

A DISSERTATION
SUBMITTED IN FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
IN COMPUTER SCIENCE AND ENGINEERING

**Deep Learning Based Segmentation and
Classification of MRI and Ultrasound Medical
Images**



by

Peng Boyuan

© Copyright by Peng Boyuan,

All Rights Reserved.

The thesis titled

*Deep Learning Application for Segmentation and Classification of
Medical Images*

by

Peng Boyuan

is reviewed and approved by:

Chief referee

Senior Associate Professor

Date

Huang Jie

Jie Huang

2025.1.10

Professor

Date

YOSHIOKA Rentaro

R. Yoshio

2025.1.10

Senior Associate Professor

Date

JING Lei

Le Jing

2025.1.10

Senior Associate Professor

Date

HASHIMOTO Yasuhiro

橋本康弘

2025.1.10

Professor

Date

Zhu Xin

朱欣

朱

2025.1.11

THE UNIVERSITY OF AIZU

Contents

Chapter 1 Background	1
1.1 Medical Image Background	1
1.2 Deep Learning Applications in Medical Imaging	2
1.3 MRI Imaging	2
1.4 Ultrasound Imaging	3
1.5 Overview of the Dissertation	4
1.6 Summary	5
Chapter 2 Deep Learning Application for Femoral MRI Images Processing(Segmentation)	7
2.1 Introduction	7
2.2 Methods	8
2.2.1 Data	8
2.2.2 Experimental procedure	8
Division of the data set	8
Models	8
2.2.3 Experimental evaluation	10
2.3 Results	10
2.4 Discussion	12
2.5 Conclusion	13
Chapter 3 Deep Learning Application for Femoral MRI Images Processing(Classification)	14
3.1 Introduction	14
3.2 Materials and Methods	16
3.2.1 Data	16
3.2.2 Experimental Design	16
3.2.3 Experiment 1	17
3.2.4 Experiment 2 Segmentation Followed by Classification	17
3.2.5 Experiment 3 Segmentation, Unilateral Femur Extraction, and Classification	20
3.2.6 Experiment 4 Direct Classification Using SAMSCNet Model	21
Branch Two: Classification Pathway	21
3.2.7 Evaluation Measures	23
3.3 Results	23
3.3.1 Overview of experimental results	23
3.3.2 Summary of the results of the four experiments	24
Experiment One: Direct Classification of Original Images	24
Experiment Two: Segmentation Followed by Classification	25

Experiment Three: Segmentation, Unilateral Femur Extraction, and Classification	25
Experiment Four: Direct Classification Using Novel Model . . .	25
3.3.3 Models Results	25
3.4 Discussion and Future work	27
3.4.1 Contribution	27
3.4.2 Limitations	29
3.4.3 Future Work	29
3.5 Conclusion	29
Chapter 4 Deep Learning Application for Transvaginal Uterine Ultrasound Images Processing(Semi-Supervised Segmentation, BCP-Mamba)	31
4.1 Introduction	31
4.2 Related Work	32
4.3 Methodology	33
4.3.1 Data Preprocessing	34
4.3.2 Bidirectional Copy-Paste (BCP) Operation	34
4.3.3 Experimental Details	34
4.3.4 Experiments	37
4.4 Results	38
4.5 Discussion	39
4.6 Conclusion	42
Chapter 5 Deep Learning Application for Transvaginal Uterine Ultrasound Images Processing(Semi-Supervised Segmentation, Multi-StudentNet)	44
5.1 Introduction	44
5.2 Methods	45
5.2.1 Dataset and preprocessing	45
5.2.2 Experimental setup	47
Division of the data set	47
Teacher Models Training	47
Student Models Training	49
Evaluation Indicators	49
5.3 Result	49
5.4 Discussion	51
5.5 Conclusion	52
Chapter 6 Deep Learning Application for Transvaginal Ultrasound Video Images Processing(Keyframe extraction)	53
6.1 Introduction	53
6.2 Related work	54
6.3 Methods	54
6.3.1 Study Design and Ethics	54
6.3.2 Dataset Collection and Processing	55
6.3.3 Segmentation	57
Data	57
Models	57
Evaluation standard	59
6.3.4 Classification	59

6.3.5	Keyframes	64
6.3.6	Comparative Experiment	66
6.3.7	Evaluation parameters	66
6.4	Results	67
6.4.1	Endometrial Segmentation Performance	67
6.4.2	Classification of Segmentation Accuracy	68
6.4.3	Keyframes and Experiment	68
	Keyframe Recognition	69
	Comparative experiments	69
	Keyframes and Experiment	69
6.4.4	The results of the comparison experiment	70
6.5	Discussion	71
6.6	Conclusion	73
Chapter 7	Discussion	74
7.1	Unified Vision: Advancing Medical Imaging Through Deep Learning .	74
7.2	Methodological Innovations	74
7.2.1	Unified Vision: Advancing Medical Imaging Through Deep Learning	75
7.2.2	Methodological Innovations	75
7.3	Challenges and Lessons Learned	76
7.4	Broader Implications	76
7.5	Summary of Doctoral Contributions	76
Chapter 8	Conclusion	78

List of Figures

Figure 2.1 Network structure diagram of PP-LiteSeg.	9
Figure 2.2 Example of results for the same set of femoral data in different models.	11
Figure 3.1 Study flowchart	16
Figure 3.2 Three Different Types of Femur	18
Figure 3.3 Flowchart of ASEC Method	19
Figure 3.4 Flowchart of SAMSCNet Method	22
Figure 3.5 Examples of Femoral Segmentation Results	24
Figure 3.6 Original Confusion Matrix	26
Figure 3.7 Segmentation Confusion Matrix	26
Figure 3.8 Unilateral Confusion Matrix	26
Figure 3.9 Example of Prediction Error	26
Figure 3.10 Results with Grad-CAM	28
Figure 4.1 Flowchart of the bidirectional copy-paste Mamba (BCP-Mamba) model	33
Figure 4.2 Mamba-UNet network model structure	35
Figure 4.3 Simple pyramid pooling module (SPPM) and Vision Mamba (VM) block	36
Figure 4.4 Results of the BCP-Net, UNet, BCP-Mamba models. Ground truth (green), predicted results (red), and the part of overlap between the ground truth and predicted results (yellow).	40
Figure 4.5 Example of endometrial prediction error	41
Figure 5.1 Dataset utilization flowchart for semi-supervised segmentation.	46
Figure 5.2 Multi-StudentNet Framework for Endometrial Segmentation.	48
Figure 5.3 Example of results for the TVUS data in different models.	50
Figure 6.1 Absolute endometrial thickness errors measured by the proposed method	55
Figure 6.2 Methods of processing	56
Figure 6.3 Schematic diagram of the structure of Resnet50-unet model modified from [1]	60
Figure 6.4 Schematic diagram of the structure of DRRNets model modified from [2]	61
Figure 6.5 Segmentation results of the same ultrasound image by six different networks. a is the original image, b is the ground truth of endometrial boundary, and c-h are the segmentation results of Resnet50_U-Net, Resnet50_segnet, U-Net, U-Net_mini, Vgg16_segnet, and DRRNets respectively.	62

Figure 6.6 keyframe	65
Figure 6.7 The results of the comparison experiment	70

List of Tables

Table 2.1	Average evaluation parameters of different models.	12
Table 3.1	Details of Training, Validation, and Test Datasets for Classification	20
Table 4.1	Comparison of evaluation parameters of the model.	43
Table 5.1	Comparison of evaluation indexes(SD) of the model.	49
Table 6.1	Five models for uterine ultrasound image segmentation.	59
Table 6.2	Evaluation Parameters.	68

Acknowledgment

I would like to express my deepest gratitude to my supervisor, Xin Zhu and Jie Huang, for his invaluable guidance, unwavering support, and insightful feedback throughout this research. His expertise and mentorship have been instrumental in shaping the direction of this work and fostering my intellectual growth as a researcher.

The accomplishment of this thesis wouldn't have been possible without the support and assistance of all esteemed professors. Special gratitude is extended to Professors Lei Jing, Rentaro Yoshioka and Yasuhiro Hashimoto. Their expert guidance, contributions to my research topic, and tireless help in navigating through numerous challenges were invaluable.

I extend my appreciation to the University of Aizu, for providing the resources and facilities necessary for carrying out this study. The collaborative environment and access to cutting-edge technologies have significantly contributed to the success of this research endeavor.

My heartfelt appreciation goes out to my friends at the University of Aizu. The experience of learning, working, and engaging in recreational activities with you all over these past few years has been genuinely thrilling. Wishing every one of you a future filled with success and prosperity.

Furthermore, I am indebted to the medical professionals and researchers who generously contributed their expertise and provided access to the clinical data essential for this study. Their dedication to advancing medical knowledge has been truly inspiring.

Last but not least, I am deeply grateful to my family and friends for their unwavering encouragement, understanding, and patience throughout this journey. Their love and support have been my source of strength and motivation.

Abstract

As the global demand for medical imaging technologies continues to grow, the rapid development of deep learning has provided new solutions for enhancing early disease diagnosis and treatment efficiency. Traditional imaging analysis methods, especially in tumor detection and staging, often face challenges such as high subjectivity, inadequate accuracy, and low processing efficiency. Therefore, developing more precise and efficient automated imaging analysis tools has become a critical focus of modern medical research. This dissertation explores two significant research projects: the analysis of femoral MRI images in patients with Non-Hodgkin's lymphoma (NHL) and the segmentation and measurement of endometrial thickness in ultrasound images, aiming to leverage deep learning technologies to advance clinical practices.

In the femoral project, we first conducted precise segmentation of femoral MRI images. This study compared the performance of four neural networks—PP-LiteSeg, U-Net, SegNet, and PspNet—using 579 MRI images from 200 patients for training and testing. The experimental results indicated that PP-LiteSeg achieved an impressive average Dice coefficient of 0.92, outperforming other models and demonstrating its potential for accurate segmentation. This phase laid a solid foundation for subsequent classification tasks. We then introduced SAMSCNet, an innovative deep learning model that integrates segmentation and classification tasks. By utilizing DenseNet169 as the encoder and employing a dual-branch strategy, SAMSCNet effectively captures both global and local features. This model not only addresses issues related to background interference and computational complexity but also significantly enhances classification accuracy, ultimately improving prognostic evaluation for NHL patients.

In the endometrial project, we focused on the segmentation of endometrial boundaries, thickness measurement, and keyframe extraction. By comparing various deep learning models, we identified deep dual-resolution networks as the optimal approach for accurate segmentation, achieving an average Dice coefficient of 0.895. Additionally, the dissertation proposes a novel automated method for measuring endometrial thickness, ensuring that 89.3% of errors fall within the clinically accepted range of ± 2 mm. To further enhance the consistency of image analysis, we developed a keyframe extraction framework aimed at standardizing assessments and mitigating the impact of variability in sonographer expertise.

In summary, this dissertation demonstrates the immense potential of deep learning in medical imaging analysis, emphasizing its critical applications in diagnosing Non-Hodgkin's lymphoma and endometrial diseases. By introducing innovative models and methodologies, we not only advance the automation of imaging analysis but also provide more reliable support for clinical decision-making. These research findings will serve as an important theoretical foundation and practical guidance for the future development of medical imaging technologies.

Chapter 1

Background

1.1 Medical Image Background

Medical imaging technologies are indispensable in modern clinical diagnostics and treatment planning [3]. From MRI (Magnetic Resonance Imaging) and CT (Computed Tomography) to ultrasound imaging, each modality serves a unique purpose in evaluating various physiological and pathological conditions [4]. MRI, renowned for its high-resolution imaging of soft tissues, plays a critical role in diagnosing tumors, musculoskeletal disorders, and neurological conditions [5] [6]. Ultrasound, with its real-time imaging capabilities, non-invasive nature, and cost-effectiveness, is widely utilized in cardiovascular, abdominal, and gynecological assessments [7]. CT, on the other hand, is celebrated for its rapid imaging capabilities and high-resolution output, making it essential in trauma, cancer screening, and pulmonary disease evaluations [8]. PET (Positron Emission Tomography), often integrated with CT or MRI, provides functional and metabolic insights critical for oncology and neurological applications [9].

Despite their clinical utility, traditional methods of interpreting medical images rely heavily on manual analysis by radiologists and clinicians. This approach is inherently subjective, time-consuming, and prone to variability [10] [11]. Studies indicate that manual interpretation can result in false-negative rates as high as 10-20% in some cases, highlighting the need for more consistent diagnostic methods [12]. Additionally, inter-observer variability remains a significant concern, particularly in complex cases like musculoskeletal or neurological disorders [13].

To address these challenges, the field has seen a paradigm shift towards computer-assisted technologies. Among these, deep learning has emerged as a transformative approach, offering innovative solutions for automating complex image analysis tasks across various imaging modalities. For instance, convolutional neural networks (CNNs) have been instrumental in detecting pulmonary nodules and segmenting liver tumors, achieving sensitivities and accuracies that exceed traditional manual methods [14]. However, challenges such as data privacy, ethical concerns, and the need for high-quality annotated datasets continue to limit the widespread adoption of AI in clinical settings [15] [16].

