarseness
	Contrast
	Busyness
	Complexity
	Strength
	...

Table 3.5 – Common descriptors of NGTDM

models like ResNet50. As illustrated in Figure 3.2, the input image first passes through a 7x7 convolutional layer, followed by max pooling, which reduces dimensions while retaining abstract features. Then it goes through four consecutive 3x3 residual blocks (layer 1 to layer 4), each contains two residual units where each residual units contains two convolutional layers. Then, the features maps are aggregated through a global average pooling layer. In the original ResNet18 architecture, it follows by a fully connected layer. However, it got removed in this study to extract deep features directly from the model without giving the classification results. These extracted

deep features would be further used in future analysis in relation to gestational age groups at birth.

3.2.5 Relationship Analysis and Texture Feature Variability

Two separate tests have been conducted after the texture feature extraction. ANOVA test was applied to determine statistically significant difference between multiple groups and Pearson correlation analysis quantifies the relationship between each texture feature and gestational age at birth.

ANOVA, or Analysis of Variance (Fisher (1925)), is a commonly used statistical method to determine possible statistically significant differences between multiple independent groups. It is used in this study because ANOVA is suitable for evaluating possible significant differences among categories, which is suitable in differentiating gestational age groups. The p-value of ANOVA test less than 0.05 indicates the statistically significant variations between groups. While pearson correlation (Pearson (1895)) is a widely used statistical measure to quantify a linear relationship between two continuous variables. The two variables are extracted quantified texture features and gestational age at birth in this study. Table 3.6 shows the definition of ranges and their corresponding interpretations. More evaluation details are explained in Section 3.3.2.

Absolute Pearson Correlation (|r|)	**Interpretation**
0.00 – 0.10	No or very weak correlation
0.10 – 0.30	Weak correlation
0.30 – 0.50	Moderate correlation
0.50 – 0.70	Strong correlation
0.70 – 1.00	Very strong correlation

Table 3.6 – Interpretation of Pearson correlation coefficients (Cohen (1988))

3.3 Experimental Setup

3.3.1 Implementation Design and Details

All the images were resized to 975 x 729 pixels before the manual segmentation step. To analyze the possible relationship between texture features and gestational age, a set of statistical and deep features were extracted from the thalamus region. More details are mentioned in Section 3.2.1.

For each texture feature, its correlation with the gestational age was calculated. Besides, all the data were divided into the following three divisions: (i) 26.4-36 weeks gestational age, (ii) 36-39 weeks gestational age, and (iii) 39-42.4 weeks gestational age, based on the data distribution. ANOVA p-value was used to evaluate the variations of texture features across these gestational age groups.

A comparative analysis between ResNet18 and VGG19 was conducted by using standard evaluation metrics: ANOVA p-value and Pearson correlation, aiming to justify the selection of ResNet18 for texture feature extraction.

3.3.2 Evaluation Metrics

As mentioned in Section 3.3.1 and Section 3.2.5, ANOVA p-value is one of the evaluation metrics used in this study. If ANOVA p-value < 0.05, this means this feature shows significant variance across different gestational groups. To correct for the risk of false positives due to multiple comparisons, Bonferroni correction was applied. For deep features extracted by using ResNet18 or VGG19, a total of 512 features were evaluated, the adjusted significance threshold was set to $\alpha = \frac{0.05}{512} \approx 9.77 \times 10^{-5}$. For statistical texture features, a total of 75 radiomic features were evaluated, the adjusted significance threshold was set to: $\alpha = \frac{0.05}{75} \approx 6.67 \times 10^{-4}$. Only features with p-values below this threshold were considered statistically significant.

Another evaluation metrics used in this study is pearson correlation. This measures the linear relationship between each texture features and gestational age. For Pearson

correlation coefficient (γ), the interpretation as follows:

- $\gamma > 0$: Positive correlation
- $\gamma < 0$: Negative correlation
- $\gamma \approx 0$: No strong linear relationship

Combining ANOVA p-values and Pearson correlation coefficient can provide a more comprehensive understanding of the relationship between texture features and gestational age.

3.4 Results

For each texture feature characteristic, including both deep feature and statistical features, the Pearson correlation coefficient was calculated between itself and gestational age at birth. The absolute Pearson correlation coefficient value measures the strength of the relationship between each quantified texture feature characteristic and gestational age at birth. The higher the value, the stronger the relationship. Besides, the ANOVA p-value was used to evaluate the possible variations across the three gestational age groups as for each texture feature. A p-value less than the threshold indicates the statistically significant differences. Table 3.7 demonstrates the results for deep features extracted by ResNet18. A total of 512 deep feature characteristics were extracted by using pretrained ResNet18. Due to the huge amount of extracted deep features, only the top 20 features are listed, ranked based on the ANOVA p-value in descending order, where a smaller ANOVA p-value indicates a greater variation across different gestational groups. Table 3.8 demonstrates the results for deep features extracted by VGG19. Similarly, Table 3.9 shows the statistics about each statistical feature in relation to the gestational age at birth.

As shown in Table 3.7, the absolute values of the Pearson Correlation coefficients for most deep texture features ranges between 0.2 and 0.3, classified as weak correlation,

but many are close to the threshold for moderate correlation (0.3). The definitions and interpretations of Pearson correlation coefficient values listed in Table 3.6. This suggests that even though individual deep feature show weak to moderate correlations with gestational age, they still have and may indicate a stronger relationship. For the ANOVA p-value, all of them show the value less than the threshold ($\alpha \approx 9.77 \times 10^{-5}$), indicating statistically significant texture variations between the given gestational age groups.

A comparison between Table 3.7 and Table 3.8 indicates that features extracted by using ResNet18 demonstrate better stronger Pearson correlations with the gestational age at birth as well as smaller ANOVA p-value. This suggests that ResNet18 is more effective in capturing developmental texture features relevant to gestational age.

For Table 3.9, two statistical texture features: 'small area low gray level emphasis' and 'small area emphasis' show weak to moderate correlation (0.2-0.3) with the gestational age. 'Small area low gray level emphasis' is one feature characteristic from GLSZM (Thibault et al. (2013)), which measures the presence of small regions with low-intensity values. The higher of this value means there's more small areas in the image that show low-intensity. This feature texture characteristic can capture subtle tissue patterns within the image based on patches or areas, might be useful for investigating the density of the tissue. Besides, 'small area emphasis' is used to calculate the presence of small uniform areas. This value will be higher if there are more small but uniform areas in the images, which will be useful for investigating uniformalty of the image. However, for other statistical texture features, most of the Pearson correlation coefficient dropped to less than 0.15, showing very weak correlation value even though ANOVA p-value still less than threshold ($\alpha \approx 6.67 \times 10^{-4}$).

3.5 Discussion

Based on the results in Section 3.4, both deep and statistical features show the capability of distinguishing between different gestational age groups at birth. Table 3.7 and Table 3.9 indicate that a greater number of deep texture features show weak to

Feature	Pearson Correlation	ANOVA p-value
ResNet18_87	0.3040	4.67×10^{-47}
ResNet18_410	0.2945	2.67×10^{-41}
ResNet18_202	-0.2677	9.27×10^{-32}
ResNet18_203	0.2411	4.55×10^{-30}
ResNet18_29	0.2530	2.75×10^{-29}
ResNet18_284	0.2466	1.73×10^{-28}
ResNet18_210	0.2508	9.84×10^{-27}
ResNet18_345	0.2378	3.04×10^{-26}
ResNet18_67	0.2479	2.94×10^{-25}
ResNet18_383	0.2192	1.32×10^{-24}
ResNet18_232	-0.2141	3.60×10^{-24}
ResNet18_220	0.2135	4.78×10^{-23}
ResNet18_289	0.2173	5.35×10^{-22}
ResNet18_107	0.2076	2.57×10^{-21}
ResNet18_374	0.1991	1.81×10^{-20}
ResNet18_478	0.2043	2.78×10^{-20}
ResNet18_218	0.1904	1.29×10^{-19}
ResNet18_136	0.2146	2.21×10^{-19}
ResNet18_90	0.1763	2.98×10^{-19}
ResNet18_208	0.1993	1.09×10^{-18}

Table 3.7 – Statistical analysis of deep features (ResNet18) in relation to gestational age: Pearson correlation and ANOVA p-values

moderate (0.2-0.3) correlation with gestational age at birth, compared to statistical texture features. In general, the correlation values of deep features are higher than those of statistical features and ANOVA p-value of deep texture features are smaller than those of statistical features. This indicates deep features have a stronger ability to capture meaningful variations related to gestational age at birth.

It is worth noting that while many deep texture features only exhibit weak to moderate (0.2-0.3) correlations with gestational age at birth, these associations may still hold clinical relevance. Each of these features is just one among hundreds of features extracted from the ultrasound image. The fact that a single feature already shows some correlation suggests a potentially stronger association when features are considered together. These subtle patterns might not be strong enough to predict on their own, but they may serve as early imaging biomarkers for future neurodevelopment or

Feature	Pearson Correlation	ANOVA p-value
VGG19_255	0.2798	1.81×10^{-34}
VGG19_405	0.2492	8.86×10^{-31}
VGG19_57	-0.2242	4.92×10^{-27}
VGG19_375	0.2405	2.32×10^{-26}
VGG19_141	0.2511	2.12×10^{-25}
VGG19_152	0.2146	1.10×10^{-24}
VGG19_33	0.2183	5.13×10^{-23}
VGG19_325	0.2198	5.66×10^{-23}
VGG19_157	0.2262	1.31×10^{-21}
VGG19_42	0.2277	1.34×10^{-21}
VGG19_260	0.2327	2.31×10^{-21}
VGG19_211	0.2153	2.88×10^{-21}
VGG19_471	0.2183	3.47×10^{-19}
VGG19_406	0.2021	3.48×10^{-19}
VGG19_116	0.1922	2.59×10^{-18}
VGG19_496	0.2075	5.46×10^{-18}
VGG19_192	0.2153	1.46×10^{-17}
VGG19_306	0.1895	1.97×10^{-17}
VGG19_419	0.1671	1.36×10^{-15}
VGG19_53	0.1551	2.64×10^{-15}

Table 3.8 – Statistical analysis of deep features (VGG19) in relation to gestational age: Pearson correlation and ANOVA p-values

the risk of preterm birth.

However, it is noticeable that the ANOVA p-value of 'small area low gray level emphasis' and 'small area emphasis' are lower than some of the deep features, which indicates the possible better capability of these two statistical texture features in distinguishing gestational age groups at birth.

These findings suggest that most of the deep features have better capability of distinguishing between different gestational age groups than statistical texture features. This align with what mentioned earlier in the introduction: Chapter 1, that deep features are supposed to capture more complex information which contains higher-order spatial relationship and deeper tissue structures. While in some cases, statistical texture features can perform good distinguish capability as well.

These results and findings align with the expectations mentioned earlier that thalamus

Feature	Pearson Correlation	ANOVA p-value
SmallAreaLowGrayLevelEmphasis	0.2674	4.24×10^{-34}
SmallAreaEmphasis	0.2249	5.22×10^{-25}
ShortRunLowGrayLevelEmphasis	0.1554	2.83×10^{-12}
SizeZoneNonUniformity	0.1242	4.16×10^{-9}
SmallDependenceLowGrayLevelEmphasis	0.1189	2.45×10^{-8}
SizeZoneNonUniformityNormalized	0.1418	3.06×10^{-8}
LowGrayLevelZoneEmphasis	0.1338	4.37×10^{-8}
SmallDependenceEmphasis	0.1045	7.95×10^{-6}
ZonePercentage	0.1033	8.36×10^{-6}
ShortRunEmphasis	0.0978	2.83×10^{-5}

Table 3.9 – Statistical analysis of statistical features in relation to gestational age: Pearson correlation and ANOVA p-values

undergoes dynamic changes during 18 to 22 weeks of gestational stage. It is also worth noting that 'small area low gray level emphasis' and 'small area emphasis' can not only show good discriminative capability as for gestational age groups at birth but also provide interpretability. In contrast, deep features only show good distinguish results but lack interpretability for further study. This indicates the complementary nature of deep features and conventional statistical features, implies the future way of combining these two for further fetal brain development study.