1.2 Deep Learning Applications in Medical Imaging

Deep learning, particularly convolutional neural networks (CNNs), has revolutionized the field of medical image analysis by enabling machines to learn complex patterns directly from data. Unlike traditional methods that rely on handcrafted features, deep learning models are capable of end-to-end training, automatically extracting meaningful representations from images. This has led to significant advancements in tasks such as image classification, segmentation, and detection [17].

In clinical settings, deep learning has already demonstrated remarkable success. For instance, in breast cancer screening, deep learning algorithms have achieved higher sensitivity and specificity than traditional approaches. A study by McKinney et al. (2020) demonstrated that a deep learning model outperformed radiologists in breast cancer detection by reducing both false positives and false negatives [18]. Similarly, in lung nodule detection, CNNs have significantly reduced false-positive rates while maintaining high accuracy, as evidenced by studies focusing on automated CT scan analysis for pulmonary nodules [19]. These successes highlight the adaptability of deep learning to diverse imaging modalities, including MRI, CT, and ultrasound, and its potential to improve diagnostic workflows.

Deep learning has also shown promise in segmentation tasks, such as tumor delineation in MRI scans and organ segmentation in CT imaging. For example, the U-Net architecture has become a benchmark for medical image segmentation due to its ability to handle limited datasets effectively while achieving high accuracy [1]. In liver tumor segmentation, models based on U-Net and its variants have achieved state-of-the-art performance by incorporating attention mechanisms and multi-scale features [20].

However, the complexity and variability of medical images pose unique challenges. Each imaging modality has distinct characteristics—MRI offers high spatial resolution and multi-sequence data, while ultrasound is characterized by low resolution and noise artifacts. Tailored deep learning solutions are therefore essential to address these modality-specific challenges and maximize clinical utility. For example, domain adaptation techniques and noise-robust architectures have been developed to improve deep learning performance in noisy ultrasound images [21]. In addition, multimodal approaches combining information from different imaging techniques, such as PET-CT or PET-MRI, have further enhanced diagnostic accuracy by leveraging complementary data [22].

Finally, the deployment of deep learning models in clinical practice faces hurdles such as data privacy, model interpretability, and the need for extensive computational resources. Federated learning and explainable AI have emerged as promising solutions to address these issues, fostering wider adoption of deep learning in healthcare [23].

1.3 MRI Imaging

MRI is a cornerstone in medical imaging, particularly for its ability to provide detailed insights into soft tissue structures. In the context of Non-Hodgkin’s Lymphoma (NHL), femoral MRI scans offer valuable information about bone marrow involvement, aiding in disease staging and prognosis. MRI’s multimodal nature, including T1-weighted, T2-weighted, and diffusion-weighted imaging (DWI), provides complementary information crucial for assessing bone marrow involvement in NHL. For instance, T1-weighted imaging is highly sensitive to marrow replacement, while DWI excels in

detecting early-stage lesions by highlighting diffusion restriction [24]. However, manual interpretation of MRI images is often time-intensive and subject to inter-observer variability, limiting its scalability in clinical practice.

This research focuses on leveraging deep learning to enhance MRI image analysis. In our femoral MRI project, we developed an innovative deep learning model, SAMSCNet, which integrates segmentation and classification tasks. This dual-branch architecture, with DenseNet169 as the encoder, effectively captures both global and local features, enabling precise delineation of the femoral region and accurate classification of disease states. The SAMSCNet model demonstrated superior performance, achieving a classification accuracy of 95.2% and a segmentation Dice coefficient of 0.92 on the test set. Compared to U-Net and SegNet, SAMSCNet showed a 12% improvement in sensitivity for detecting bone marrow involvement, addressing key clinical needs for early and accurate NHL staging [?].

Developing SAMSCNet required a robust dataset of femoral MRI scans, meticulously annotated by experienced radiologists. The dataset comprised over 2,000 scans, including various disease states and imaging modalities, ensuring model generalizability and robustness. Semi-automated tools like ITK-SNAP were utilized to streamline the annotation process, significantly reducing manual effort [25].

Beyond NHL, the architecture of SAMSCNet could be extended to other clinical applications, such as detecting metastases in solid tumors or assessing inflammatory conditions in the femoral bone. The model's ability to handle high variability in imaging data makes it a promising tool for broader clinical applications. Integration of SAMSCNet into clinical workflows involves real-time processing of MRI scans, providing automated segmentation and classification outputs that can assist radiologists in decision-making. This reduces the workload and minimizes diagnostic delays. Additionally, explainability techniques, such as saliency maps, were incorporated to ensure clinicians could trust and interpret the model's predictions [26].

1.4 Ultrasound Imaging

Ultrasound imaging is a cornerstone of real-time diagnostic evaluation, particularly in fields like gynecology and obstetrics. One critical application is the measurement of endometrial thickness, which plays a pivotal role in assessing uterine health and diagnosing conditions such as infertility and endometrial abnormalities. Ultrasound's non-invasive nature, real-time imaging, and cost-effectiveness make it an invaluable tool in routine clinical practice. However, despite these advantages, ultrasound images are prone to inherent variability due to factors such as patient body type, operator-dependent expertise, and equipment settings, leading to inconsistent and often subjective assessments.

To address these challenges, this research developed a comprehensive pipeline for automated endometrial analysis, incorporating cutting-edge deep learning techniques to improve the accuracy and consistency of ultrasound image interpretation. A deep dual-resolution network was employed to achieve highly accurate segmentation of endometrial boundaries, achieving an average Dice coefficient of 0.895, demonstrating the model's robustness in delineating the uterine layers. This dual-resolution approach allowed the model to capture both fine details at the local level and broader structural information at the global level, effectively enhancing the segmentation performance.

In addition to segmentation, we proposed a novel automated method for endometrial thickness measurement. This approach ensured that 89.3% of measurement errors remained within a clinically acceptable range of ± 2 mm, which is crucial for ensuring diagnostic reliability and reproducibility. This threshold aligns with established clinical standards for assessing uterine health, making the tool directly applicable to real-world clinical environments.

To further standardize the evaluation process and mitigate the variability introduced by differences in sonographer technique, we introduced a keyframe extraction framework. This framework automatically selects representative frames from the ultrasound video sequence, reducing the impact of motion artifacts and inconsistent image quality. By focusing on keyframes that best represent the endometrial boundaries, we ensured more consistent measurements and enhanced reproducibility across different operators. This method not only streamlines the analysis workflow but also offers a more objective approach to ultrasound-based assessments, thereby reducing diagnostic errors and improving clinical decision-making.

By integrating these innovations, our approach paves the way for more standardized, efficient, and reliable ultrasound evaluations, providing healthcare providers with a powerful tool for the assessment of endometrial health.

1.5 Overview of the Dissertation

This dissertation explores the application of deep learning in medical imaging, addressing critical challenges across MRI and ultrasound modalities. It focuses on segmentation, classification, and keyframe extraction tasks, with each chapter presenting a research project that builds on the previous one to demonstrate methodological advancements and clinical relevance.

Chapter 1: Femoral Segmentation of MRI Images Using PP-LiteSeg This chapter lays the foundation for the dissertation by introducing PP-LiteSeg, a lightweight segmentation model for femoral MRI images. Segmentation of the femoral region is crucial for analyzing Non-Hodgkin's Lymphoma (NHL) and other conditions affecting the bone marrow. Using a dataset of 579 MRI images, the study evaluates PP-LiteSeg against established models like U-Net and SegNet. Experimental results highlight the superior performance of PP-LiteSeg, achieving a Dice coefficient of 0.92. This chapter provides the groundwork for integrating segmentation with classification tasks in the subsequent chapter.

Chapter 2: A Novel Dual-Branch Deep Learning Model for Segmentation and Classification in Femoral MRI Analysis for Patients with Non-Hodgkin's Lymphoma Building upon Chapter 1, this chapter introduces SAMSCNet, a dual-branch deep learning model designed to integrate segmentation and classification tasks for femoral MRI images. By utilizing a DenseNet169-based encoder and a dual-branch architecture, SAMSCNet effectively captures global and local features, addressing challenges like background interference and computational complexity. The model enhances diagnostic accuracy and staging precision for NHL patients, providing a robust framework for advancing MRI-based clinical applications.

Chapter 3: Bidirectional Copy-Paste Mamba for Enhanced Semi-Supervised Segmentation of Transvaginal Uterine Ultrasound Images Transitioning to ultrasound imaging, this chapter presents the Bidirectional Copy-Paste Mamba (BCP-Mamba) model

for semi-supervised segmentation of the parametrium in transvaginal ultrasound images. Ultrasound imaging poses unique challenges, including diverse textures and limited pixel-level annotations. The BCP-Mamba model addresses these issues through a novel visual state space (VSS) module integrated into a U-shaped architecture. Using a dataset of 1,940 transvaginal ultrasound images, the model achieves a Dice coefficient of 86.55%, outperforming U-Net and BCP-Net. This approach significantly reduces the reliance on manual annotations while maintaining high segmentation accuracy, alleviating the workload for clinical experts.

Chapter 4: Keyframe Extraction from Transvaginal Ultrasound Video Images Using Deep Learning This chapter focuses on automated keyframe extraction for endometrial thickness (ET) measurement from transvaginal ultrasound videos. Accurate ET measurement is essential for diagnosing gynecological conditions but can be challenging for less experienced sonographers. This study introduces EndoUSScan, a system combining MSNet, a DenseNet169-based candidate image selection model, with a keyframe detection mechanism. Evaluations involving 976 videos and 82,000 images demonstrate that EndoUSScan achieves 94.7% accuracy and significantly improves ET measurement speed and consistency compared to junior sonographers. The system bridges the expertise gap in clinical settings, enhancing diagnostic workflows.

This dissertation systematically addresses the challenges of segmentation, classification, and keyframe extraction in MRI and ultrasound imaging. Starting with foundational work on femoral segmentation in MRI (Chapter 1), it progresses to advanced dual-branch models for segmentation and classification (Chapter 2). The focus then shifts to ultrasound imaging, with semi-supervised segmentation (Chapter 3) and keyframe extraction for endometrial thickness measurement (Chapter 4).

Through the development of innovative models like PP-LiteSeg, SAMSCNet, BCP-Mamba, and EndoUSScan, this dissertation demonstrates the versatility and clinical potential of deep learning. By addressing both technical and clinical challenges, the research not only enhances diagnostic accuracy and efficiency but also provides scalable, practical solutions for real-world medical imaging applications. These contributions represent a significant advancement in the integration of artificial intelligence into healthcare, offering a strong foundation for future research and clinical adoption.

1.6 Summary

This dissertation bridges the gap between advanced deep learning techniques and practical medical imaging applications, with a focus on MRI and ultrasound modalities. Through two distinct projects, it addresses critical challenges in medical imaging: precise segmentation and classification in MRI for Non-Hodgkin's Lymphoma (NHL), and automated thickness measurement and keyframe extraction in ultrasound for gynecological applications. These studies highlight the adaptability of deep learning to diverse imaging contexts, paving the way for more standardized and reliable diagnostic workflows.

In the MRI project, the development of the SAMSCNet model demonstrated the ability to integrate segmentation and classification tasks within a unified architecture, leveraging DenseNet169 for feature extraction. This innovation improved diagnostic accuracy, particularly in staging NHL, where detailed analysis of femoral bone marrow is critical. The model's dual-branch design, capable of capturing both global and local

image features, effectively addressed challenges such as background noise and variability in MRI datasets. These advancements not only contribute to more efficient and accurate NHL diagnosis but also showcase the potential for deep learning to transform other areas of MRI-based diagnostics.

For ultrasound imaging, the proposed automated pipeline tackled the inherent variability and operator dependency of traditional assessments. By employing a dual-resolution network for endometrial segmentation, the study achieved a high Dice coefficient, ensuring precise delineation of anatomical boundaries. Furthermore, the introduction of a novel automated thickness measurement method reduced errors to clinically acceptable levels, while the keyframe extraction framework standardized the analysis process. These contributions underscore the capability of deep learning to enhance diagnostic consistency and reliability in real-time ultrasound applications.

Across both projects, a key strength of this work lies in its emphasis on modality-specific optimizations. By tailoring models to the unique characteristics of MRI and ultrasound, the research demonstrated how deep learning can overcome challenges posed by imaging artifacts, noise, and variability. This approach ensures not only high performance but also clinical relevance, making the models directly applicable in real-world healthcare settings.

Beyond addressing immediate challenges, this dissertation lays the groundwork for future advancements in medical imaging and diagnostics. The innovative methodologies developed here can serve as a blueprint for extending deep learning applications to other imaging modalities and clinical domains. For instance, multimodal approaches that integrate MRI, ultrasound, and other techniques, such as CT or PET, hold promise for even more comprehensive diagnostic solutions. Furthermore, the focus on automation and standardization aligns with the growing trend toward personalized medicine, where AI-driven tools can provide tailored insights based on individual patient data.