This study also suggests that the fetal development not only are visible based on structural size and shape dimensions, but also structural and textural changes, paving the way for future radiomics-based texture analysis on fetal brain US images.

This study has some limitations. While individual texture features, both deep and statistical features show the good capability of distinguishing gestational age groups at birth, there's not an effective way used in this study to combine these features into a more comprehensive but interpretable representation. Simply adding these features together or use dimensionality reduction techniques like PCA would further reduce the interpretability. In this case, it would be great to develop a better way to combine the texture features together for further study. In the future, it will also be great to try to use these thalamus texture features to predict gestational age at birth, which could bring more clinical value.

Building on the success of this study, radiomics-based texture analysis will be conducted on not just thalamus but also other brain structures such as cerebellum and brainstem to gain a more comprehensive view of the relationship between their texture features and fetal brain growth.

3.6 Conclusion

This study utilized deep texture features and statistical features from the fetal thalamus to investigate the relationship between texture features and gestational age at birth. The findings indicate that there are statistically significant variations in texture features across different gestational age groups at birth and these features also show a weak to moderate correlation with gestational age at birth. Compared to conventional clinical biometric measurement assessments, the proposed texture-based approach provides an additional biomarker to capture tissue characteristics that are not related to size or shape. This capability and the results in this study also suggest that texture-based analysis can provide insightful results in relation to fetal brain maturation and development, which is beneficial for future research in fetal neurodevelopment and growth.

Currently, this approach has not been implemented in clinical practice as the texture analysis proposed in this study is in an early preliminary research stage and without clinical trial. As future work, the next step will be to use the extracted texture features from other sub-brain regions to predict gestational age or the potential of neurodevelopmental issues. More validation experiments and tests on multiple datasets, as well as improving robustness to different ultrasound devices and qualities, will be required before clinical adoption.

Chapter 4

Fetal Ultrasound Image Analysis: Sub-Region Segmentation

4.1 Introduction

Accurate manual segmentation of US images is challenging due to their low signal to noise ratio (SNR) and poor contrast, especially when identifying and segmenting sub-regions of the images. Furthermore, the process of manual segmentation are both time-consuming and labor-intensive.

Fortunately, various automated methods for US segmentation have emerged. Traditional methods such as active contour models (Kass and Witkin (1988)), the watershed algorithms (Beucher and Meyer (1993)) and graph cuts (Boykov and Jolly (2001)) have shown good segmentation results. However, these traditional segmentation methods have limitations. Studies claim that active contour only works well when the segmentation boundary is clear and watershed algorithms usually shows over-segmentation situation in regions with big gray-level gaps (Wang et al. (2023)). Graph cuts highly relies on the manually defined energy functions, making them less transferrable for different datasets (Boykov and Funka-Lea (2006)). More recently, deep learning methods such as U-Net (Ronneberger et al. (2015)), a U-shape encoder-decoder convolutional neural network (CNN) primarily designed for medical image

segmentation have shown significantly improved segmentation performance across various medical imaging modalities, including CT, US and MRI. This is due to CNN and its ability to learn from large amount of data and automatically extract hierarchical details and complex spatial patterns from the feature maps. For US images, U-Net has demonstrated its effectiveness in tasks such as breast tumor segmentation (Tang and Zhang (2023)) and laternal bone segmentation (Z. Wang (2021)). Moreover, variants of U-Net have been released, each introducing unique modifications to enhance the performance. For example, Oktay et al. (2018) proposed Attention U-Net, which incorporates attention gates to the original U-Net model. These attention gates augment the skip connections in the original U-Net architecture, allowing the model to focus on the most relevant region of the feature map while still leveraging the direct connections between the encoder and the decoder. Zhou and Siddiquee (2018) introduced U-Net++, an enhanced variant of U-Net featuring a nested and more densely connected architecture. U-Net++ refines the original U-Net by using a series of intermediate convolutions to capture multi-scale features. Researchers have conducted experiments with numerous variants of these deep learning based models to segment regions in various types of US images. For example, Zhao and Dai (2022) proposed an improved residual U-Net, where residual blocks were introduced to mitigate the vanishing gradient problem. By integrating residual blocks and channel attention mechanism, the model was effective in segmenting breast tumors from US images.

Despite the advancements in US image segmentation methods, most of these approaches primarily focused on segmenting larger anatomical structures, such as the liver, heart and brain. However, analyzing sub-regions, defined as smaller structures within these regions, such as the thalamus or cerebellum in the brain is equally critical in many clinical applications. For instance, in fetal brain transcerebellar US scan, the cerebellum segment might provide valuable information about fetal cerebellar growth and its potential impact on future neurodevelopment such as language and motor functions. As demonstrated in Limperopoulos and Robertson (2016), abnormalities in cerebellar development such as malformations or injuries, are associated

with long-term neurocognitive deficits.

This study aims to develop a segmentation model capable of accurately segmenting the target sub-region: cerebellum while effectively minimizing errors caused by the presence of other regions with similar shapes or features. Even though existing models such as Attention U-Net can apply attention weights to the sub-regions like cerebellum in the fetal brain US images, there still have limitations in its attention mechanism. One limitation of Attention U-Net is that the attention weights are assigned to each location in the feature map independently, rather than ensuring they sum up to a single distribution across the entire feature map. This might bring us situations when multiple regions end up with high attention weights, making it harder to focus on the specific sub-region aimed to segment. As a result, this can lead to less accurate segmentation of the target sub-region, especially in complex and noisy US images. Moreover, the attention gates in the Attention U-Net only capture feature maps from the same layer without incorporating multi-scale information. This limits the model's ability to capture both fine details and broader context, which are essential for accurately segmenting complex sub-regions in US images. Without leveraging multi-scale features, the model may struggle to differentiate the target sub-regions from surrounding structures with similar appearances, thereby reducing segmentation precision.

To improve this, an adaptive attention gate was proposed to generate a spatially interdependent attention distribution, which enables the model to assign attention in a probability-like distribution so that all spatial weights are interdependent. This helps the model focus more precisely on important regions. Furthermore, to better handle feature maps of varying dimensions, a multi-fusion gate was introduced before the adaptive attention gate, forming an adaptive multi-scale attention (AMFA) mechanism. This AMFA could ideally maximize the usage of feature maps, thus enhancing segmentation performance.

Segmenting sub-regions from US images requires global context at different imaging scale, such as large contextual information and small structural details. Considering about this, deep supervision (Lee and Xie (2015)) was incorporated to the proposed

model. Deep supervision is a technique to apply auxiliary loss functions at multiple layers of the network rather than only at the final layer. It can introduce loss at intermediate layers of the network, helps the model to learn useful features across multiple levels, thus improving the sub-region segmentation performance. Deep supervision has been successfully used in medical image segmentation (Chen and Liu (2023); Fu and Li (2023)).

To demonstrate the effectiveness of the proposed model, experiments on two separate US datasets were conducted: the private transcerebellar fetal (more details in Section 4.3.1) brain dataset and public Breast Ultrasound Images (BUSI) dataset (Shah (2019)), where the cerebellum and breast tumor were segmented respectively. The results demonstrate that the proposed model can not only achieves accurate segmentation of sub-regions in US images but also performs effectively on larger, more general regions. It ourperformed existing segmentation models overall and effectively reduced false segmentation in regions with similar appearances in the images.

4.2 Related Work

4.2.1 Deep Learning-Based Segmentation Models

This section provides an overview of the segmentation models benchmarked or used as a baseline in this study. It introduces their architecture, motivations and key improvements over standard baseline models, and their relevant segmentation performance.

U-Net

As mentioned in Section 2.1, U-Net is one of the first CNN architectures specifically optimized for biomedical image segmentation. Proposed by Ronneberger et al. (2015), U-Net is a symmetric architecture, which made up of an encoder, an decoder and a bottleneck layer as demonstrated in Figure 4.1.

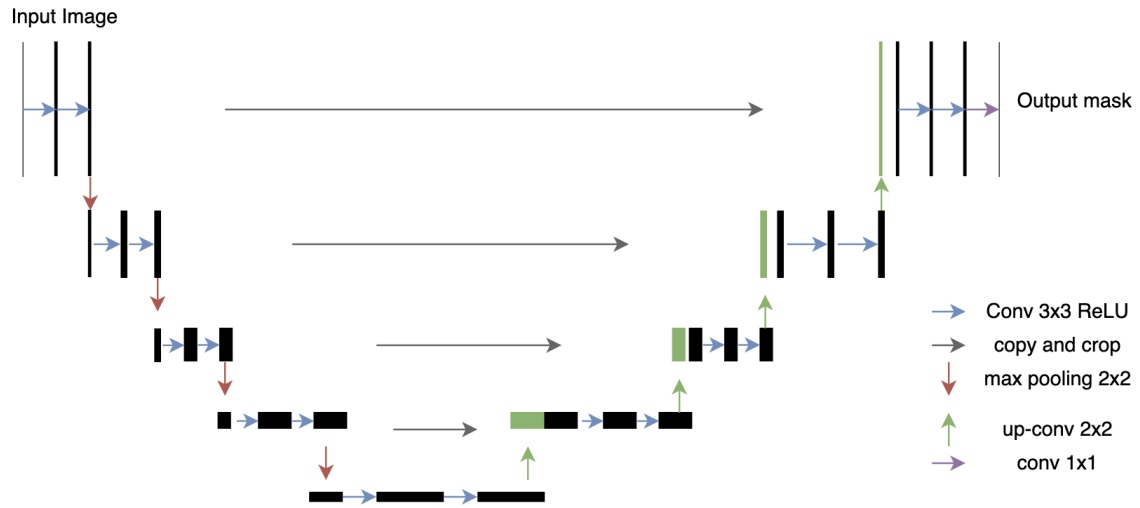


Figure 4.1 – U-Net architecture (Ronneberger et al. (2015))

The input image first goes through the encoder layer by layer and gets downsampled by conducting the maxpooling process. This step reduces the dimensions of the image so that the model can learn more abstract features of the image. Then it will goes through a bottleneck layer in which represents the most compressed and high-level feature representation, serving as the bridge between the encoder and the decoder. In the decoder, the downsampled feature maps are gradually upsampled through multiple layers until they match the same size of the original image. Additionally, skip connections concatenate feature maps from the encoder to decoder, which helps mitigate spatial information loss during the downsampling process.

This encoder-decoder architecture enables the model to learn high-level feature maps from the input image while preserving spatial details through skip connections, thus providing a more accurate segmentation result.