These contributions underscore the transformative potential of artificial intelligence in modern healthcare. By enhancing diagnostic accuracy, consistency, and efficiency, this work demonstrates how deep learning can address the complex demands of contemporary medical imaging, ultimately improving patient outcomes and supporting the transition toward a more automated, data-driven healthcare paradigm.

Chapter 2

Deep Learning Application for Femoral MRI Images Processing(Segmentation)

2.1 Introduction

Recently, the incidence of tumors around the world is increasing annually. In addition to the well-known solid tumors such as liver cancer, stomach cancer, and lung cancer, a variety of hematological malignancies such as lymphoma, leukemia and multiple myeloma have increasing incidences around the world [27].

Lymphomas are malignant tumors that originate in lymph nodes or lymphoid tissues. Lymphomas are mainly classified as non-Hodgkin's lymphoma and Hodgkin's lymphoma. A common symptom of both types of lymphoma is painless swelling of one or more lymph glands. Hodgkin's lymphoma is more common in the neck, armpits, and chest, while non-Hodgkin's lymphoma is usually found in lymph nodes throughout the body [28].

Currently, clinical practice relies on routine site bone marrow biopsy to diagnose lymphoma. However, biopsy is an invasive operation with sampling errors. MRI can compensate for the deficiency of biopsy and improve the detection rate of lymphoma. In addition, MRI may have predictive value for the prognosis of lymphoma patients [29].

However, using MRI images to analyze hematological malignancies is not only time-consuming and labor-intensive but also requires extensive experience. Diagnosis is essential for effective treatment of Hematological malignancies. With the rapid development of artificial intelligence, deep learning have been widely used in computer-aided detection and analysis. Accurate segmentation of femur from MRI images is crucial to the analysis of femoral marrow. Yun et al. proposed a fully automated method for femoral segmentation using pelvic CT, and their method was accurate and improved the subsequent measurement of fracture determination [30]. Yue et al. analyzed the role of whole-body MRI versus PET/CT in the diagnosis and prognosis of bone marrow infiltration in lymphoma [31]. In our previous work, we also proposed a method for the automatic segmentation of femur [32]. PP-LiteSeg is a novel lightweight model for the real-time semantic segmentation task. PP-LiteSeg incorporates an attention mechanism module that generates attention weights and fuses input features with the weights [33].

In this research, we proposed a new method based on PP-LiteSeg for femur segmentation in MRI images from patients with hematological malignancies. In addition,

we compared the segmentation results of PP-LiteSeg with those of traditional CNN models.

2.2 Methods

2.2.1 Data

200 patients (579 MRI image files) with hematologic malignancies (109 males, 91 females, mean age 55-85) were consecutively recruited from Aizu Medical Center, Fukushima Medical University in 2012-2020. The data are T1-enhanced images acquired by a 1.5T MRI machine (MAGNETOM Avanto Siemens) using a T1_se_cor (T1 Turbo spin-echo Coronal) sequence on a tomographic scan on the coronal plane of femur. This study has been approved by the Institutional Review Board of Aizu Medical Center, Fukushima Medical University.

2.2.2 Experimental procedure

Division of the data set

The training set consists of 447 images, the validation set consists of 6 images, and the test set consists of 126 images. The training and validation sets are used for the parameters of the deep neural network, while the test set is used to test the training results. Both training and test sets are from different patients.

Models

In this study, we compare four models, PP-LiteSeg, U-Net, SegNet, and PspNet. U-Net, SegNet, and PspNet are the most common and effective models used for medical image segmentation. U-Net is mainly used in medical image segmentation and is fast becoming the baseline for most medical image semantic segmentation tasks [1].

U-Net consists of a total of 23 convolutional layers, the first half is feature extraction and the second half is upsampling. All features of medical images are important, so both low-level features and high-level semantic features are important, so the U-shaped structure of skip connection structure is better to come in handy.

SegNet is a symmetric network consisting of encoder (left) and decoder (right). The encoder part of SegNet uses the first 13 layers of VGG16 convolutional network, each encoder layer corresponds to a decoder layer, and the final decoder output is fed to a soft-max classifier to generate class probabilities for each pixel independently. [34].

The core module of PspNet is the pyramid pooling module, which aggregates contextual information from different regions, This improves the ability to obtain global information [35].

PP-LiteSeg [33] is a new lightweight real-time model to be segmented for the task and proposed a flexible and lightweight decoder (FLD) to reduce the computational overhead of previous decoders. To enhance the feature representation, the authors also propose a unified attention fusion module (UAFM), which uses spatial attention and channel attention to generate an attention weight, and then fuses the input features with the weight. In addition, a simple pyramid pooling module (SPPM) is proposed to aggregate global contextual information at a low computational cost.

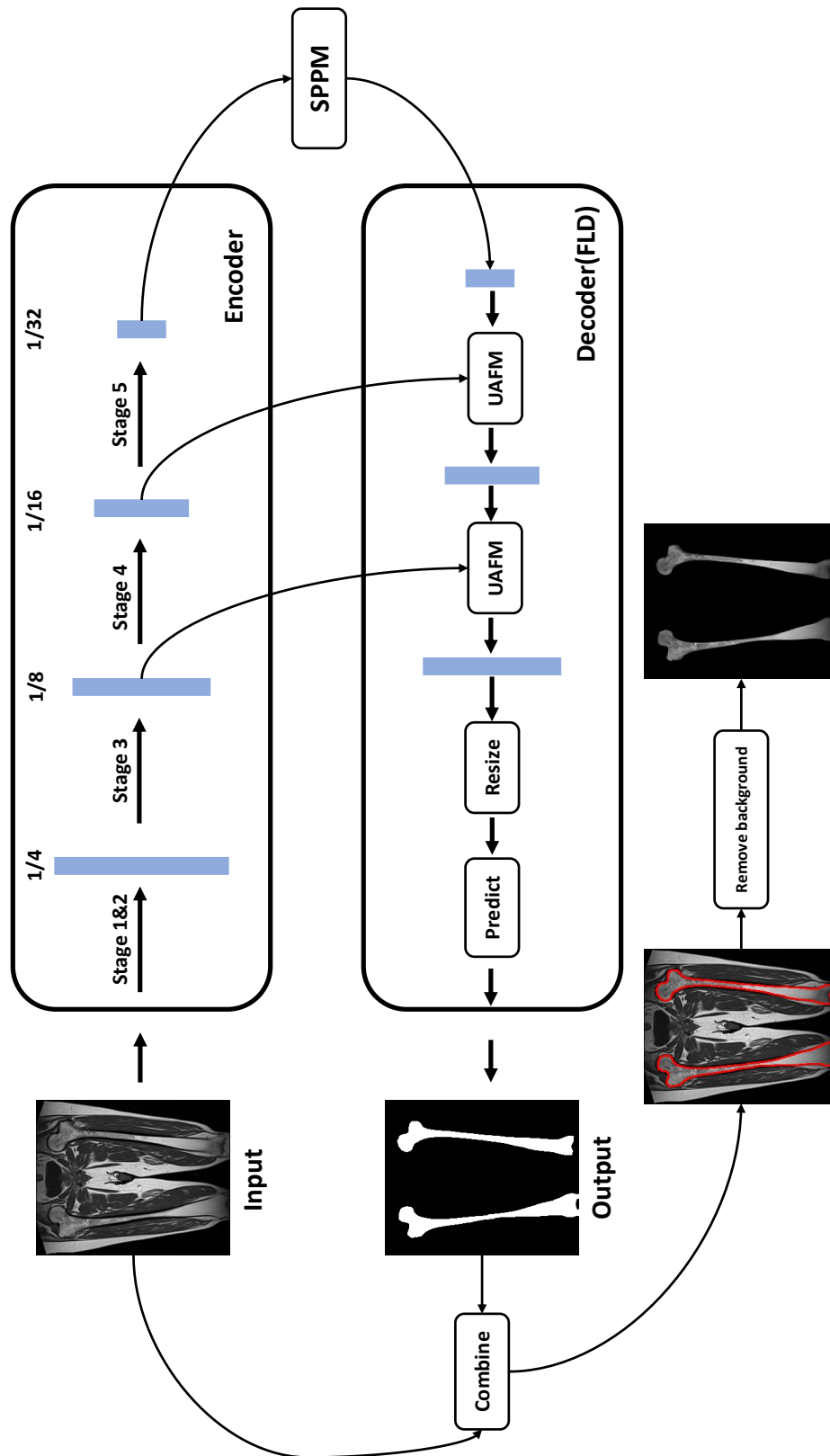


Figure 2.1: Network structure diagram of PP-LiteSeg.

In Fig. 2.1, the encoder module is the normal feature extraction part, fusing the branches of the different layers connected to the decoder. At the end of the encoder, an SPPM module is connected, this is a lightweight feature pyramid, which then enters the decoder structure (FLD), which contains the UAFM structure and acts as an attention mechanism.

FLD (Flexible and lightweight decoder) module is a lightweight decoder with gradually decreasing channel and increasing space size features. FLD can improve the efficiency of the overall model.

SPPM (The Simple Pyramid Pooling Module) has three global averaging pooling operations of size 1×1 , 2×2 , and 4×4 . The features are obtained, convolved, and upsampled, and finally, these upsampled features are added up and the convolution operation is applied to produce the subsequent features.

The UAFM (Unified Attention Fusion Module) uses the attention module to generate weights to obtain attention-weighted features. Finally, the UAFM sums the attention-weighted features element by element and outputs the fused features.

2.2.3 Experimental evaluation

Our proposed method was evaluated on MRI images of the femur. The segmentation results of the four models were compared using precision, specificity, recall, and dice coefficient(DSC). The difference between the results and the manual measurements was evaluated using DSC. These evaluation parameters are defined as.:

$$DSC = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (2.1)$$

$$Precision = \frac{TP}{FP + TP} \quad (2.2)$$

$$Recall = \frac{TP}{FN + TP} \quad (2.3)$$

$$Specificity = \frac{TN}{FP + TN} \quad (2.4)$$

(TP: True positive, FP: false positive, TN: true negative, FN: false negative)

DSC is usually used to calculate the similarity of two samples with a value threshold of [0, 1]. It is often used in medical images for image segmentation, where the best result of segmentation is 1 and the worst time result is 0 [36].

2.3 Results

In this study, we compared four neural networks for femoral segmentation. The evaluation indexes were listed in table 6.2. PP LiteSeg had the best performance, with an average DSC reaching 0.9195. The average DSC of U-Net, Segnet and Pspnet were 0.8783, 0.8638, and 0.8031, respectively. As shown in the same table, the other evaluation indexes (average Recall, Specificity, and Precision) are positively proportional to the average DSC.

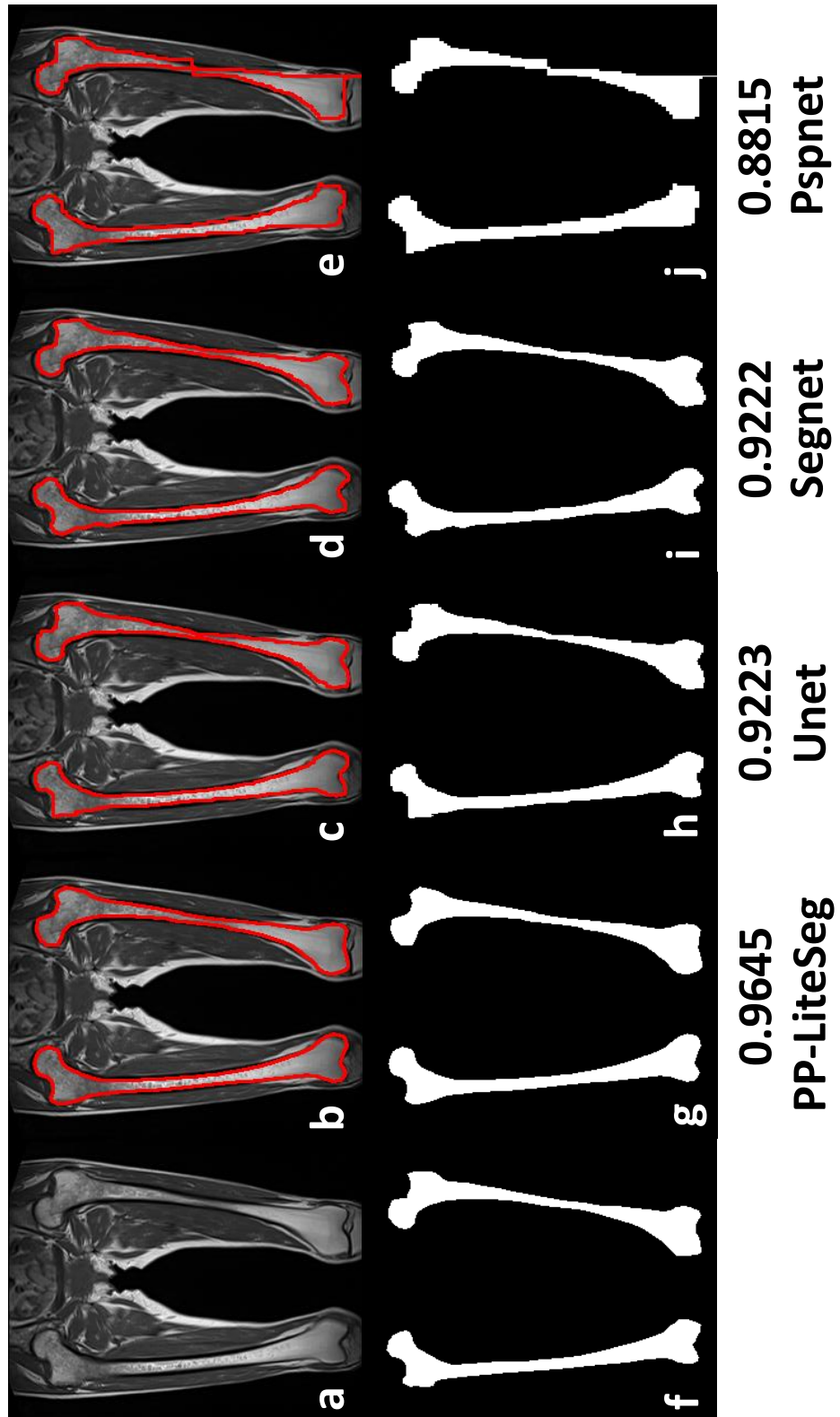


Figure 2.2: Example of results for the same set of femoral data in different models.

Table 2.1: Average evaluation parameters of different models.

Model Names	Avg_Recall	Avg_Specificity	Avg_Precision	Avg_DSC
PP-LiteSeg	0.9354	0.9909	0.9100	0.9195
U-Net	0.8925	0.9877	0.8738	0.8783
SegNet	0.8696	0.9872	0.8677	0.8638
PspNet	0.8050	0.9824	0.8249	0.8031

Fig. 5.3 shows a segmentation example using different models. Figs. 2a and 2f are the original image and the groundtruth of femur. Figs. 2g-2j are the segmentation results of PP-LiteSeg, U-Net, SegNet, and PspNet, respectively. The corresponding DSCs are 0.9645, 0.9223, 0.9222, and 0.8815, respectively. It is observed that the segmentation using PP-LiteSeg is the closest to groundtruth.

2.4 Discussion

In this paper, we compared four neural network models for femoral segmentation, and the results show that PP-LiteSeg with the addition of the attention mechanism module has the best performance on femur segmentation. The DSC reaches 0.92. In addition, the models were not chosen randomly. We tried about eight models in combination with others' studies and selected the four with better results for analysis. Segmentation plays a key role in medical projects. The accuracy of segmentation is a guarantee for further clinical trials.

PP-LiteSeg is a lightweight real-time model with the advantages of small size and fast computing speed, it is easier to deploy in mobile or embedded engineering machines. PP-LiteSeg proposes three innovative modules: Flexible Decoding Module (FLD), Attention Fusion Module (UAFM), and Simple Pyramid Pooling Module (SPPM). The FLD module flexibly adjusts the number of channels in the decoding module to balance the computation of the encoding and decoding modules, making the whole model more efficient; the UAFM module effectively enhances the feature representation and better improves the accuracy of the model; the SPPM module reduces the number of channels in the intermediate feature map and removes the jump connection, which further improves the model performance. It is based on the design and improvement of these modules that PP-LiteSeg achieves a balance between accuracy and speed, achieving the best results in the models.

In the introduction section we mentioned that our previous work also used neural networks to segment femurs from MRI images of patients with hematological malignancies. The reason of we researched the segmentation of the femur again is that, medical diagnosis requires a high degree of accuracy, and PP LiteSeg outperforms all previously used models in terms of performance. This facilitates further clinical work.

The limitations of this study are the small amount of data and the uneven number of different hematological malignancies. We will increase the number of training and testing data.

In the future, we also expect to propose a method on the automatic classification of different diseases based on the segmentation results of femur.

2.5 Conclusion

In this paper, PP-LiteSeg incorporating an attention mechanism module is used to segment femurs from MRI images. The results show that the PP-LiteSeg method is more accurate compared to conventional neural networks. The accurate segmentation is beneficial for further classification studies. PP-LiteSeg significantly saves time and computational cost compared with manual segmentation. Therefore, the proposed method may be useful in the clinical practice.

Chapter 3

Deep Learning Application for Femoral MRI Images Processing(Classification)

3.1 Introduction

Non-Hodgkin's Lymphoma (NHL) is a malignant neoplasm that affects the lymph nodes, lymphatic systems, and extra-nodal tissue, manifesting throughout the body [28]. It represents a significant public health concern, as reflected by the United States Cancer Statistics from 2015-2019, which reported 359,204 new cases and 101,438 deaths attributed to NHL. These numbers rank NHL as the eighth most common cancer in incidence and the ninth leading cause of cancer-related mortality [37]. The disease primarily affects individuals under 30 years old and aged 55 years or older. Progress has been made in NHL management over recent decades, and successful treatments have enabled at least 80% of patients to achieve remission [38] [39]. Diagnosing and treating NHL requires a thorough pathological analysis of bone marrow (BM) involvement. Conventional methods of obtaining BM samples through blind biopsy and aspiration offer limited information on the complete extent of BM infiltration and the specific type of infiltration. By contrast, magnetic resonance imaging (MRI) is an important imaging modality for the analysis of NHL in patients due to its ability to provide detailed and high-spatial-resolution information about the disease. MRI offers excellent soft tissue contrast, enabling the visualization of lymph nodes and extra-nodal structures with high precision [40]. It is particularly useful in assessing bone marrow involvement for lymphoma diagnosis and staging. In addition, MRI has emerged as a powerful tool for comprehensive NHL evaluation, allowing the detection of both localized and disseminated disease manifestations in various body regions. Furthermore, MRI is non-invasive and does not involve ionizing radiation, making it a safe option for repeated evaluations and follow-up assessments. These advantages make MRI an indispensable component of the diagnosis and treatment planning for NHL patients, facilitating disease characterization and guiding personalized therapeutic strategies [41] [42]. Early and accurate diagnosis of BM infiltration is crucial for guiding treatment decisions and assessing disease prognosis [43]. Diagnosing and treating NHL requires a thorough pathological analysis of bone marrow (BM) involvement. Conventional methods of obtaining BM samples through blind biopsy and aspiration offer limited information on the complete extent of BM infiltration and the specific type of infiltration. By contrast,

magnetic resonance imaging (MRI) is an important imaging modality for the analysis of NHL in patients due to its ability to provide detailed and high-spatial-resolution information about the disease. MRI offers excellent soft tissue contrast, enabling the visualization of lymph nodes and extra-nodal structures with high precision [40]. It is particularly useful in assessing bone marrow involvement for lymphoma diagnosis and staging. In addition, MRI has emerged as a powerful tool for comprehensive NHL evaluation, allowing the detection of both localized and disseminated disease manifestations in various body regions. Furthermore, MRI is noninvasive and does not involve ionizing radiation, making it a safe option for repeated evaluations and follow-up assessments. These advantages make MRI an indispensable component of the diagnosis and treatment planning for NHL patients, facilitating disease characterization and guiding personalized therapeutic strategies [41] [42]. Early and accurate diagnosis of BM infiltration is crucial for guiding treatment decisions and assessing disease prognosis [43]. However, visual classification of MRI patterns by physicians is subjective and time-consuming, leading to intra- and inter-differences in diagnosis and treatment planning. This may result in delays in appropriate therapies, thereby affecting patient outcomes. Moreover, the growing number of NHL cases demands more efficient and precise diagnostic tools to support clinicians in managing their patients effectively [44]. Therefore, to address the challenges posed by the requirement of more accurate analysis, recent studies have explored the integration of deep learning techniques in MR image processing and analysis. For instance, Hemanth et al. proposed an improved deep convolutional neural network (DCNN) to classify MR images of brain tumors accurately as benign or malignant. Their method achieved higher accuracy (96.4%) than traditional approaches (94.5%) [45]. Deniz et al. presented a 2-D and 3-D CNN-based method for segmenting the proximal femur, demonstrating improved performance over conventional methods [46]. Moreover, Irmakci et al. introduced a novel Deep-HLNet framework utilizing few-sample learning for the automatic classification of sensorineural hearing loss (SNHL) from MR images. Their method exhibited remarkable classification performance, with an overall accuracy of 96.62% and an average sensitivity of over 92% for all three balance categories [47]. Krishnapriya et al. investigated migration learning in the classification of brain MR images using a series of pre-trained models. The experimental results revealed that the pre-trained VGG-19 model with migration learning exhibited the best classification performance [48]. Funayama et al. proposed an improved model-based deep learning (iMoDL) network to confirm the quality of reconstructed images through training with abdominal MR images from 122 patients [49].

In this study, we propose an advanced method based on PP-LiteSeg for the segmentation of femoral MRI images in patients with hematological malignancies. We aim to demonstrate that our model effectively improves segmentation accuracy while addressing the limitations of traditional convolutional neural networks (CNNs). We also introduce SAMSCNet, a novel deep learning model designed to improve the classification of femoral MRI images of NHL patients. Our model utilizes DenseNet169 as an encoder with a dual-branching strategy that integrates segmentation and classification to leverage both global and local features. Due to the need for more accurate and efficient diagnostic tools, we aim to address the limitations of existing methods, such as background interference and computational complexity. We aim to provide reliable tools for clinical applications in the diagnosis and treatment of NHL.

3.2 Materials and Methods

3.2.1 Data

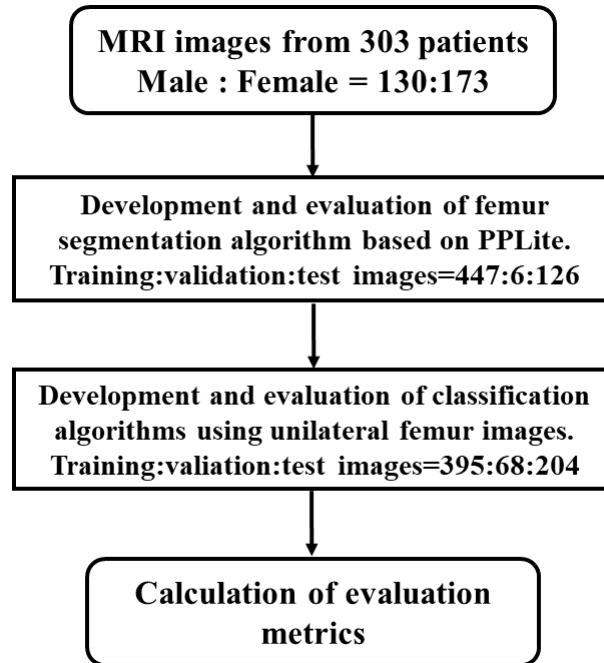


Figure 3.1: Study flowchart

As shown in Figure 3.1, this retrospective cohort study recruited 303 consecutive patients, including 130 males and 173 females, aged between 55 and 85 years old, diagnosed with NHL in 2012-2020 at Aizu Medical Center, Fukushima Medical University. T1-weighted images (T1WI) were obtained using a 1.5T MRI system (MAGNETOM Avanto 1.5T, Siemens, Germany) with a T1WI Turbo spin-echo coronal sequence. Ethical approval for the study was obtained from the Institution Review Board of Fukushima Medical University, and the study adhered to the relevant guidelines and regulations outlined in the Declaration of Helsinki. Before enrollment, all participants provided written informed consent after an explanation of the study by doctors.

3.2.2 Experimental Design

We systematically designed and conducted four distinct experiments to rigorously evaluate the performance of various classification methodologies for femoral MR images. Each experiment encompassed a comparative analysis of multiple deep learning models to ascertain the most productive approach. The principal innovation of this study is encapsulated in Experiment Four, wherein we introduce and apply a novel classification model directly to the original MR images, leveraging the strengths identified in previous experiments. The femur classification was based on the categorization from

a previous clinical study [44] and included three types. The first type was a Uniform femur with a homogeneous low signal region in the bone marrow extending from the proximal to the distal femur; the second was a Non-Uniform femur with a heterogeneous, nodular, and scattered low signal region in the bone marrow; and the third was a Normal femur with an absence of a distinct low signal region in the bone marrow. Two experienced physicians (SI and ST) reviewed and classified the training, validation, and test data through discussion. Details of training, validation, and test datasets for classification are shown in Table 3.1.

Figure 3.2(a)–(c) illustrate samples of the Uniform femur (Uniform), Non-Uniform femur (Non-Uniform), and Normal femur (Normal), respectively. Figure 3.2(d)–(f) are the corresponding segmentation results of Figure 3.2(a)–(c), respectively. Figures 3.2(g)–(i) are the samples of the unilateral femur patches.