U-Net++

U-Net++ (Zhou and Siddiquee (2018)) is an enhanced model based on the original U-Net and designed to improve segmentation accuracy through nested and dense skip connections. The primary motivation of this enhanced segmentation model is to use

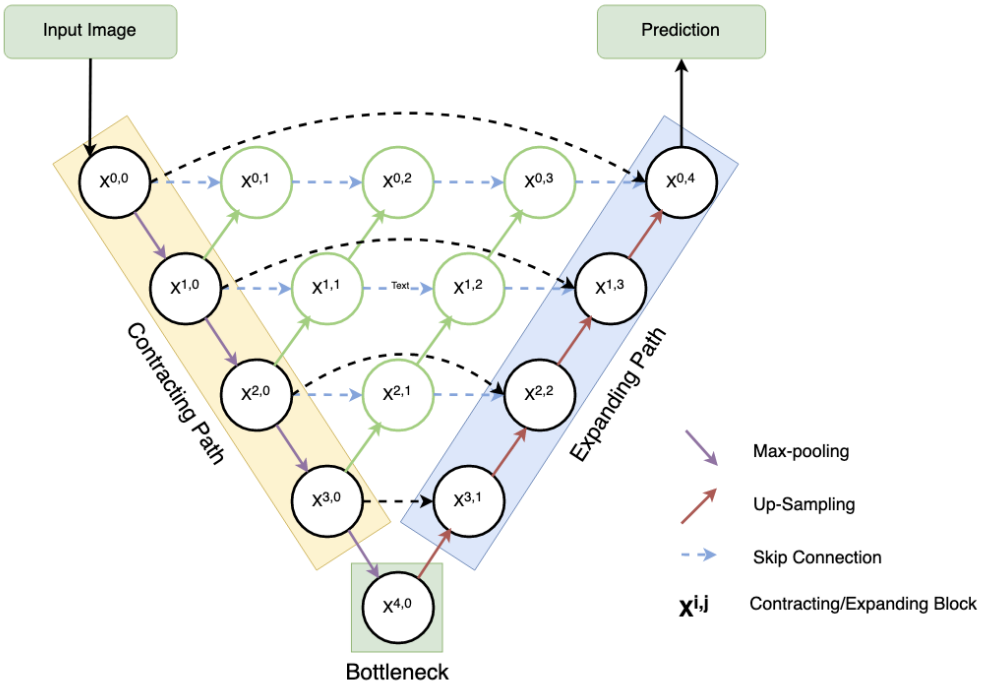


Figure 4.2 – U-Net++ architecture (Zhou and Siddiquee (2018))

intermediate feature fusion from multiple levels to bridge the semantic gap between the encoder and decoder.

The architecture of U-Net++ is demonstrated in Figure 4.2. It retains the encoder-decoder symmetric architecture from U-Net but incorporates the nested and dense skip connections to the original skip connections. More specifically, instead of directly connecting encoder and decoder feature maps through simple skip connections, in U-Net++, each encoder level is progressively connected to multiple decoder layers at different depths, allowing the decoder to receive the feature maps from different dimensions with more fine-grained details to minimizing the semantic map between the encoder and decoder, thus providing more accurate segmentation performance. Zhou and Siddiquee (2018) evaluated the segmentation performance on several different datasets such as nodule segmentation in chest CT scans and liver segmentation in abdominal CT scans, and they all show U-Net++ provides more accurate segmentation results than the original U-Net model.

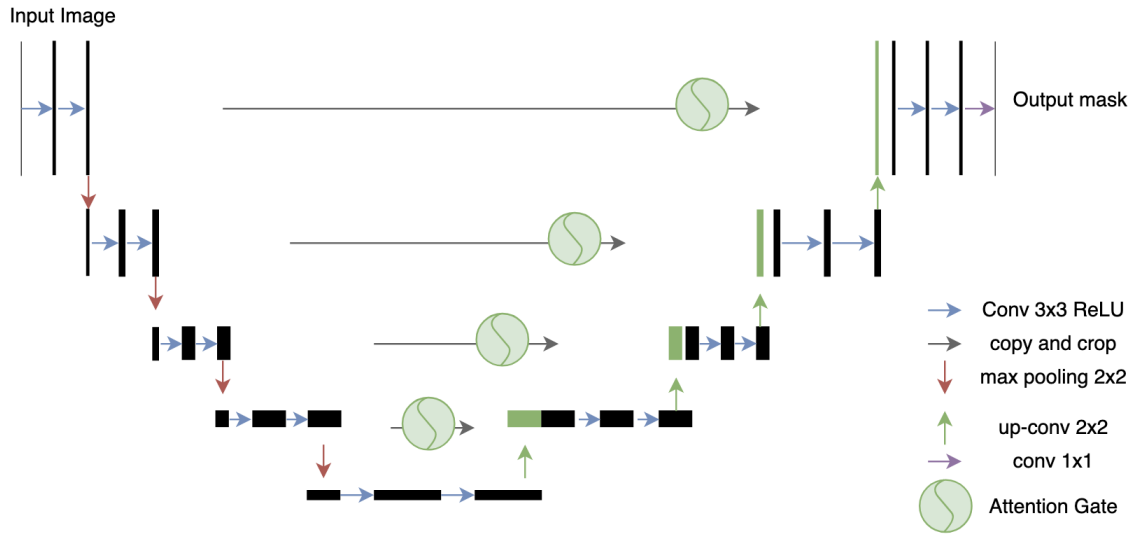


Figure 4.3 – Attention U-Net architecture (Oktay et al. (2018))

Attention U-Net

Attention U-Net (Oktay et al. (2018)) is another enhanced variant of U-Net model. Unlike U-Net++ which used nested and dense skip connections to bridge the gap between the encoder and decoder, Attention U-Net tries to improve segmentation accuracy by using an attention mechanism to make the model focus on the relevant regions of interest as well as suppressing irrelevant features.

Figure 4.3 shows the architecture of Attention U-Net, Oktay et al. (2018) integrated an attention gate to each layer's of skip connection, which can effectively refine the feature maps before they are concatenated to the decoder.

The architecture of the attention gate (AG) is shown in Figure 4.4. The encoder feature map x^l and the gating signal g first goes through 1 x 1 convolutions with learnable weights W_x and W_g respectively. Then the transformed feature maps go through an element-wise summation to combine the spatial information from both sources. Next, the fused features go through a ReLU activation function, a 1 x 1 convolution, and sigmoid activation function to generate attention coefficient whose values between 0 and 1 to be used as a weighting mask for the feature maps. To summarize, AG takes inputs from encoder feature map and decoder gating signal to

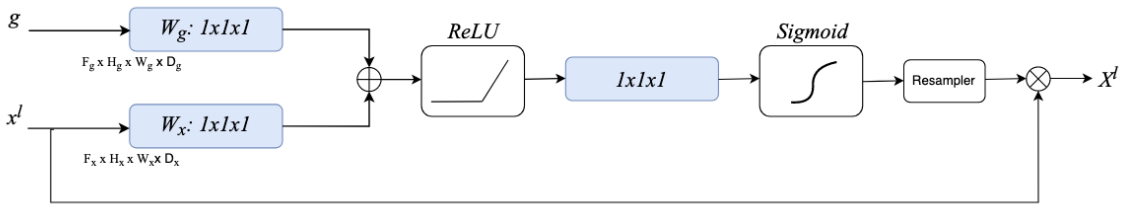


Figure 4.4 – Attention gate in Attention U-Net (Oktay et al. (2018))

generate weights on the feature maps so that regions of interests will be applied with more weights while irrelevant parts will be applied with less weights, thus providing more accurate segmentation results.

Oktay et al. (2018) compared the segmentation results between Attention U-Net and U-Net model on two CT datasets and they both show better performance when using Attention U-Net.

TransUNet

TransUNet (Chen et al. (2021)) is a hybrid segmentation architecture that integrates Transformer to U-Net in order to capture both local and global dependencies in medical images. Traditional convolutional networks like U-Net are effective in capturing local spatial information but have difficulty capturing long-range dependencies. This limitation is the motivation for the work by Chen et al. (2021), which embedded Transformer modules into U-Net and proposed: TransUNet.

In TransUNet, the encoder is a hybrid of CNN and Transformer, which first uses CNN layers to extract features, then the feature map will be flattened into 1 x 1 patches and passes through patch embedding layer before being fed into the Transformer encoder. They demonstrated that the hybrid encoder performs better than a pure Transformer encoder. In decoder, TransUNet follows the classic decoder from U-Net architecture, with skip connections and feature concatenation from the corresponding encoder.

The results show that TransUNet improves segmentation accuracy compared to other state-of-art methods. This improvement is attributed to its ability to capture long-

range dependencies through the Transformer in encoder, while still preserving spatial details from the CNN backbone. This characteristic makes it particularly work better where the segmentation highly relies on the global context of the medical image.

Swin-Unet

Swin-Unet (Cao et al. (2021)) is another Transformer-based segmentation model which utilizes Swin Transformer as its backbone. Compared to TransUNet, Swin Transformer uses shifted window attention to capture global context with a hierarchical feature representation. This design can significantly improve computational efficiency while capturing global dependencies.

Swin-Unet follows the U-shape architecture of U-Net but replacing the standard CNN blocks with Swin Transformer blocks. The encoder extracts multi-scale features through patch merging and Swin Transformer blocks, while the decoder constructs feature maps by using patch expanding and Swin Transformer blocks accordingly.

Compared to TransUNet, Swin-Unet provides a better trade-off between performance and computational efficiency and this makes it more suitable for high-resolution medical images.

Mamba U-Net

Proposed by Wang et al. (2024), Mamba U-Net is one of the most state-of-art segmentation model for medical images. Instead of making conventional changes to the original U-Net architecture, Wang et al. (2024) got the inspiration from the Mamba architecture, a recent advancement in State Space Models (SSMs), more details of Mamba mentioned in Section 2.1.

In Mamba U-Net, the most important change is that the standard convolutions are replaced due to their limitation in receptive fields. Instead, Visual State Space (VSS) layers are applied to encoder, bottleneck, and decoder layer to capture both long-range

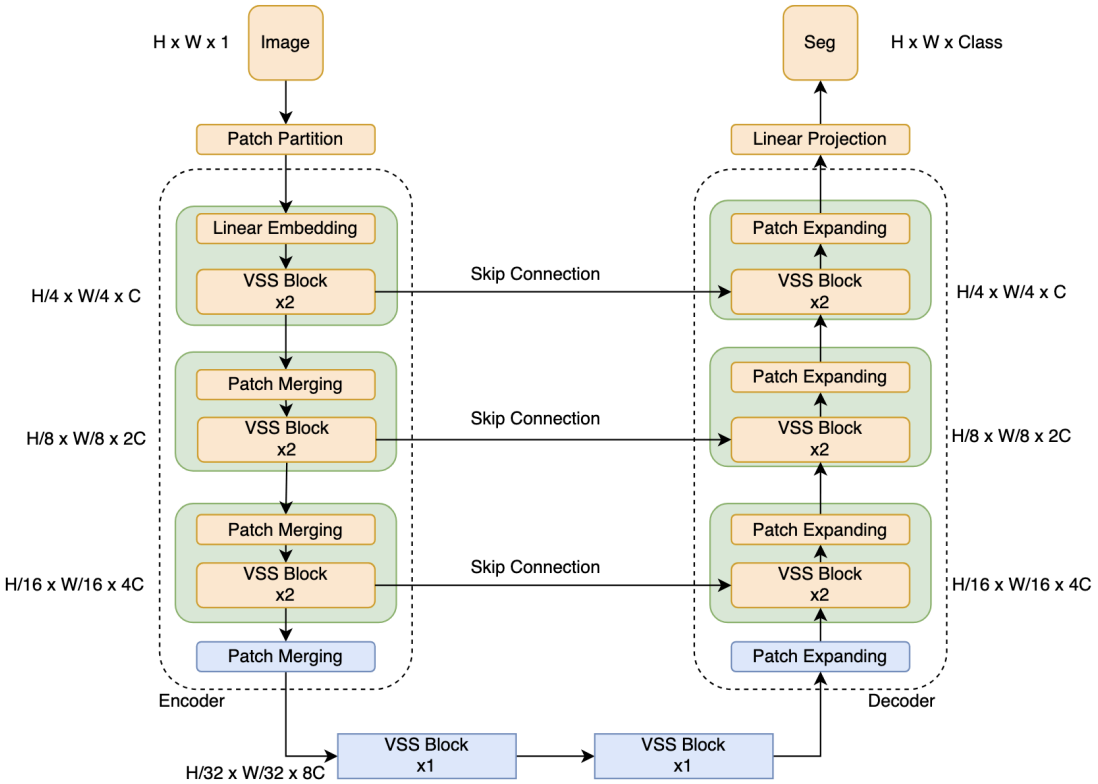


Figure 4.5 – Mamba U-Net architecture (Wang et al. (2024))

dependencies and local spatial information efficiently. The architecture of Mamba U-Net is shown in Figure 4.5, more architectural details are provided in Section 4.3.2. Results listed in Wang et al. (2024) shows that the proposed Mamba U-Net provides more accurate segmentation outcome on both the ACDC MRI Cardiac and the Synapse CT Abdomen segmentation dataset, proves the feasibility of replacing convention convolutions by VSS blocks.