3.2.3 Experiment 1

In this experiment, the MR images were directly classified without any prior segmentation. This approach allows for the observation of the entire image, but it is susceptible to background interference. The training, validation, and test datasets comprised 234, 38, and 118 images from 121, 21, and 63 patients, respectively (the third step in Figure 3.1). Images containing implanted bone fixation devices or incomplete femurs were excluded from the study. We compared several classification models, including ResNet50, ResNet 101, DenseNet121, and DenseNet 169. The model parameters were as follows: batch size of 16, input size of 320×320, 100 epochs, initial warm-up of 5 epochs, and an initial learning rate of 0.001. Stochastic gradient descent was used to update the models with a cross-entropy loss function.

In this study, DenseNet169 was modified to perform a three-category classification. The batch size, input size, epoch number, initial warm-up epoch number, and initial learning rate were set to 16, 320, 100, 5, and 0.001, respectively. A stochastic gradient descent method was used to update the models with a cross-entropy loss function (Equation 3.1). In equation 3.1, M is the number of categories; y_{ic} a sign equation (0 or 1), equaling 1 if the true category of sample i is equal to c , 0 otherwise; p_{ic} the predicted probability that the observed sample i belongs to category c .

$$L = \frac{1}{N} \sum_i L_i = -\frac{1}{N} \sum_i \sum_{c=1}^M y_{ic} \log(p_{ic}) \quad (3.1)$$

3.2.4 Experiment 2 Segmentation Followed by Classification

The second experiment involved segmenting the MR images to isolate the femoral region before classification. This method reduces background noise, focusing on the target area. The first stage involves femoral segmentation from MR images using PP-LiteSeg. PP-LiteSeg was selected based on our previous research, wherein it demonstrated superior performance compared with other methods such as U-Net, SegNet, and PspNet, achieving an average Dice coefficient of 0.92 [50]. PP-LiteSeg utilizes an encoder architecture comprising three novel modules, a Flexible and Lightweight Decoder (FLD), a Unified Attention Fusion Module (UAFM), and a Simple Pyramid Pooling Module (SPPM). The FLD module progressively reduces channel numbers while increasing feature space sizes, balancing computational complexities and improving

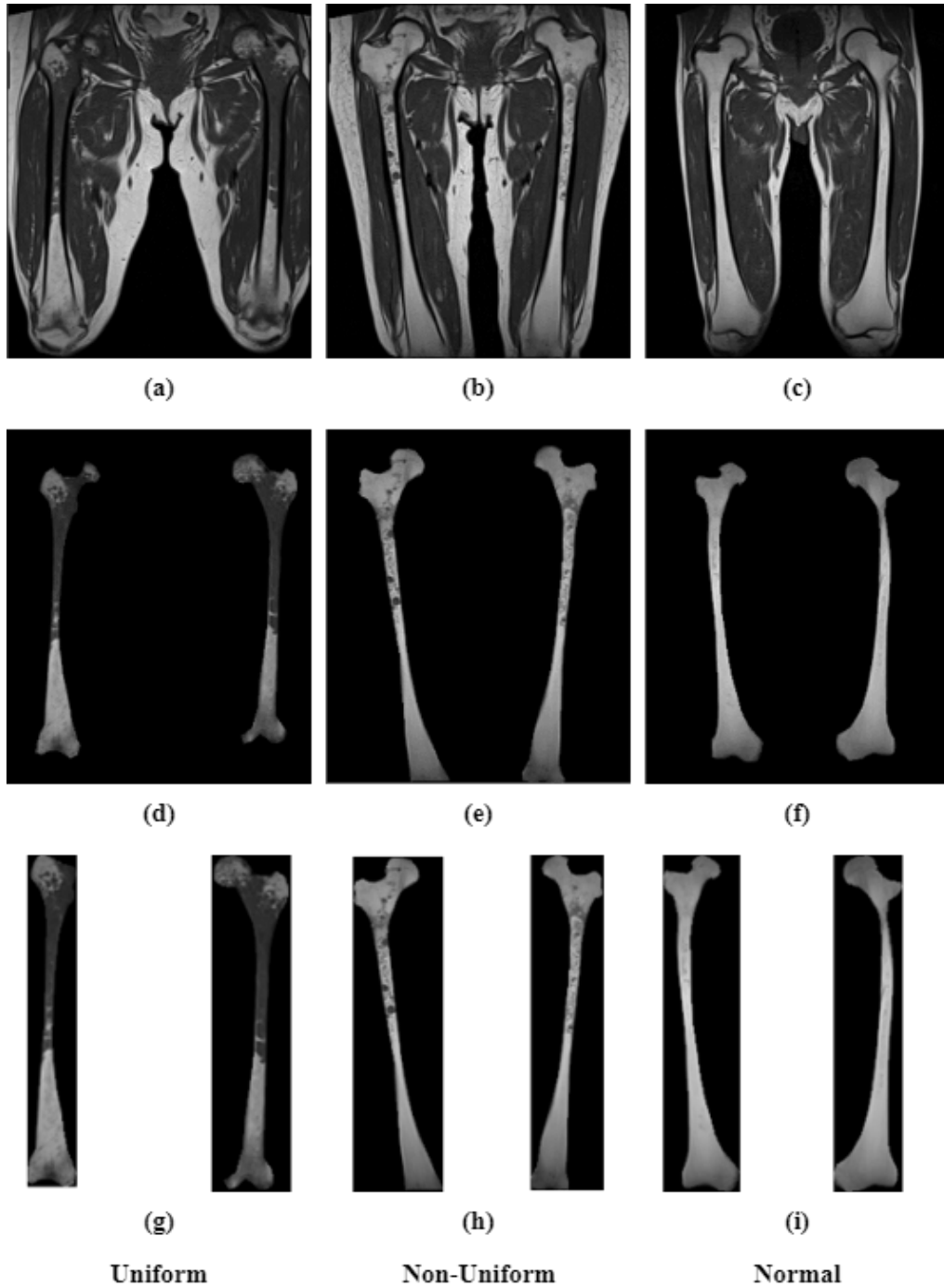


Figure 3.2: Three Different Types of Femur

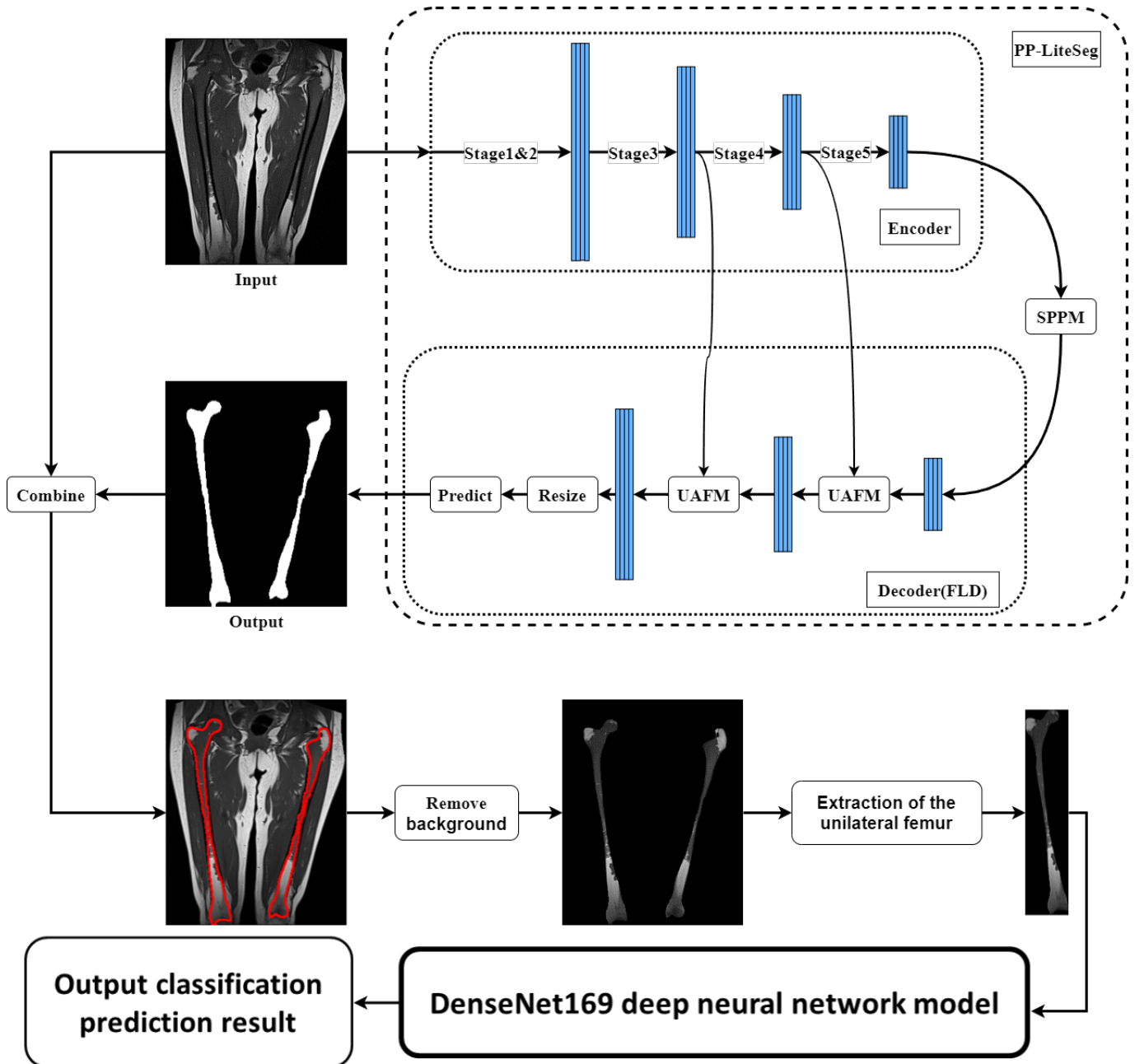


Figure 3.3: Flowchart of ASEC Method

Table 3.1: Details of Training, Validation, and Test Datasets for Classification

Dataset	Non-Uniform	Normal	Uniform	Total
Original and Segmentation images				
Train	82	82	70	234
Validation	15	12	11	38
Test	45	40	33	118
Unilateral images				
Train	141	145	110	396
Validation	28	21	19	68
Test	77	69	58	204

model efficiencies. The UAFM module employs both channel and spatial attention mechanisms to enhance feature representation, and accurate segmentation is achieved through multilevel feature fusion. SPPM, compared with the traditional PPM model, reduces intermediate and output channels, replaces the cat operation with the add operation, and eliminates the shortcut operation, thus increasing its effectiveness [33]. Developing the segmentation method based on PP-LiteSeg involved 579 MR images (the second step in Figure 3.1). The training, validation, and test datasets consisted of 447, 6, and 126 images from 149, 1, and 37 patients, respectively. There was no patient overlap among the three datasets. In this study, PP-LiteSeg was modified to perform a two-class segmentation (background and femur). Batch size, input size, epoch number, initial warm-up epoch number, and initial learning rate were set to 6, 960×720 , 100, 5, and 0.001, respectively. Stochastic gradient descent was used to update the weights of the model with an OHEM-CrossEntropyLoss function [51] (Equation 3.2). Because the number of background samples is usually much more than the target class samples, it leads to the problem of imbalance in data distribution, and some difficult samples are challenging for the training of the network. Equation 3.2 is designed to solve these problems. The core idea of this loss function is to select only those difficult samples with high loss values for gradient updating during the training process, thus focusing more on the samples that are difficult to classify, which helps the network to adapt better to these samples and improve the performance of the model.

$$\text{OHEM-CrossEntropyLoss} = -\frac{1}{N} \sum_{i=1}^N \begin{cases} \log(p_{\text{target}}) & \text{if } y_{\text{target}} = 1 \text{ (Target class sample)} \\ \log(1 - p_{\text{target}}) & \text{if } y_{\text{target}} = 0 \text{ (Background class samples} \\ & \text{and loss values above thresholds)} \\ 0 & \text{otherwise} \end{cases} \quad (3.2)$$

For the classification step, we utilized DenseNet169, the model that achieved the best results in Experiment One.

3.2.5 Experiment 3 Segmentation, Unilateral Femur Extraction, and Classification

In Experiment III, the magnetic resonance image is first segmented and then the unilateral femur is extracted before classification. Figures 3.2(g)–(i) are the samples of the unilateral femur patches. Extracting the unilateral femur combines the advantages of Experiment II, both in terms of reducing background noise and further increas-

ing the amount of data. The same segmentation and classification models are used as in Experiments I and II. Although this approach enhances the focus on the target region, it is computationally expensive. PP-LiteSeg was again used for segmentation and DenseNet169 for classification.

3.2.6 Experiment 4 Direct Classification Using SAMSCNet Model

In Experiment Four, we proposed a classification model, SAMSCNet, applied directly to the original MR images. This model aims to optimize both accuracy and computational efficiency. SAMSCNet leverages DenseNet169 as the encoder, processing input images through the initial convolutional layer and four dense blocks to generate a series of feature maps at different resolutions, as illustrated in Figure 3.4.

Following the encoder, the model employs a dual-branch strategy: **Branch One: Segmentation Pathway** In the first branch, we implemented a U-Net-like upsampling structure with skip connections. These skip connections combine features from the initial convolutional layer and each of the four dense blocks. Through a series of upsampling and convolution operations, this branch generates the segmentation result, `logits_seg`. This result represents the probability of each pixel belonging to the target class. To map these values to the $[0, 1]$ range, `logits_seg` is passed through the `torch.sigmoid` function, producing an attention map. To optimize the segmentation task, we combine Tversky Loss and Focal Loss. Tversky Loss effectively handles class imbalance by adjusting the penalty for false positives and false negatives, making it particularly suitable for our medical imaging application where the target regions are often small. Focal Loss, on the other hand, focuses on hard-to-segment regions by down-weighting the loss contribution of well-classified examples. This combination ensures that the model pays more attention to difficult cases and improves the overall segmentation performance.

Branch Two: Classification Pathway

In the second branch, the attention map generated in the first branch is interpolated to match the size of the final feature map from the last dense block. This interpolation ensures alignment between the attention map and the feature map. The final feature map is then weighted by the attention map, enhancing regions with high probability and suppressing those with low probability. This weighted feature map is passed through a fully connected layer to generate the final classification result.

To optimize the classification task, we use Label Smoothing Cross Entropy Loss. This loss function prevents the model from becoming overconfident in its predictions by assigning a small probability to incorrect classes, thus improving generalization and robustness.

This dual-branch strategy allows the model to leverage both pixel-level segmentation and global image features, enhancing classification accuracy and robustness. By capturing global context while focusing on relevant local features, DenseU-Net+ effectively balances computational efficiency and performance. This design enables the model to make more informed decisions by emphasizing important regions identified during the segmentation process, thereby improving the overall classification results.

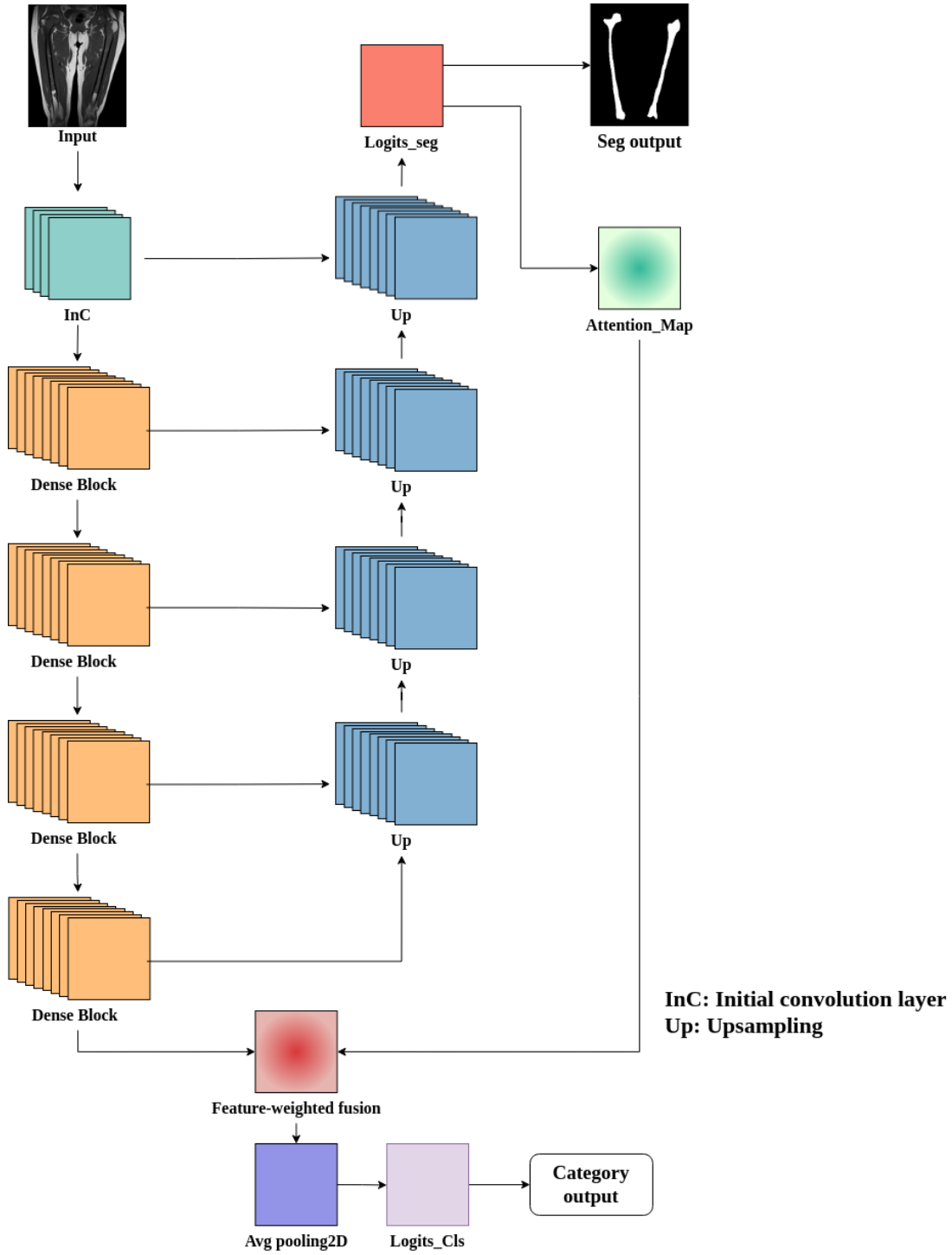


Figure 3.4: Flowchart of SAMSCNet Method

3.2.7 Evaluation Measures

Accuracy, Precision, Sensitivity, Specificity, Macro-Precision, Macro-Sensitivity, Macro-Specificity, and Macro-F1 defined below are used to evaluate the accuracy of the four classification models.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (3.3)$$

$$Precision = \frac{TP}{TP + FP} \quad (3.4)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (3.5)$$

$$Specificity = \frac{TN}{TN + FP} \quad (3.6)$$

$$Macro - Precision = \frac{Precision_{Original} + Precision_{Segmentation} + Precision_{Unilateral}}{3} \quad (3.7)$$

$$Macro - Sensitivity = \frac{Sensitivity_{Original} + Sensitivity_{Segmentation} + Sensitivity_{Unilateral}}{3} \quad (3.8)$$

$$Macro - Specificity = \frac{Specificity_{Original} + Specificity_{Segmentation} + Specificity_{Unilateral}}{3} \quad (3.9)$$

$$Macro - F_1 = 2 \cdot \frac{Macro - Precision \cdot Macro - Sensitivity}{Macro - Precision + Macro - Sensitivity} \quad (3.10)$$

where TP is true positive, FP false positive, TN true negative, and FN false negative. In addition, we used a confusion matrix to compute the accuracy of the model. Confusion matrix, also known as error matrix, is a standard format for representing accuracy evaluation in the form of a matrix with n rows and n columns. In Artificial Intelligence, confusion matrices (confusions) are visualization tools, especially used in supervised learning. Each row of the matrix represents instances in the predicted class, whereas each column represents instances in the actual class (and vice versa). The name derives from the fact that it is easy to see if the system confuses the two classes [52].

3.3 Results

3.3.1 Overview of experimental results

In this section, we present the results of four experiments, focusing on the performance of the best model in each experiment. Details of the comparison of the results

of the segmentation classification models in Experiments I, II, and III we placed in the Supplementary Material, and the model we present in Experiment IV has a higher accuracy.

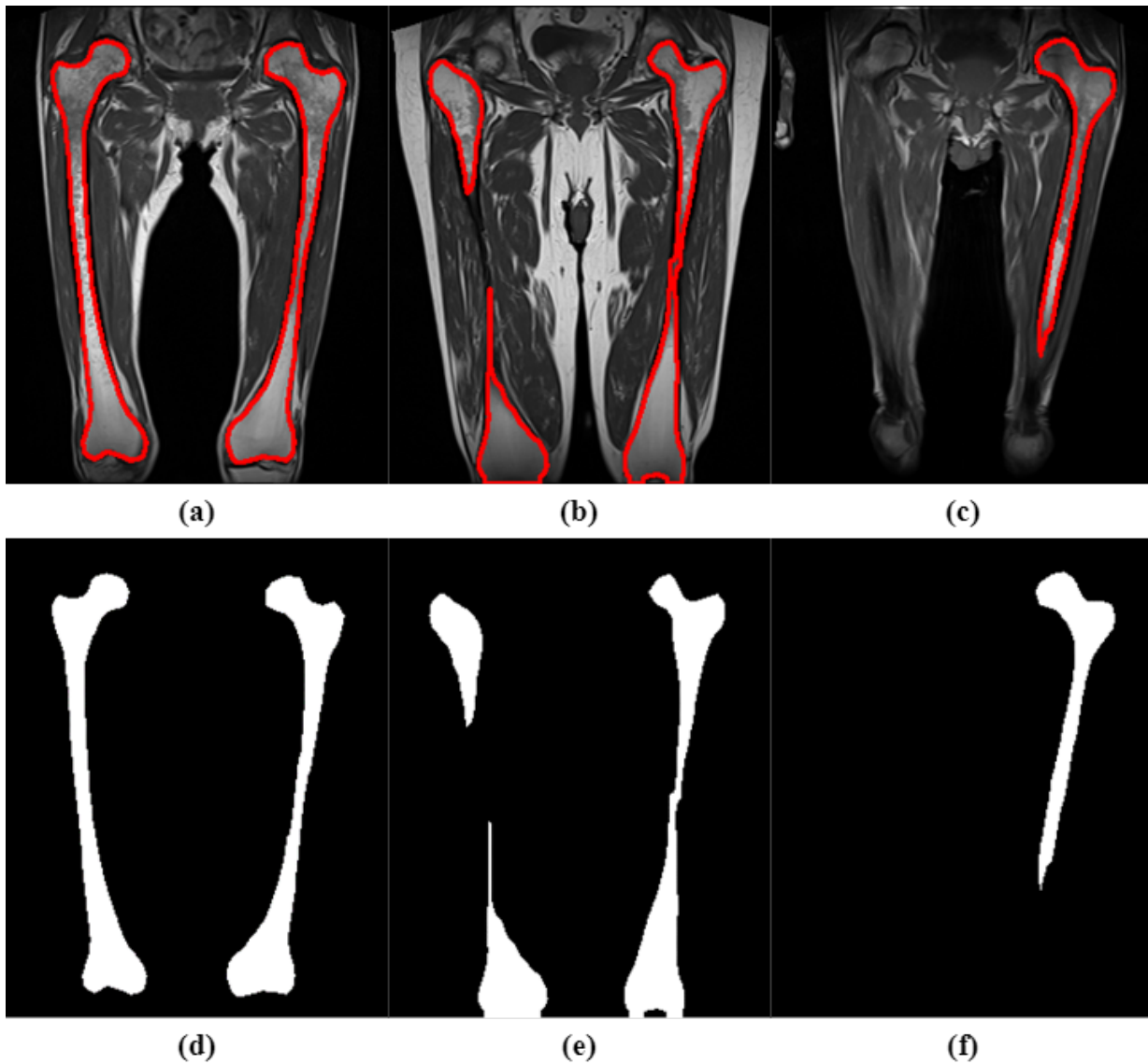


Figure 3.5: Examples of Femoral Segmentation Results

3.3.2 Summary of the results of the four experiments

The comparative results of the four experiments are summarized in Table 2 and visualized in Figure 4???. These comparisons demonstrate the better performance of our proposed model in Experiment Four.

Experiment One: Direct Classification of Original Images

In Experiment One, the best-performing model was DenseNet169. This model achieved an accuracy of 77.8%, a precision of 75.6%, a recall of 80.9%, and an F1-score of 78.2%. Despite the simplicity of directly classifying original MR images, the presence of background noise affected the overall performance.

Experiment Two: Segmentation Followed by Classification

In Experiment Two, we used the PP-LiteSeg model for segmentation followed by classification with DenseNet169. This approach resulted in improved performance with an accuracy of 80.5%, a precision of 82.4%, a recall of 76.5%, and an F1-score of 79.4%. The segmentation step effectively reduced background noise, allowing the classifier to focus on the femoral region.

Experiment Three: Segmentation, Unilateral Femur Extraction, and Classification

Experiment Three further enhanced classification performance by segmenting the MR images and extracting the unilateral femur before classification. Using the same models (PP-LiteSeg for segmentation and DenseNet169 for classification), this method achieved an accuracy of 81.0%, a precision of 84.7%, a recall of 82.7%, and an F1-score of 83.7%. This approach increased the data quantity and further reduced background noise, albeit with a higher computational cost.

Experiment Four: Direct Classification Using Novel Model

Experiment Four introduced our novel classification model, applied directly to the original MR images. This model integrated features from the first three experiments through a weighted combination approach, optimizing both accuracy and computational efficiency. The proposed model achieved the highest performance with an accuracy of 84.2%, a precision of 86.0%, a recall of 85.3%, and an F1-score of 85.6%. This result highlights the model's ability to capture global features while minimizing background interference, proving its robustness and reliability.