4.3 Method

4.3.1 Datasets

Two separate datasets were used in the experiments to evaluate the effectiveness of the proposed AMFA network: the private fetal brain US dataset and the public Breast

Ultrasound Images (BUSI) dataset.

Fetal Brain US images

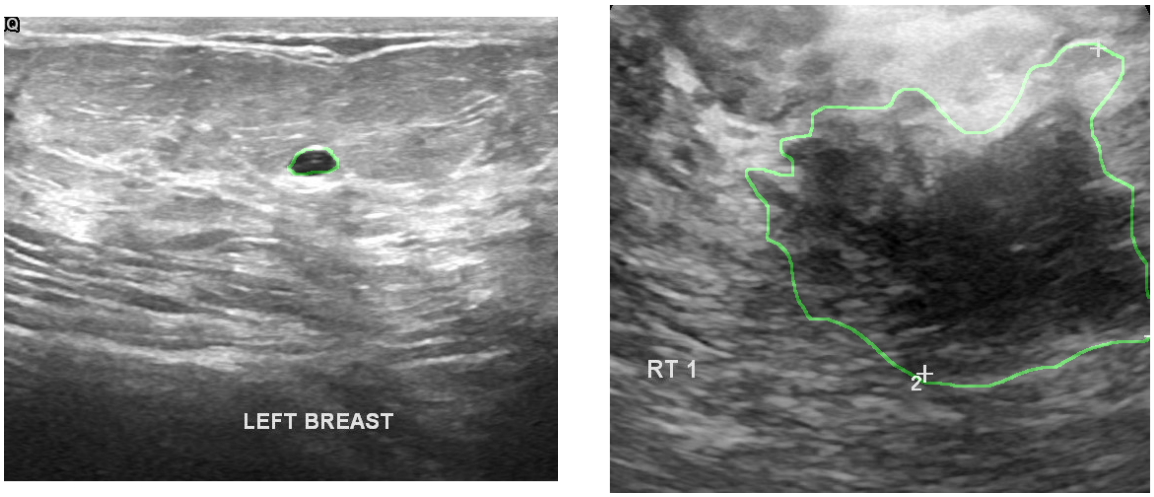
US images used in this dataset were collected from the Nepean Hospital, Sydney, Australia, and it contains 1086 images. The Human Research Ethics Committee (HREC) of the Nepean Blue Mountains Local Health District (NBMLHD) has waived the informed consent fo the use of de-identified images in this retrospective study (HREC Study Reference No: Study 07/002). All images were captured on the transcerebellar plane (plane explained in Section 2.3.1) between 18 and 22 weeks of gestation. The targeted segmentation region is the cerebellum, which annotated as the two circles demonstrated in Figure 1.2. Importantly, the definitions of the cerebellum region were verified and confirmed by experienced ultrasonographers.

All annotations were performed manually and subsequently reviewed by a clinician to ensure rigor and precision. A random selection of 100 manually labeled images were confirmed by experienced ultrasonographers to ensure the precision of the annotated data.

To improve the robustness and generalizability of the segmentation model, data augmentation techniques were applied to training images and ground truth masks. The augmentations included resizing, all images were resized to 256 x 256. This resizing step helped standardize the data format and allowed the model to process the images efficiently. No other data augmentation techniques were applied.

Breast Ultrasound Images (BUSI) dataset

Breast Ultrasound Images (BUSI) is a public dataset which includes breast ultrasound images from women between 25 and 75 years old. It has 780 scans in total, with 437 benign, 210 malignant and 133 normal cases. Since the focus of this research is on segmentation performance, only 437 benign and 210 malignant cases were used. The example scans for the benign and malignant breast tumor segmentations are shown in Figure 4.6.



(a) (b)

Figure 4.6 – Example scans in the BUSI breast tumor dataset: (a) benign and (b) malignant

4.3.2 Baseline: Mamba U-Net

Overall architecture

The Mamba U-Net architecture is shown in Figure 4.5. It consists of an encoder, a bottleneck layer, and a decoder. In the encoder, input images first pass through the patch partition layer, where each image is divided into smaller patches. These patches are then passed through a linear embedding layer, transforming them into a 1D sequence. Afterward, the sequence is processed through two consecutive VSS blocks to learn features and maintain dimension. The patch merging layer acts as downsampling process in this encoder, gradually captures hierarchical features while increases the feature dimensions. The encoder uses multiple pairs of patch merging and VSS blocks, allowing the inputs to reach the bottleneck layer with enriched feature representations.

The bottleneck layer consists of two VSS blocks, which connects the encoder and decoder. In the decoder, patch expanding is used as upscaling process to restore and reconstruct the dimension of the images. Following a structure similar to the encoder, two consecutive VSS blocks are used for feature reconstruction.

VSS block

The architecture of the VSS block is shown in Figure 4.7, primarily based on Visual Mamba by Liu et al. (2024). In one VSS block, input will first undergoes layer normalization layer (LN). Then the normalized input is split into two separate paths. One of them passes through a linear layer with SiLU activation, followed by a depth-wise convolution (DWCNN), then moves into the SS2D module and a subsequent layer normalization (LN). Another path will simply go through a linear convolution layer with SiLU activation.

The outputs of these two separate paths are then fused together through element-wise multiplication and the fused result is passed through another linear layer to refine the output features. Finally, the initial input is directly added to the output to preserve input information.

Mamba U-Net is chosen as the baseline of AMFA network due to its lightweight architecture and its strong capability for long-range dependencies. More specifically, the VSS blocks enable efficient sequence representation instead of self-attention mechanisms from Transformers. This structure retains the ability to capture long-range dependencies and global context but in a much more efficient way, which is more suitable for grayscale images like ultrasound with speckle noise and low contrast.

4.3.3 AMFA Network

Figure 4.8 shows the overall architecture of the proposed AMFA network. It follows the baseline of Mamba U-Net, retains its encoder-decoder structure, and essential components such as patch partition, linear embedding, VSS blocks as mentioned in Section 4.3.2.

For the connection between the encoder and decoder, the original skip connections in the Mamba U-Net was replaced with the proposed AMFA gate, enabling it to incorporate feature maps from adjacent layers as well as generating more effective weights for the feature map. In the AMFA network, each AMFA gate, except for the

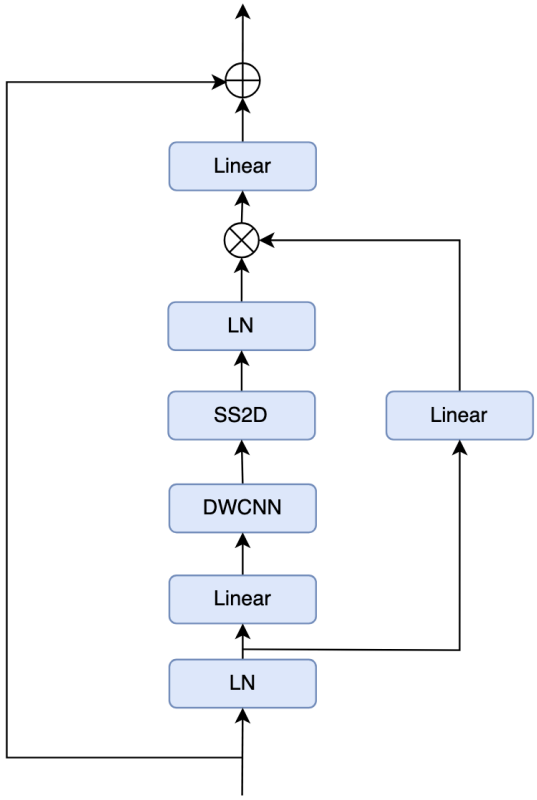


Figure 4.7 – VSS block (Wang et al. (2024))

deepest layer, receives two inputs: one from the VSS block at the current layer and another from a deeper layer in the encoder. The deepest layer only receives input from the VSS block at the same layer as there's no deeper layer in the model. These inputs are processed through the AMFA gate to generate attention or weight signals, which are then applied to the VSS block at the corresponding layer in the decoder. Besides, deep supervision was incorporated into the decoder of each intermediate layer of the network so that the loss function of the proposed network will be collaboratively decided by multiple layers instead of decided by only one output layer.

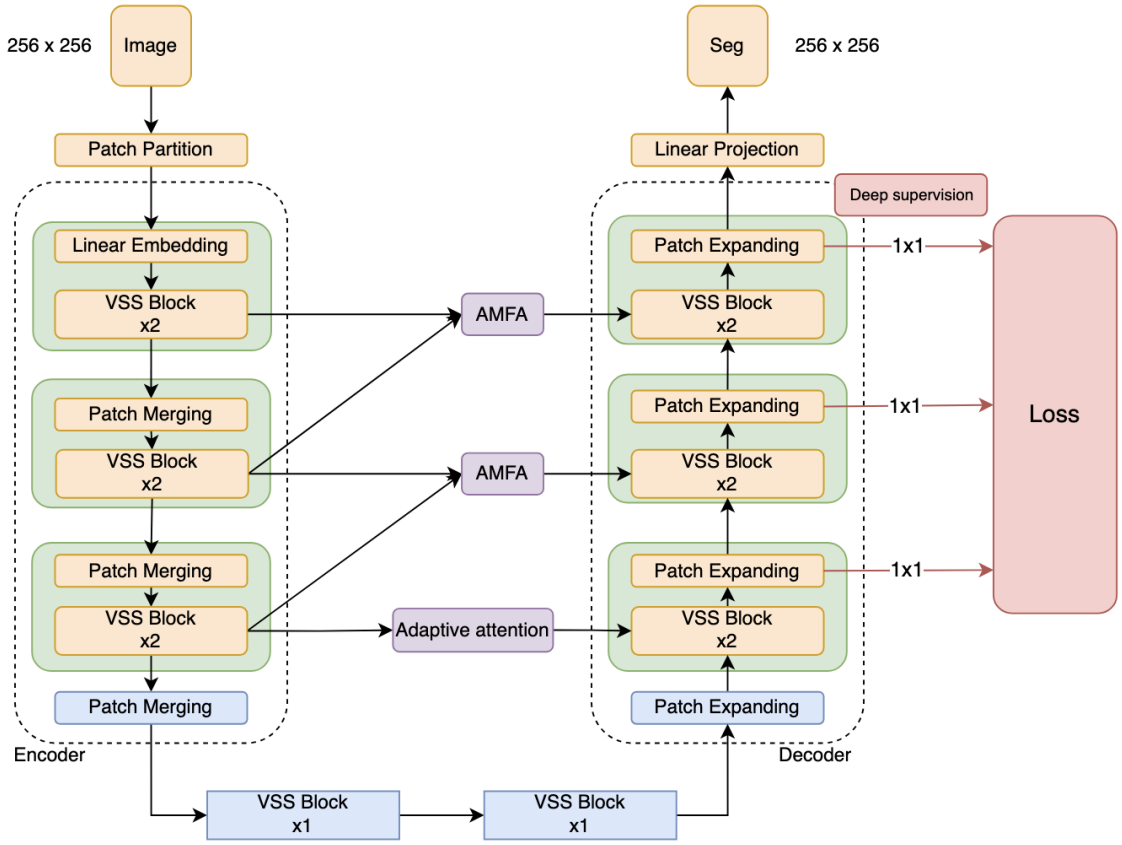


Figure 4.8 – Overall architecture of AMFA network

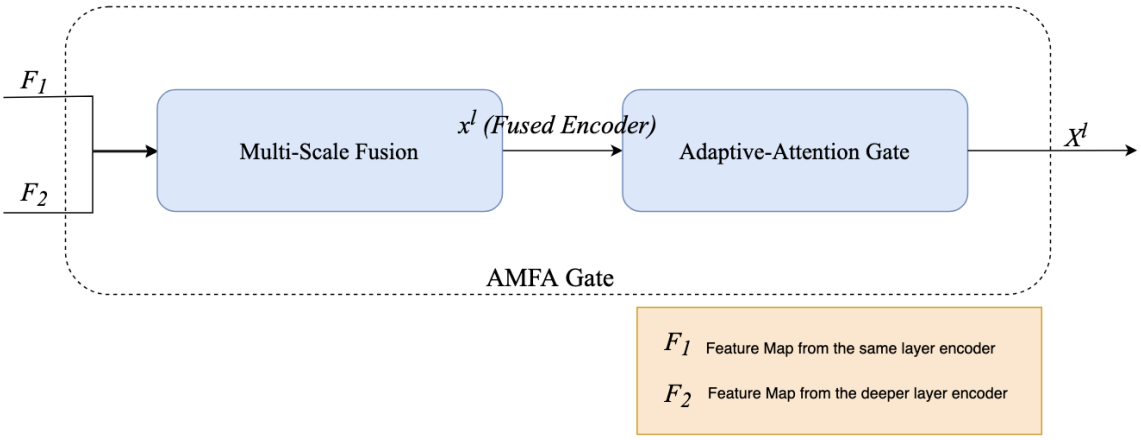


Figure 4.9 – AMFA gate

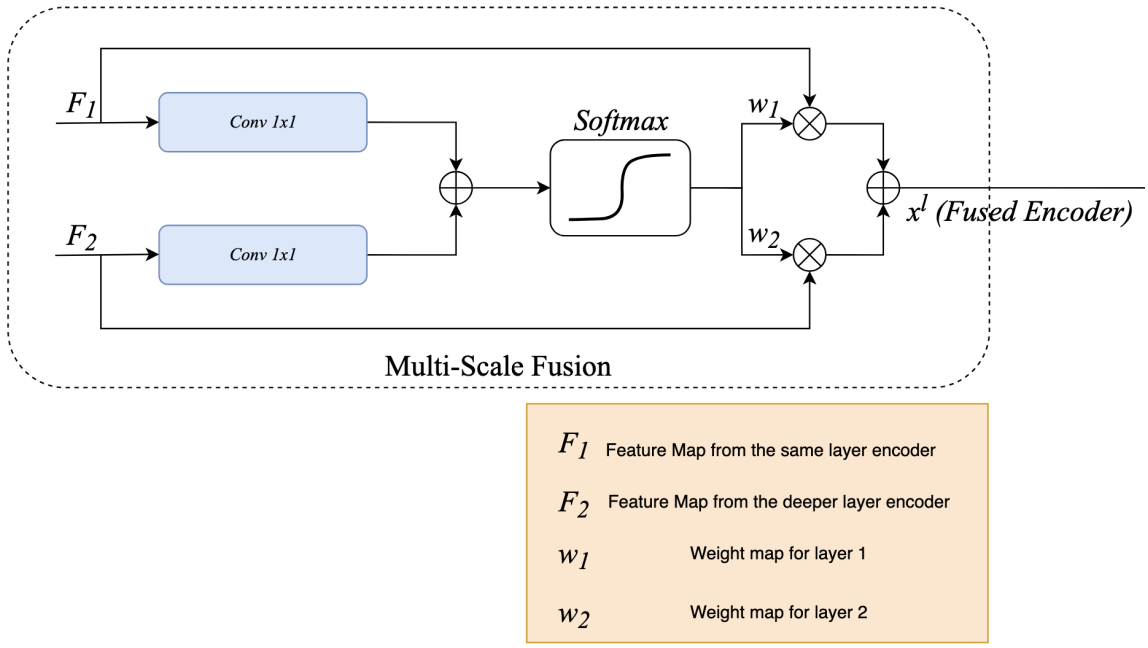


Figure 4.10 – Multi-scale fusion in AMFA gate

4.3.4 AMFA Gate

AMFA Gate

Figure 4.9 shows the main components and flow of each AMFA gate. It consists of two components: multi-scale fusion and an adaptive attention gate. In simple terms, the multi-scale fusion component takes two inputs F_1 and F_2, combines them into a single representation, and passes it to the adaptive attention gate. The adaptive attention gate then re-weights different regions of the feature map, assigning higher weights to the focused sub-region while reducing the weights for less relevant regions. This helps generates a weighted feature map, with the focused region receiving greater emphasis. Equation (4.1) to Equation (4.4) present the complete mathematical formulations of the proposed AMFA gate. Further details are discussed in the subsequent sections.

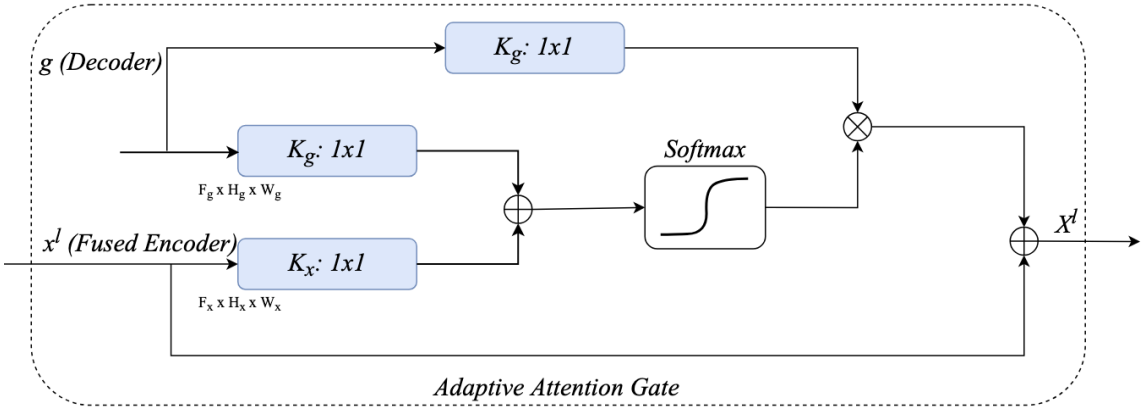


Figure 4.11 – Adaptive attention in AMFA gate

Multi-scale Fusion

The multi-scale fusion module (Figure 4.10) takes inputs from VSS block of the current layer(F_1) and a deeper layer of the encoder(F_2). Equations for the multi-scale fusion module is in (4.1) and (4.2). Each of these inputs passes through a 1x1 convolution, primarily to reduce the number of channels and compress the features, making the subsequent feature merging more efficient. The outputs of these convolutions are then fused using element-wise addition, combining information from both levels. Afterwards, this fused map is processed by a softmax operation, which generates two weights, w_1 and w_2, corresponding to the feature maps from the current layer and the deeper layer respectively. Once the weights are determined, they are applied to their respective feature maps through element-wise multiplication, adjusting the contributions based on the learned importance of each feature map. Finally, the weighted feature maps are combined through addition, generating the final fused feature map.

Adaptive Attention Gate

The proposed adaptive attention gate (Figure 4.11) is modified based on the attention gate in Attention U-Net proposed by Oktay et.al.. This component builds upon the segmentation work originally developed in the author's undergraduate thesis, with further enhancements made as part of this study. The equations governing the

adaptive attention gate are presented in (4.3) and (4.4). Two key modifications were introduced in the adaptive attention gate compared to the original attention gate.

First, the original ReLU + sigmoid weighing activation function was replaced by one softmax to generate a real weight distribution of the feature map. Second, the feature map from the encoder is concentrated into the weighted feature to retain more fine-grained feature map details.

In the original attention gate (Figure 4.4), the ReLU function retains values in the feature map that are greater than zero and replaces any negative values with zero. Then, the sigmoid function compresses the feature map, ensuring that the weights are between 0 and 1. However, the weights generated for each region are independent of one another, and the total weight across the feature map does not sum to 1.

Differently, in the modified adaptive attention gate (Figure 4.11), the softmax activation converts a set of values into a probability distribution over the feature map, ensuring that the total sum of weights equals 1. This makes the weights of each feature more interdependent and constrained by each other: when the weight of one region increases, the weights of other regions decrease accordingly.

$$x^l_{\text{Fused Encoder}} = w_1 \cdot \text{Conv}_{1\times1}(x^l_{\text{same}}) + w_2 \cdot \text{Conv}_{1\times1}(x^l_{\text{deeper}}). \tag{4.1}$$

where:

$$[w_1, w_2] = \text{Softmax}(\text{Conv}_{1\times1}(x^l_{\text{same}}) + \text{Conv}_{1\times1}(x^l_{\text{deeper}})). \tag{4.2}$$

In (4.1), $x^l_{\text{Fused Encoder}}$ is the fused feature map, with w_1 and w_2 as attention weights computed by the softmax function in (4.2). The terms x^l_{same} and x^l_{deeper} represent feature maps from the same and deeper encoder layers, respectively.

$$\alpha = \text{Softmax}(K_g \cdot g + K_x \cdot x^l_{\text{Fused Encoder}}), \tag{4.3}$$

$$x^l = \alpha \cdot g + x^l_{\text{Fused Encoder}}. \tag{4.4}$$

In (4.3), α is the attention weight applied to the decoder feature map g. Equation (4.4) defines x^l as the output feature map after combining the weighted decoder feature map and the fused encoder feature map $x^l_{\text{Fused Encoder}}$.

where:

- g is the feature map from the decoder.
- $x^l_{\text{Fused Encoder}}$ is the fused feature map from the encoder.
- K_g and K_x are 1×1 convolutions applied to g and $x^l_{\text{Fused Encoder}}$, respectively.
- α is the attention weight computed using the softmax function in (4.3).
- x^l is the output feature map after combining the weighted decoder feature map and the fused encoder feature map as shown in (4.4).

Comparison with Conventional Attention Gates

As discussed in Section 4.3.4, the conventional attention gate from Attention U-Net only contains a single encoder feature map and a decoder gating signal to calculate the attention weights. However, the proposed AMFA gate integrates both multi-scale encoder features and an enhanced adaptive attention mechanism, which significantly enriches the final feature representation. In this case, the resulting feature map becomes more informative and structurally clearer, allowing models for better segmentation accuracy even with high speckle noise and low contrast.

4.4 Experimental Setup

4.4.1 Implementation Details

All scans were resized to 256 x 256 pixels to be compatible with the proposed AMFA network architecture. 10-fold cross-validation were employed to ensure the robustness and reliability of the training outcomes. The network was implemented in PyTorch, utilizing the Adam optimizer for efficient parameter optimization. Deep supervision was applied to add auxiliary loss functions at different layers.

4.4.2 Evaluation Metrics

For the private fetal brain dataset, Dice Score (DSC), mIoU, precision, and recall were used to evaluate segmentation performance. For the BUSI dataset, DSC was selected as the evaluation metric as it aligns with the metrics used in most other research to facilitate easier benchmarking and comparison.