3.3.3 Models Results

The segmentation performance of four different models, U-Net, PspNet, SegNet, and PP-LiteSeg, was compared. PP-LiteSeg demonstrated the best performance, achieving significantly better results in all the evaluation criteria. The average sensitivity, specificity, and precision were 92.0%, 93.5%, 99.1%, and 91.0%, respectively [50]. Samples of segmentation results are shown in Figure 3.5.

Figures 3.6, 3.8 illustrate the Normalized confusion matrix using the classification results of original, segmentation, and unilateral images, respectively. According to the statistics of the confusion matrix, the error types of Uniform and Normal classes are classified as Non-Uniform, whereas the error types of Non-Uniform classes are mainly classified as Normal. Examples of classification errors are shown in Figure 3.9. Uniform femoral bone marrow in Figure 3.9(a) was wrongly classified as a Non-Uniform femoral bone marrow. The corresponding prediction values are Uniform: 1.9980, Non-Uniform: 2.6046, and Normal: 4.5905. The difference between Uniform and Non-Uniform prediction values is small. This may be caused by the discontinuity of the Uniform infiltration in Figure 8(a). In Figure 3.9(b), the predicted result is Non-Uniform and the actual result is Normal. After analysis, the vascular grooves (the dark thin lines) in the image that affect the classification result. Figure 3.9(c), the predicted result is Normal and the actual result is Non-Uniform. The prediction was wrong because the signal of the proximal and distal femur was better and no infiltration was observed.

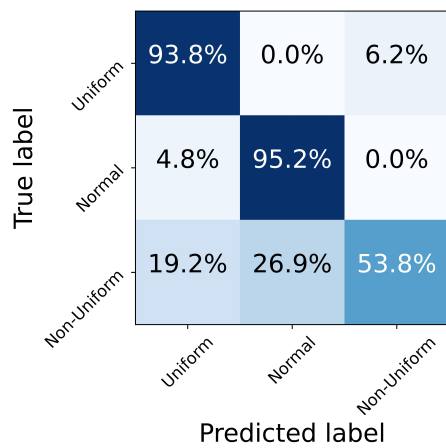


Figure 3.6: Original Confusion Matrix

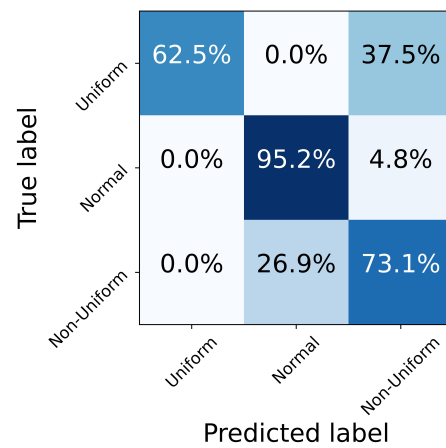


Figure 3.7: Segmentation Confusion Matrix

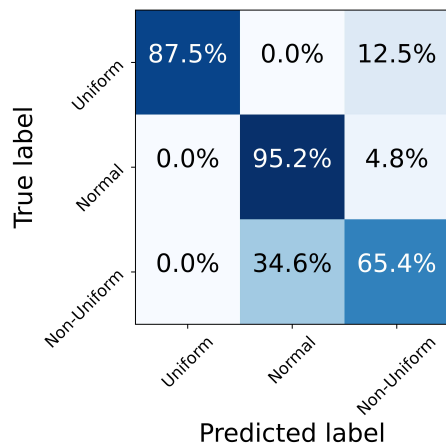


Figure 3.8: Unilateral Confusion Matrix

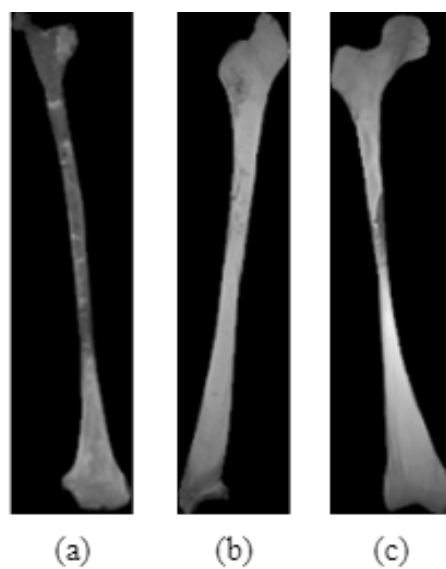


Figure 3.9: Example of Prediction Error

Gradient-weighted Class Activation Mapping (Grad-CAM) was also performed to confirm the regions important for image segmentation and classification [53]. As shown in Figure 3.10, the proximal end and body of femurs were mainly used for image segmentation and classification by the models employed. Examples of classification errors are shown in Figure 3.9. Uniform femoral bone marrow in Figure 3.9(a) was wrongly classified as a Non-Uniform femoral bone marrow. The corresponding prediction values are Uniform: 1.9980, Non-Uniform: 2.6046, and Normal: 4.5905. The difference between Uniform and Non-Uniform prediction values is small. This may be caused by the discontinuity of the Uniform infiltration in Figure 8(a). In Figure 3.9(b), the predicted result is Non-Uniform and the actual result is Normal. After analysis, the vascular grooves (the dark thin lines) in the image that affect the classification result. Figure 3.9(c), the predicted result is Normal and the actual result is Non-Uniform. The prediction was wrong because the signal of the proximal and distal femur was better and no infiltration was observed. Gradient-weighted Class Activation Mapping (Grad-CAM) was also performed to confirm the regions important for image segmentation and classification

[53]. As shown in Figure 3.10, the proximal end and body of femurs were mainly used for image segmentation and classification by the models employed.

3.4 Discussion and Future work

In this study, we aimed to enhance the accuracy and efficiency of classifying femoral MR images in NHL patients by developing and comparing several deep learning models. The primary objective was to address the limitations of existing methods by introducing a novel model, SAMSCNet, which integrates the strengths of previous approaches while minimizing their weaknesses. This research holds significant clinical relevance, as accurate classification of bone marrow involvement is crucial for guiding treatment decisions and assessing prognosis in NHL patients.

3.4.1 Contribution

We conducted four experiments to evaluate different classification methodologies. Experiment One involved direct classification on original MR images using DenseNet169, which provided a baseline performance but was susceptible to background noise. Experiment Two improved accuracy by segmenting the femoral region before classification, using PP-LiteSeg followed by DenseNet169, effectively reducing background interference but at the cost of increased computational complexity. Experiment Three further refined this approach by segmenting and extracting the unilateral femur before classification, enhancing focus on the target area and increasing data quantity, though it remained computationally intensive. Experiment Four, which introduced our novel model SAMSCNet, achieved the best overall performance with high accuracy, precision, recall, and F1-score. While the improvement in accuracy over Experiment Three was modest, SAMSCNet significantly reduced computational requirements and preserved global image context. This balance of efficiency and performance highlights the model's practical advantages for clinical application. SAMSCNet's superior performance can be attributed to its dual-branch architecture. By leveraging DenseNet169 as an encoder and implementing a U-Net-like upsampling structure with skip connections,

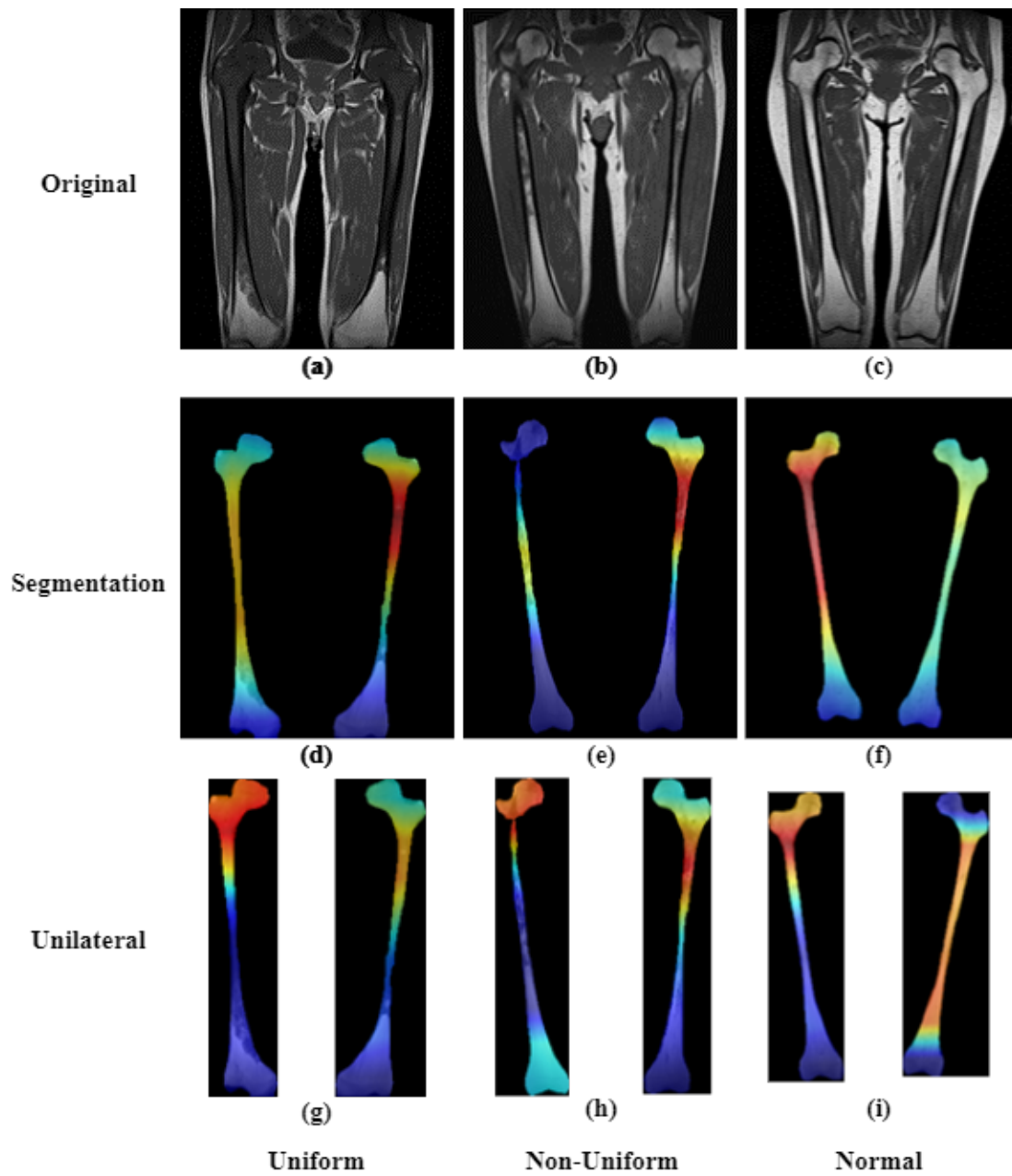


Figure 3.10: Results with Grad-CAM

SAMSCNet effectively combines pixel-level segmentation and global image features. The attention map generated in the first branch refines the feature map in the second branch, enhancing regions with high probability while suppressing low-probability areas. This integrated approach allows SAMSCNet to capture both local and global features, improving classification robustness and accuracy. Moreover, the use of Tversky Loss and Focal Loss for segmentation addresses class imbalance and focuses on challenging regions, while Label Smoothing Cross Entropy Loss for classification prevents overconfidence in predictions. These combined strategies ensure that SAMSCNet delivers consistent and reliable performance across different types of femoral MR images.

3.4.2 Limitations

This study has several key limitations. The first limitation is the insufficient and imbalanced data, which negatively impacts the training outcomes. Additionally, there is an issue of incomplete data information. As illustrated in Figure 3.5, Figure 3, panel (b) shows an incomplete representation of the left femur, while panel (c) depicts a scenario where the left femur is entirely obscured. This issue arises from the potential curvature of the patient's femur, which may prevent capturing both femurs entirely within a single plane during an MRI scan. Another significant limitation is that we only used data from a single medical center. Current classification models often achieve better results on single-center datasets, but their performance may not generalize well to datasets from other centers without retraining. Furthermore, although the computational efficiency of SAMSCNet has been improved, it still requires validation in real-time clinical settings to ensure its practical applicability. Additionally, the reliance on manual annotations for training data introduces a potential source of bias and variability.

3.4.3 Future Work

Because single-layer MR images can only provide limited information, future investigations will involve constructing 3-D images for classification using multilayer MR images, from which 3-D features of the infiltrated area can be obtained to improve the performance of classification. Moreover, the spatial localization and quantification of the infiltrated area in 3-D images could offer physicians valuable insights into precise treatment planning, aiding more effective therapeutic strategies. Future research should focus on validating SAMSCNet on larger and more diverse datasets from multiple medical centers to enhance its generalizability. Further optimization of the model architecture and training procedures could yield even better performance and efficiency. Exploring semi-supervised or unsupervised learning techniques may reduce the reliance on manual annotations, thereby mitigating potential biases. Additionally, integrating SAMSCNet with other imaging modalities and clinical data could provide a more comprehensive tool for diagnosing and managing NHL.