4.4.3 Benchmark on the Fetal Brain Dataset

To evaluate the performance of AMFA network and its effectiveness of segmenting the sub-region in US images, experiments on the private transcerebellar fetal brain dataset were conducted. Comparisons were performed against several existing segmentation methods, including U-Net (Ronneberger et al. (2015)), U-Net++ (Zhou and Siddiquee (2018)), TransUNet (Chen et al. (2021)), Swin-Unet (Cao et al. (2021)), and Mamba U-Net (Wang et al. (2024)). As mentioned in Section 4.2, U-Net serves as a widely used baseline and U-Net++ introduces nested skip connections for improved segmentation. TransUNet and Swin-Unet are Transformer-based architectures which incorporate self-attention from Transformer to capture long-range dependencies in medical images. Specifically, compared to TransUNet, Swin-Unet improves computational efficiency. Mamba U-Net represents a state-of-the-art approach for ultrasound

imaging. This benchmark selection ensures a comprehensive comparison across basic, enhanced, and cutting-edge methods.

4.4.4 Benchmark on the Public BUSI Breast Tumor Dataset

To demonstrate that the proposed AMFA network is also effective in segmenting larger, general regions in US images while primarily designed for sub-region segmentation, additional experiments were conducted by using the BUSI breast tumor dataset. Benchmark with a few published results on this dataset were also performed. Proposed AMFA network was compared with U-Net, Attention U-Net, and U-Net++, using the segmentation results reported in Ejiyi et al. (2024), where these models were trained and evaluated on the BUSI dataset using a consistent experimental setup. In addition, SegmentNet (Ejiyi et al. (2024)) was also included in the comparison, as it is a segmentation framework that can effectively integrates and handles both global and local contextual information simultaneously. Moreover, the proposed model was compared with an optimal U-Net++ model (Shang and Lai (2024)) specifically designed and tested on the BUSI dataset. This model, which achieved strong segmentation results according to the original study, utilizes ResNet18 as its backbone and leverages the nested skip connection of U-Net++ to enhance segmentation performance. Finally, AMFA network was also compared with a variant of U-Net: SK-U-Net (Byra et al. (2020)), which is a model designed to provide more accurate segmentation results for breast lesion images, addressing the heterogeneity in breast mass sizes and image characteristics.

4.4.5 Ablation Study

To systematically evaluate the contribution of the proposed AMFA module, an ablation study was conducted by comparing different model configurations on the fetal brain dataset. The original attention gate was first tested in Mamba U-Net, then replaced it with the adaptive attention gate (Adaptive-Att Mamba) to examine how

Model	mIoU (%)	DSC (%)	Recall (%)	Precision (%)
U-Net (Ronneberger et al.)	59.35	70.29	73.41	75.15
U-Net++ (Zhou and Siddiquee)	71.29	81.89	83.74	84.04
TransUNet (Chen et al.)	59.48	74.31	70.58	78.96
Swin-Unet (Cao et al.)	54.83	70.61	68.31	73.77
Mamba U-Net (Wang et al.)	64.28	78.22	77.57	79.04
Attention Mamba (w/ ds)	70.73	82.85	82.88	82.86
Adaptive-Att Mamba (w/o ds)	71.37	83.28	84.05	82.57
Adaptive-Att Mamba (w/ ds)	72.89	84.31	**85.05**	83.60
AMFA Network (w/ ds)	**73.10**	**84.45**	84.55	**84.38**

Table 4.1 – Segmentation performance benchmark on the transcerebellum US scan, *w/ ds* and *w/o ds* indicate with and without deep supervision, respectively.

dynamic weight between encoder feature maps affects segmentation performance. Finally, the full AMFA network, which integrates multi-scale fusion into the adaptive attention mechanism, was evaluated to further enhance segmentation performance.

4.5 Results

4.5.1 Fetal Brain Dataset Results

As presented in Table 4.1, the proposed AMFA network(w/ ds) outperformed existing methods such as U-Net, U-Net++, TransUNet, Swin-Unet, and Mamba U-Net, when segmenting the cerebellum region out of the fetal brain US images, achieving a mIoU of 73.10%, DSC of 84.45%, recall of 84.55% and precision of 84.38%. While the original Mamba U-Net baseline did not perform well on this dataset, it still surpassed the performance of the U-Net architecture, reaching a mIoU of 64.38% compared to U-Net's mIoU of 59.35%. It is noticeable that TransUNet slightly outperformed U-Net with a mIoU of 59.48% while Swin-Unet underperformed the original U-Net architecture with only 54.83% and 70.61% as its mIoU and DSC respectively. Moreover, Mamba U-Net significantly outperformed both TransUNet and Swin-Unet. By adding the attention gate proposed by Oktay et al. (2018) to the Mamba U-Net

Model	DSC (%)
U-Net (Ejiyi et al.)	57.57
Attention U-Net (Ejiyi et al.)	47.89
U-Net++ (Ejiyi et al.)	55.87
SegmentNet (Ejiyi et al.)	64.79
Optimal U-Net++ (Shang and Lai)	69.70
SK U-Net (fine-tuned) (Byra et al.)	70.90
Mamba U-Net (Wang et al.)	71.18
AMFA Network (w/ ds)	**73.93**

Table 4.2 – Segmentation performance on the BUSI breast tumor dataset

(listed as Attention Mamba (w/ ds) in Table 4.1), the segmentation performance had a significant increase at around 6% for each evaluation metrics and the mIoU reached to 70.73%. This demonstrates the feasibility and effectiveness of adding an attention mechanism to the Mamba U-Net network.

4.5.2 BUSI Breast Tumor Dataset Results

Table 4.2 lists out the segmentation results of the proposed AMFA network in comparison with several external models on the BUSI dataset. The results demonstrated that the AMFA network achieves a DSC of 73.93%, demonstrating better performance than the baseline Mamba U-Net, which achieves 71.18% DSC. Representative segmentation results by using Mamba U-Net and AMFA network are illustrated in Figure 4.13 and it is noticeable that the AMFA network provides more accurate segmentation then Mamba U-Net on the both benign and malignant breast tumors.

It is noticeable that both Mamba U-Net and AMFA Network outperforms the original U-Net, Attention U-Net or U-Net++. Compared to other state-of-art methods, AMFA network achieves competitive results, outperforming 64.79% by SegmentNet (Ejiyi et al. (2024)), 69.70% by Optimal U-Net++ (Shang and Lai (2024)), and 70.90% by SK U-Net (Byra et al. (2020)).

4.5.3 Ablation Results

To evaluate the contribution in the proposed model, an ablation study was conducted on the fetal brain dataset (Table 4.1). Replacing the original attention gate (Attention Mamba (w/ ds)) by the proposed adaptive attention gate (Adaptive-Att Mamba (w/ ds)) resulted in a 2% performance improvement, making the mIoU and DSC reach to 72.89% and 84.31% respectively. Moreover, incorporating multi-scale fusion into the adaptive attention gate (AMFA network) further enhanced segmentation performance. The mIoU improved from 72.89% to 73.10%.

Additionally, to evaluate the effectiveness of deep supervision in the Mamba model, comparisons were made between Adaptive-Att Mamba with and without deep supervision (listed in table as Adaptive-Att Mamba (w/o ds) and Adaptive-Att Mamba (w/ ds)). The results indicate that incorporating deep supervision yields approximately a 1% improvement across all evaluation metrics.

Figure 4.12 visualize some comparison results between Mamba U-Net, Attention Mamba, Adaptive-Attention Mamba and AMFA network.

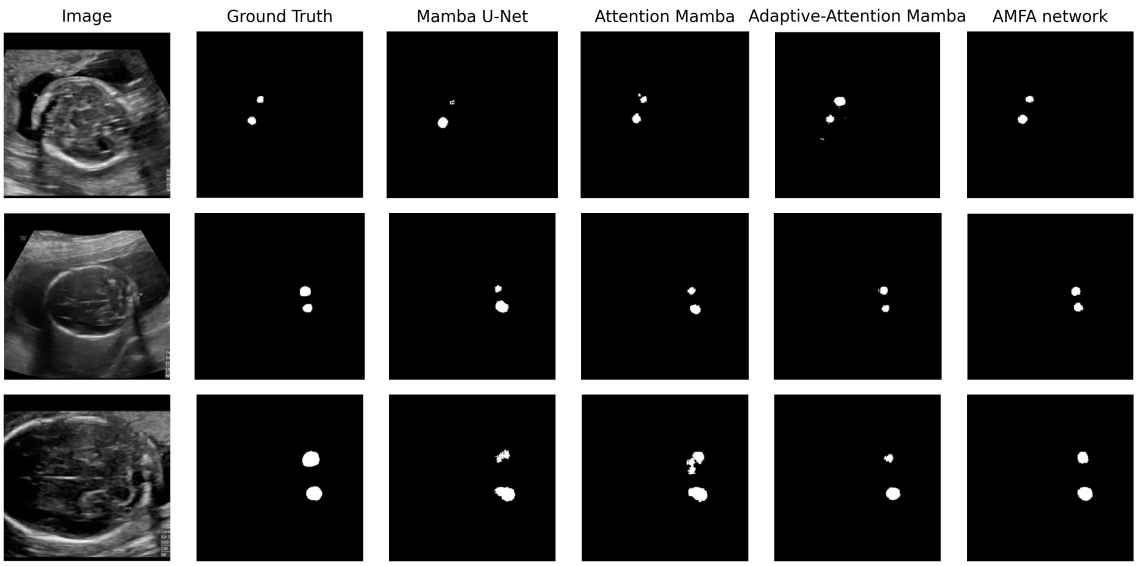


Figure 4.12 – Segmentation results comparison on the fetal brain dataset

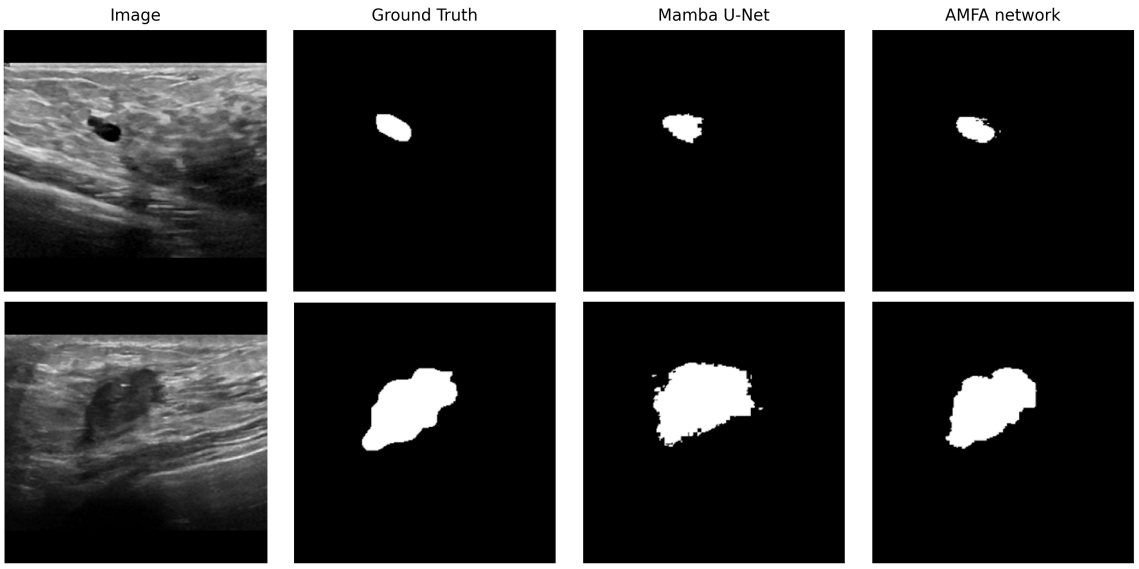


Figure 4.13 – Segmentation results comparison on the BUSI breast tumor dataset

4.6 Discussion

4.6.1 Fetal Brain Dataset

Based on the results mentioned above, the main findings of this study are as follows: (i)replacing ReLU + sigmoid with softmax in the adaptive attention gate improves segmentation accuracy by preventing irrelevant regions from receiving high attention weights; (ii) incorporating multi-scale fusion enhances feature representation by capturing fine-grained details from adjacent layers; and (iii) using deep supervision improves segmentation performance by refining gradient flow and stablizing training.

One of the key insights from this study is the impact of activation function on attention-based segmentation. The original ReLU + sigmoid activation function tends to assign high weights to multiple regions within a image, even when the model should focus on a specific sub-region. This might be problematic when segmenting sub-regions in a scan, as it can result in confusing or irrelevant regions being assigned high weights as well, which could negatively affect the segmentation performance. In contrast, the use of softmax activation in this algorithm generates a true distribution

of weights within the feature map, where the weights sum up to 1. This ensures the region that require the most attention stands out, while the less relevant areas receive very low weights, reducing the issue of multiple regions simultaneously having high weights.

Another important finding is the role of multi-scale fusion in improving segmentation quality. By incorporating information from adjacent layers, the network captures both fine-grained and high-level structural details, leading to better boundary delineation. Results confirm that this strategy contributes to measurable improvements in mIoU and DSC, making the model more effective at handling complex ultrasound textures.

Visual segmentation results also proves the proposed AMFA network can effectively reduce false segmentation. According to the example segmentation comparison showed on Figure 4.12, a comparison between column three and four demonstrates that the original attention gate does add some attentions to the sub-regions while it struggles to distinguish them from surrounding tissues, thus giving some inaccurate segmentation results with vague boundary. With the help of adaptive attention gate(without fusion), the segmentation becomes clearer and with AMFA network, which integrates multi-fusion mechanism, shows the best performance with clear segmentation boundary and accurate sub-region attention.

Besides, it is noticeable that although the Adaptive Attention Mamba (w/ ds) achieved the highest recall at 85.05%, its precision of 83.60% is relatively lower. In contrast, by incorporating multi-scale fusion, which aggregates features from different encoder layers to enhance fine-grained details, the AMFA network demonstrated a higher mIoU and a more balanced performance between recall and precision. This further validates that AMFA gate effectively captures features from adjacent layers and improves attention to target regions.

Finally, experiments show that deep supervision helps stabilize training and improve segmentation performance, particularly in deeper architectures like the Mamba-based network.

In addition, as is shown in Table 4.1 in results, Transformer-based models like Tran-

sUNet and Swin-Unet showed poor performance when segmenting the cerebellum region from the fetal brain US images. Swin-Unet even underperformed the original U-Net architecture. This may be attributed to the grayscale nature of the ultrasound images, the limited size of ultrasound dataset, and the noise and low contrast of the images. Transformer-based models usually requires large datasets and high resolution medical images to capture global dependencies effectively. In contrast, Mamba U-Net performed significantly better as the Mamba's lightweight architecture and its stable state-space design, which makes it more suitable for smaller dataset and images with noise. This further validates the selection of use Mamba U-Net as the baseline of AMFA network.

4.6.2 BUSI Breast Tumor Dataset

For breast tumor segmentation on the public BUSI breast tumor dataset, as shown in Table 4.2, the proposed AMFA network (w/ ds) outperforms the Mamba U-Net by around 3% in mIoU and dice score (DSC), further demonstrating the effectiveness of adding AMFA gate to the original Mamba U-Net, consistent with the results observed in the first dataset. By comparing the segmentation results obtained by using original U-Net and U-Net++ with those from Mamba U-Net and the AMFA Network, it is noticeable that incorporating Mamba to U-Net can provide better segmentation result and the proposed AMFA network can further enhance the performance. This aligns with the findings when evaluating on the first fetal brain dataset, proving the effectiveness of the proposed approach.

Experimental results were compared against published benchmarks from studies employing various segmentation methods. The proposed method surpassed the performance of Shang and Lai (2024) and Byra et al. (2020), achieving an approximate 2% improvement in dice score. Compared to SegmentNet in Ejiyi et al. (2024), the proposed AMFA Network can provide around 10% increase in the segmentation results.

The generalizability of AMFA on a different dataset can be explained as follows. First, the adaptive attention mechanism can enhance feature selection by enabling

the model to focus on the most relevant regions, regardless of anatomical structure or dataset variation. Second, the multi-scale fusion recovers the fine-grained details that might lose in the past convolutions, provides richer feature representations to the model, thus generating a more accurate segmentation.

All these results validate the effectiveness of the AMFA Network, proving is ability to provide robust segmentation performance when segmenting general and larger regions from the US images. This supports and demonstrates the generalizability of the AMFA Network, implying its potential segmentation capability across different US datasets and various US imaging tasks.

4.7 Summary

In this study, an enhanced segmentation model was proposed specifically for US sub-region segmentation: Adaptive Multi-fusion Attention (AMFA) network. In this model, AMFA gates are integrated into the Mamba U-Net baseline to effectively fuse feature maps from adjacent encoder layers, generating more accurate weights for improved segmentation. The proposed AMFA could effectively spot the sub-region and segment it more accurately, making it very suitable for segment sub-regions or regions from a relatively complex scan. This proposed architecture demonstrates superior performance compared to existing segmentation methods such as U-Net, U-Net++ on the fetal brain dataset and the BUSI breast tumor dataset serparately. Future work includes testing the AMFA on various baseline segmentation models to evaluate its feasibility and effectiveness. The aim is to develop the AMFA into a widely applicable component for sub-region segmentation, making it adaptable for integration into various segmentation models to enhance their performance.

Chapter 5

Conclusion

US imaging remains a routine diagnostic tool during pregnancy globally. In this thesis, two studies were conducted based on the transcerebellar fetal brain US images from 18-22 weeks of gestational age.

The first study explored the use of texture features instead of traditional size or shape-based analysis on the fetal brain US images. Deep learning and conventional statistical texture extraction methods were used to extract the features from the thalamus, using quantified textures to investigate the possible variations across three gestational age groups: 26.4-36 weeks, 36-39 weeks, and 39-42.4 weeks. The results show that for several deep features and a few statistical texture features, there exist statistically significant variations across the gestational age groups. It also shows weak to moderate (0.2-0.3) correlation between those single texture feature characteristic and gestational age at birth. These results imply the possible correlation between the texture features from fetal brain sub-region such as thalamus and the fetal growth and development, which could be clinically valuable for fetal neurodevelopment assessment and gestational age estimation.

In another study, a novel Adaptive Multi-Fusion Attention (AMFA) Network was proposed, which integrates a novel adaptive attention gate with multi-fusion mechanism, built upon the Mamba U-Net baseline. It is specifically designed for sub-region segmentation in noisy US images. The proposed AMFA Network was evaluated on

segmenting the cerebellum sub-region from the fetal brain US images. The results show that AMFA network outperforms other state-of-art models. Its performance was evaluated on the public BUSI dataset and it showed better results as well. These findings not only highlight the effectiveness of the proposed AMFA Network on the sub-region segmentation tasks but also its generalizability across multiple datasets.

In the future, the first study will be extended to radiomics-based texture analysis of other brain sub-regions such as cerebellum and brainstem, to gain a more comprehensive understanding of their relationship with fetal brain growth. The proposed US sub-region segmentation model from the second study of this thesis will be used in the segmentation step in the texture analysis pipeline for a seamless, automated, and effective workflow.

List of References

A. P. Souka and A. Pilalis and A. Antsaklis. Assessment of fetal anatomy at the 11–14-week ultrasound examination. 2004. Accessed: 2025-03-10.

Marcos A. M. Almeida and Iury A. X. Santos. Classification Models for Skin Tumor Detection Using Texture Analysis in Medical Images. *Journal of Imaging*, October 2020. Accessed: 2024-03-07.

M. Amadasun and R. King. Textural features corresponding to textural properties. *IEEE Transactions on Systems, Man, and Cybernetics*, June 1989. Accessed: 2024-03-06.

Javeria Amin, Muhammad Sharif, and Mussarat Yasmin. A New Approach for Brain Tumor Segmentation and Classification Based on Score Level Fusion Using Transfer Learning. *Journal of Medical Systems*, October 2019. Accessed: 2024-03-09.

Delayehu Bekele, Wondimu Gudu, Mekitie Wondafrash, Abdulfetah Abdulkadir Abdosh, and Abraham Fessehaye Sium. Utilization of third-trimester fetal transcerebellar diameter measurement for gestational age estimation: a comparative study using Bland-Altman analysis. *AJOG Global Reports*, 4(1): 100307, January 2024.

S. Beucher and F. Meyer. The morphological approach to segmentation: The watershed transformation. *Opt. Eng. Opt. Sci.*, January 1993. Accessed: 2024-03-02.

Shai Bookstein, Noy Nachmias, and Eldad Katorza. Agreement between Fetal Brain Ultrasonography and Magnetic Resonance Imaging in the Measurements of the Corpus Callosum and Transverse Cerebellar Diameter. *Diagnostics (Basel).*, Feburary 2024. Accessed: 2024-03-09.

Y. Boykov and M.-P. Jolly. Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images. *Proc. 8th IEEE Int. Conf. Comput. Vis.*, January 2001. Accessed: 2024-03-02.

Yuri Boykov and Gareth Funka-Lea. Graph cuts and efficient n-d image segmentation. *International Journal of Computer Vision*, November 2006.

Claudia Brusasco, Gregorio Santori, and Guido Tavazzi. Second-order grey-scale texture analysis of pleural ultrasound images to differentiate acute respiratory distress syndrome and cardiogenic pulmonary edema. *Journal of Clinical Monitoring and Computing*, October 2020. Accessed: 2024-03-07.

Xavier P. Burgos-Artizzu, David Coronado-Gutiérrez, and Kilian Vellvé. Analysis of maturation features in fetal brain ultrasound via artificial intelligence for the estimation of gestational age. *American Journal of Obstetrics Gynecology MFM*, September 2021. Accessed: 2024-03-09.

M. Byra, P. Jarosik, and A. Szubert. Breast mass segmentation in ultrasound with selective kernel U-Net convolutional neural network. *Biomedical Signal Processing and Control*, March 2020. Accessed: 2024-02-27.

Hu Cao, Yueyue Wang, and Yueyue Wang. Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation. *Computer Vision – ECCV 2022 Workshops*, May 2021. Accessed: 2024-03-04.

Lina Chato and Shahram Latifi. Machine Learning and Deep Learning Techniques to Predict Overall Survival of Brain Tumor Patients using MRI Images. *IEEE 17th International Conference on Bioinformatics and Bioengineering (BIBE)*, October 2017. Accessed: 2024-03-09.

G. Chen and Y. Liu. DSEU-net: A novel deep supervision SEU-net for medical ultrasound image segmentation. *Expert Systems with Applications*, June 2023. Accessed: 2024-03-02.

Gongping Chen, Lei Li, and Lei Li. AAU-Net: An Adaptive Attention U-Net for Breast Lesions Segmentation in Ultrasound Images. *IEEE Trans Med Imaging*, January 2022. Accessed: 2024-03-04.

Jieneng Chen, Yongyi Lu, and Yongyi Lu. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. *arXiv:2102.04306*, January 2021. Accessed: 2024-03-04.

Jacob Cohen. *Statistical Power Analysis for the Behavioral Sciences*. Routledge, New York, NY, 2nd edition, 1988. ISBN 978-0-12-179060-8.

Angel Cruz-Roa, John Arévalo, and Alexander Judkins. A method for medulloblastoma tumor differentiation based on convolutional neural networks and transfer learning. *11th International Symposium on Medical Information Processing and Analysis*, December 2015. Accessed: 2024-03-09.

Naznin Daginawala, Baojun Li, and Karen Buch. Using texture analyses of contrast enhanced CT to assess hepatic fibrosis. *European Journal of Radiology*, October 2016. Accessed: 2024-03-07.

Pegah Dehbozorgi, Oleg Ryabchykov, and Thomas W. Bocklitz. A comparative study of statistical, radiomics, and deep learning feature extraction techniques for medical image classification in optical and radiological modalities. *Computers in Biology and Medicine*, January 2025. Accessed: 2024-03-09.

Carlos Eduardo, Sílvio Leão, and Sílvio Leão. Breast tumor segmentation in ultrasound images: comparing U-net and U-net++. *Research on Biomedical Engineering*, January 2025. Accessed: 2024-03-04.

Chukwuebuka Joseph Ejiyi, Zhen Qin, Ariyo Oluwasanmi, Mugahed A Al-antari, and Olusola Bamisile. A unified 2d medical image segmentation network (segmentnet) through distance-awareness and local feature extraction. *Biocybernetics and Biomedical Engineering*, September 2024.

R. A. Fisher. Statistical methods for research workers. *Pediatric pharmacology (New York, NY)*, 1925.

Z. Fu and J. Li. Deep supervision feature refinement attention network for medical image segmentation. *Engineering Applications of Artificial Intelligence*, June 2023. Accessed: 2024-03-02.

Mary M. Galloway. Texture analysis using gray level run lengths. *IEEE Transactions on Systems, Man, and Cybernetics*, June 1975. Accessed: 2024-03-06.

Albert Gu and Tri Dao. Mamba: Linear-Time Sequence Modeling with Selective State Spaces. *Machine Learning (cs.LG)*, May 2024. Accessed: 2024-03-02.

Robert M. Haralick, K. Shanmugam, and K. Shanmugam. Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics*, November 1973. Accessed: 2024-03-06.

K. He, X. Zhang, and Shaoqing Ren. Deep Residual Learning for Image Recognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016. Accessed: 2024-03-06.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, December 2015. Accessed: 2024-03-09.

Gao Huang, Zhuang Liu, and Kilian Q. Weinberger. Densely Connected Convolutional Networks. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, December 2018. Accessed: 2024-03-09.

International Society of Ultrasound in Obstetrics and Gynecology. Practice Guidelines for Performance of the Routine Mid-Trimester Fetal Ultrasound Scan. *ISUOG*, June 2022. Accessed: 2024-02-23.

M. Kass and A. Witkin. Snakes: Active contour models. *International Journal of Computer Vision*, January 1988. Accessed: 2024-03-02.

Tricia Kinman. Fetal Echocardiography: What It Is, When It's Done, and What to Expect. *Healthline*, July 2018. Accessed: 2024-02-23.

Kriti, Jitendra Virmani, and Ravinder Agarwal. Deep feature extraction and classification of breast ultrasound images. *Multimedia Tools and Applications*, 2020.

ZhiFei Lai and HuiFang Deng. Medical Image Classification Based on Deep Features Extracted by Deep Model and Statistic Feature Fusion with Multilayer Perceptron. *Comput Intell Neurosci.*, September 2018. Accessed: 2024-03-14.

C.-Y. Lee and S. Xie. Deeply-supervised nets. *Proc. 18th Int. Conf. Artif. Intell. Stat. (AISTATS)*, June 2015. Accessed: 2024-03-02.

Wesley Lee, Scott Barton, and Christine. Transverse cerebellar diameter: A useful predictor of gestational age for fetuses with asymmetric growth retardation. *American Journal of Obstetrics and Gynecology*, May 2013. Accessed: 2024-02-23.

S. Limperopoulos and E. Robertson. The cerebellum's role in fetal development: Motor and cognitive implications. *Seminars in Fetal and Neonatal Medicine*, June 2016. Accessed: 2024-03-02.

Pengfei Liu, Huaici Zhao, and Huaici Zhao. Automated classification and measurement of fetal ultrasound images with attention feature pyramid network. *Ultrasound Q.*, May 2020. Accessed: 2024-03-04.

Y. Liu, Y. Tian, Y. Zhao, and H. Yu. VMamba: Visual State Space Model. *Medical Image Analysis (DLMIA) Workshop*, July 2024. Accessed: 2024-02-23.

Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully Convolutional Networks for Semantic Segmentation. *CVPR*, March 2014. Accessed: 2024-03-02.

Mohammed A. Maraci, Mohammad Yaqub, and Mohammad Yaqub. Toward point-of-care ultrasound estimation of fetal gestational age from the trans-cerebellar diameter using CNN-based ultrasound image analysis. *J Med Imaging (Bellingham)*, May 2021. Accessed: 2024-03-04.

Marion, MD, and Cathy Cartwright. Hydrocephalus in Pregnancy. *Hydrocephalus Association*, February 2023. Accessed: 2024-02-23.

Raghav Mehta, Tal Arbel, and Tal Arbel. 3D U-Net for Brain Tumour Segmentation. *nternational MICCAI Brainlesion Workshop*, January 2019. Accessed: 2024-03-02.

Joel Than Chia Ming, Norliza Mohd Noor, and Omar Mohd Rijal. Lung disease classification using glcm and deep features from different deep learning architectures with principal component analysis. *2018 2nd International Conference on BioSignal Analysis, Processing and Systems (ICBAPS)*, July 2018.

T. Mul, M. Mongelli, and J. Gardosi. A comparative analysis of second-trimester ultrasound dating formulae in pregnancies conceived with artificial reproductive techniques. *Ultrasound Obstet Gynecol.*, December 1996. Accessed: 2024-03-09.

Ozan Oktay, Jo Schlemper, and Loic Le Folgoc. Attention U-Net: Learning Where to Look for the Pancreas. *MIDL'18*, May 2018. Accessed: 2024-03-02.

Karl Pearson. Note on regression and inheritance in the case of two parents. *Royal Society*, 1895.

Queensland Health. Ultrasounds during pregnancy. 2020. Accessed: 2025-03-10.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. *MICCAI*, July 2015. Accessed: 2024-02-23.

Olga Russakovsky, Jia Deng, and Hao Su. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, April 2015. Accessed: 2024-03-09.

Michael Schumacher, Philippe Liere, and Abdelmoumen Ghoumari. Progesterone and fetal-neonatal neuroprotection. *Best Pract Res Clin Obstet Gynaecol.*, September 2020. Accessed: 2024-02-23.

Julia A Scott, Piotr A Habas, and Kio Kim. Growth trajectories of the human fetal brain tissues estimated from 3D reconstructed in utero MRI. *International Journal of Developmental Neuroscience*, May 2012. Accessed: 2024-03-14.

A. Shah. Breast ultrasound images dataset. *Kaggle*, June 2019. Accessed: 2024-03-02.

J. Shang and Y. Lai. An Optimal U-Net++ Segmentation Method for Dataset BUSI. *Seminar on Artificial Intelligence, Networking and Information Technology (AINIT)*, March 2024. Accessed: 2024-02-27.

Karen Simonyan and Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. *3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, October 2015. Accessed: 2024-03-09.

Aryan Singh, Pepijn Van de Ven, and Pepijn Van. Image Segmentation: Inducing graph-based learning. *arXiv preprint arXiv:2501.03765v2*, January 2025. Accessed: 2024-03-04.

Vishal Singh, Pradeeba Sridar, and Pradeeba Sridar. Semantic Segmentation of Cerebellum in 2D Fetal Ultrasound Brain Images Using Convolutional Neural Networks. *IEEE Access*, May 2021. Accessed: 2024-03-04.

Pradeeba Sridar, Ashnil Kumar, and Changyang Li. Automatic Measurement of Thalamic Diameter in 2-D Fetal Ultrasound Brain Images Using Shape Prior Constrained Regularized Level Sets . *IEEE J Biomed Health Inform.*, July 2017. Accessed: 2024-03-09.

Pradeeba Sridar, Narelle J Kennedy, and Jinman Kim. Normative ultrasound data of the fetal transverse thalamic diameter derived from 18 to 22 weeks of gestation in routine second-trimester morphology examinations. *Australas J Ultrasound Med.*, 2(4):349–356, 2020a.

Pradeeba Sridar, Narelle J Kennedy, Ann E Quinton, Kristy Robledo, Jinman Kim, and Ralph Nanan. Normative ultrasound data of the fetal transverse thalamic diameter derived from 18 to 22 weeks of gestation in routine second-trimester morphology examinations. *Australas J Ultrasound Med*, January 2020b. Accessed: 2024-02-23.

Pradeeba Sridar, Ashnil Kumar, and Ashnil Kumar. Automatic Measurement of Thalamic Diameter in 2-D Fetal Ultrasound Brain Images Using Shape Prior Constrained Regularized Level Sets. *IEEE J Biomed Health Inform.*, May 2021. Accessed: 2024-03-04.

Z. Tang and H. Zhang. DBU-Net: Dual Branch U-Net for Tumor Segmentation in Breast Ultrasound Images. *PLOS ONE*, October 2023. Accessed: 2024-03-02.

Guillaume Thibault, Jesús Angulo, and Fernand Meyer. Advanced statistical matrices for texture characterization: application to cell classification . *IEEE Transactions on Systems, Man, and Cybernetics*, November 2013. Accessed: 2024-03-06.

Helena R. Torres, Helena R. Torres, and Bruno Oliveira. A review of image processing methods for fetal head and brain analysis in ultrasound images. *omput Methods Programs Biomed.*, January 2022. Accessed: 2024-03-09.

Vathy and Ilona. Prenatal opiate exposure: long-term CNS consequences in the stress system of the offspring. December 2002.

Villiger, JW, Taylor, KM, and Gluckman. Ontogenesis of opiate receptors in regions of the ovine brain. *Pediatr Pharmacol (New York).*, 2(4):349–356, 1982.

Yuanbo Wang, Unaiza Ahsan, and Hanyan Li. A comprehensive review of modern object segmentation approaches. *Now Foundations and Trends*, 2023.

Z. Wang, J.-Q. Zheng, and Y. Zhang. Mamba-unet: Unet-like pure visual mamba for medical image segmentation. arXiv preprint arXiv:2402.05079. 2024. Accessed: 2024-02-23.

Joan S. Weszka, Charles R. Dyer, and Charles. A Comparative Study of Texture Measures for Terrain Classification. *IEEE Transactions on Systems, Man, and Cybernetics*, November 1976. Accessed: 2024-03-06.

Hai Ye, Kaiping Feng, and Yueyue Wang. Ultrasonic Image Segmentation Method based on Improved Fully Convolution Network. *ICMEIT 2019*, May 2019. Accessed: 2024-03-04.

J. Zhang Z. Wang. Light-Convolution Dense Selection U-Net (LDS U-Net) for Ultrasound Spine Segmentation. *Medical Imaging: Advanced Techniques and Applications*, November 2021. Accessed: 2024-03-02.

Jing Zhang, Longzheng Tong, and Lei Wang. Texture analysis of multiple sclerosis: a comparative study . *Magnetic Resonance Imaging*, October 2008. Accessed: 2024-03-07.

Y. Zhang, Y. Wang, and J. Zhang. AAU-Net: An Adaptive Attention U-Net for Breast Lesions Segmentation in Ultrasound Images. *IEEE*, March 2022. Accessed: 2024-02-27.

Chen Zhao, Jungang Han, and Jungang Han. Lung Nodule Detection via 3D U-Net and Contextual Convolutional Neural Network. *International Conference on Networking and Network Applications (NaNA)*, January 2018. Accessed: 2024-03-02.

T. Zhao and H. Dai. Breast tumor ultrasound image segmentation method based on improved residual U-Net network. *Computational Intelligence and Neuroscience*, June 2022. Accessed: 2024-03-02.

Zongwei Zhou and Md Mahfuzur Rahman Siddiquee. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. *Medical Image Analysis (DLMIA) Workshop*, July 2018. Accessed: 2024-02-23.