3.5 Conclusion

This study presents SAMSCNet, a novel deep learning model designed to improve the classification of femoral MR images in NHL patients. By integrating DenseNet169 with a dual-branch strategy that combines segmentation and classification, SAMSCNet

captures both global and local features, achieving superior accuracy and efficiency. Our experiments show that SAMSCNet significantly outperforms traditional methods, with an accuracy of 84.2%. This success is due to the model's ability to use segmentation-derived attention maps to enhance relevant features while reducing background noise. As a result, SAMSCNet provides a robust and reliable tool for the accurate classification of femoral MR images, essential for diagnostic decisions and prognostic predictions in NHL. Despite these advancements, the study has limitations, including a single-center dataset and reliance on manual annotations. Future work should validate SAMSCNet on larger, multi-center datasets and explore semi-supervised or unsupervised learning techniques to reduce potential biases. In summary, SAMSCNet represents a significant step forward in medical image classification, offering substantial potential for clinical applications in NHL diagnosis and management. Future research will focus on optimizing the model further and integrating it with other imaging modalities for comprehensive diagnostic capabilities.

Chapter 4

Deep Learning Application for Transvaginal Uterine Ultrasound Images Processing(Semi-Supervised Segmentation, BCP-Mamba)

4.1 Introduction

Medical image segmentation is crucial in various diagnostic and therapeutic procedures to help clinicians make accurate diagnoses and develop treatment plans [54]. The uterine peritoneum is the outer plasma membrane of the uterus, equivalent to the peritoneum of the abdomen [55]. Similar to the abdominal peritoneum, the uterine peritoneum provides structural support and protection for the uterus [56]. Specially, the accurate segmentation of the peritoneum is critical for the diagnosis of various uterine pathologies such as fibroids, adenomyosis, and endometrial abnormalities [57]. In addition, the disruption of the uterine peritoneum can affect fertility and pregnancy outcomes. This makes accurate segmentation of the uterine peritoneum clinically important for reproductive health management. Supervised learning methods have shown effectiveness in this task [58]. However, considering the tedious and costly task of manual contour drawing for medical image annotation, semi-supervised segmentation has received increasing attention in recent years and has been widely used in medical image analysis.

In general, labeled and unlabeled data share the same distribution in semi-supervised medical image segmentation. However, in the real world, it is difficult to estimate the exact distribution of limited labeled data. Therefore, there is always an empirical distribution mismatch between the large amount of unlabeled data and the tiny amount of labeled data. To overcome this problem, semi-supervised segmentation methods always try to train labeled and unlabeled data symmetrically consistently [59–62]. Vision Mamba (VM) [63] was recently proposed for image segmentation with location-aware visual recognition through location embedding, making the model more robust in dense prediction tasks. Mamba is a new architecture for LLMs that can handle long sequences more efficiently than traditional models such as Transformers. Mamba’s efficiency comes from its bi-directional state space model, which theoretically allows for faster processing of image data compared to that of the traditional Transformer model.

In this study, we propose a bidirectional copy-paste Mamba (BCP-Mamba) for the

enhanced semi-supervised segmentation of transvaginal ultrasound uterine images. The BCP-Mamba architecture can efficiently utilize the limited labeled data while leveraging the rich unlabeled data to improve segmentation accuracies. Comprehensive experiments were also performed to compare the proposed method with the well-established semi-supervised models U-Net [1] and BCP-Net [64] and demonstrate its efficacy in the segmentation of transvaginal uterine ultrasound images.

4.2 Related Work

Previous research in medical image segmentation has explored various supervised and semi-supervised learning approaches. Supervised methods typically rely on annotated data for training convolutional neural networks (CNNs) to segment the target structures in medical images accurately. However, the scarcity of labeled data poses a significant challenge in medical imaging tasks, motivating the exploration of semi-supervised approaches [65] [66]. Semi-supervised learning methods aim to leverage both labeled and unlabeled data to improve model performance, often through techniques such as consistency regularization, pseudo-labeling, and data augmentation [67].

Many efforts have been made in semi-supervised medical image segmentation. Entropy minimization (EM) and consistency regularization (CR) stand out as two commonly employed loss functions. Additionally, researchers have extended the mean teacher framework in various ways. For instance, SASSNet [68], leverages unlabeled data to impose geometric shape constraints on segmentation outputs, whereas DTC [69], introduces a dual-task consistency framework by explicitly incorporating task-level regularization. SimCVD [70], explicitly models geometric structures and semantic information, constraining them within the teacher and student networks. These methods employ geometric constraints to monitor the network outputs. UA-MT [71], utilizes uncertainty information to guide the student network towards meaningful and reliable goals established by the teacher network. [72] Combines image-intelligent and patch-intelligent representations are combined to explore more intricate similarity cues, ensuring output consistency across different input sizes. CoraNet [73], proposes a model that generates both deterministic and indeterministic regions, with the student network assigning varying weights to regions from the teacher’s network. UMCT [74], utilizes diverse viewpoints of the network to predict the same image from different angles, employing prediction and the corresponding uncertainty to generate pseudo-labels for supervised prediction of unlabeled images.

In recent years, significant advancements have been made in the field of supervised and semi-supervised ultrasound image segmentation. Notably, Oktay et al. introduced the Attention U-Net, which enhances segmentation performance by focusing on relevant regions of the image using attention mechanisms [75]. Similarly, Li et al. presented the H-DenseUNet, a hybrid model that combines dense connections with the U-Net architecture to improve segmentation accuracy for liver and tumor segmentation from CT volumes [76]. These studies underscore the ongoing efforts to improve segmentation accuracy and robustness, providing valuable insights and methodologies that inform our approach.

These approaches significantly enhance the efficacy of semi-supervised medical image segmentation. However, they often overlook the process of learning generic semantics from labeled to unlabeled data. Treating labeled and unlabeled data separately

frequently impedes knowledge transfer from labeled to unlabeled data.

Furthermore, very few studies have focused on semi-supervised segmentation of the perimetrium. This study used the BCP-Manba framework for semi-supervised medical image segmentation, which effectively utilizes unlabeled data by leveraging image translations and reconstructions. This framework is designed to correctly segment the perimetrium. The BCP-Net architecture on which it is based has on demonstrated promising results across various medical imaging modalities, laying the foundation for further advancements in semi-supervised segmentation.

4.3 Methodology

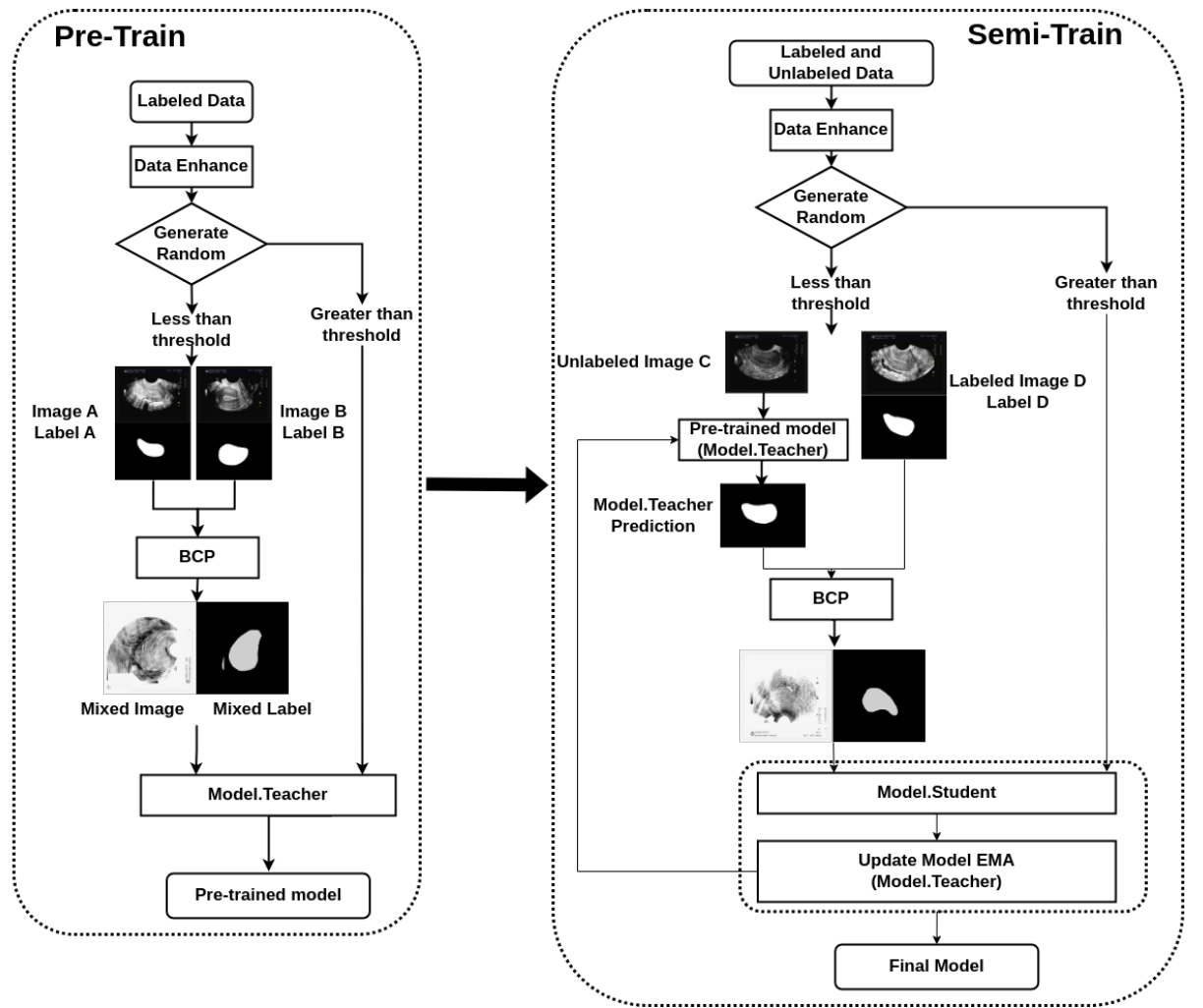


Figure 4.1: Flowchart of the bidirectional copy-paste Mamba (BCP-Mamba) model

Our semi-supervised segmentation method is based on the BCP-Net framework. It helps to integrate labeled and unlabeled data, thereby improving segmentation performance. We have made some modifications to the BCP-Net framework. As shown in Figure 4.1, we added a Generate Random judgment, which enables data from different modalities to participate in the training and avoids a single input in the BCP mode. A detailed description is in the following subsection. The focus is on improving the effi-

ciency and effectiveness of the segmentation process, making it particularly suitable for the task of segmenting transvaginal ultrasound uterine images.

4.3.1 Data Preprocessing

Before delving into the segmentation process, it is essential to preprocess the data appropriately. We start by collecting a dataset comprising both labeled and unlabeled transvaginal ultrasound uterine images. The labeled image contains manually labeled areas corresponding to the perimetrium, whereas unlabeled images lack such annotations. A total of 1940 transvaginal ultrasound images were acquired from Tongji Hospital, affiliated with Huazhong University of Science and Technology, for this study. The patient cohort encompassed individuals spanning an age range from 19 to 80 years.

4.3.2 Bidirectional Copy-Paste (BCP) Operation

The BCP operation is the core component of our methodology, and it facilitates the fusion of labeled and unlabeled images to improve segmentation accuracy. We employ a bidirectional approach, where we select foreground regions from both labeled and unlabeled images and copy them onto the background regions of the opposing images. This process allows the unlabeled images to learn common semantic information from the labeled images, thereby enhancing the segmentation performance.

Furthermore, we employ the BCP operations to generate supervised signals for training the student network. By inputting unlabeled images into the teacher network and applying dynamic soft label filtering, we generate refined pseudo-labels that guide the training process. Additionally, we introduce a confidence filtering mechanism to refine the pseudo-labels further, ensuring that only high-confidence predictions are utilized for training.

This bidirectional copy-paste technique is based on previous research and is effective in generating diverse training data by creating more realistic variations. [64]

4.3.3 Experimental Details

We used 1300 images (from 152 patients) as training set, 70 images (from 10 patients) as validation set, and 570 images (from 67 patients) as test set. The training set used for Pre-train included 130 images with labels, whereas that used for Semi-train comprised 130 images with labels, and 1170 images without labels.

Our training strategy involves several key steps. Firstly, we pre-train a model using the labeled data, establishing a baseline for segmentation performance. Subsequently, we utilize this pre-trained model to generate pseudo-labels for the unlabeled data. These pseudo-labels serve as approximations of the ground truth masks and are crucial for leveraging the unlabeled data during training.

During each training iteration, we optimize the parameters of the network using the Adaptive Moment Estimation (Adam) [77]. Simultaneously, we update the parameters of the teacher network using an exponential moving average (EMA) [78] of the student network’s parameters. This dual optimization process ensures that both networks learn from the available data effectively.

To further improve the segmentation performance, we introduce two key enhancements to the model architecture. As shown in Figure 4.2 and Figure 4.3, we first inte-

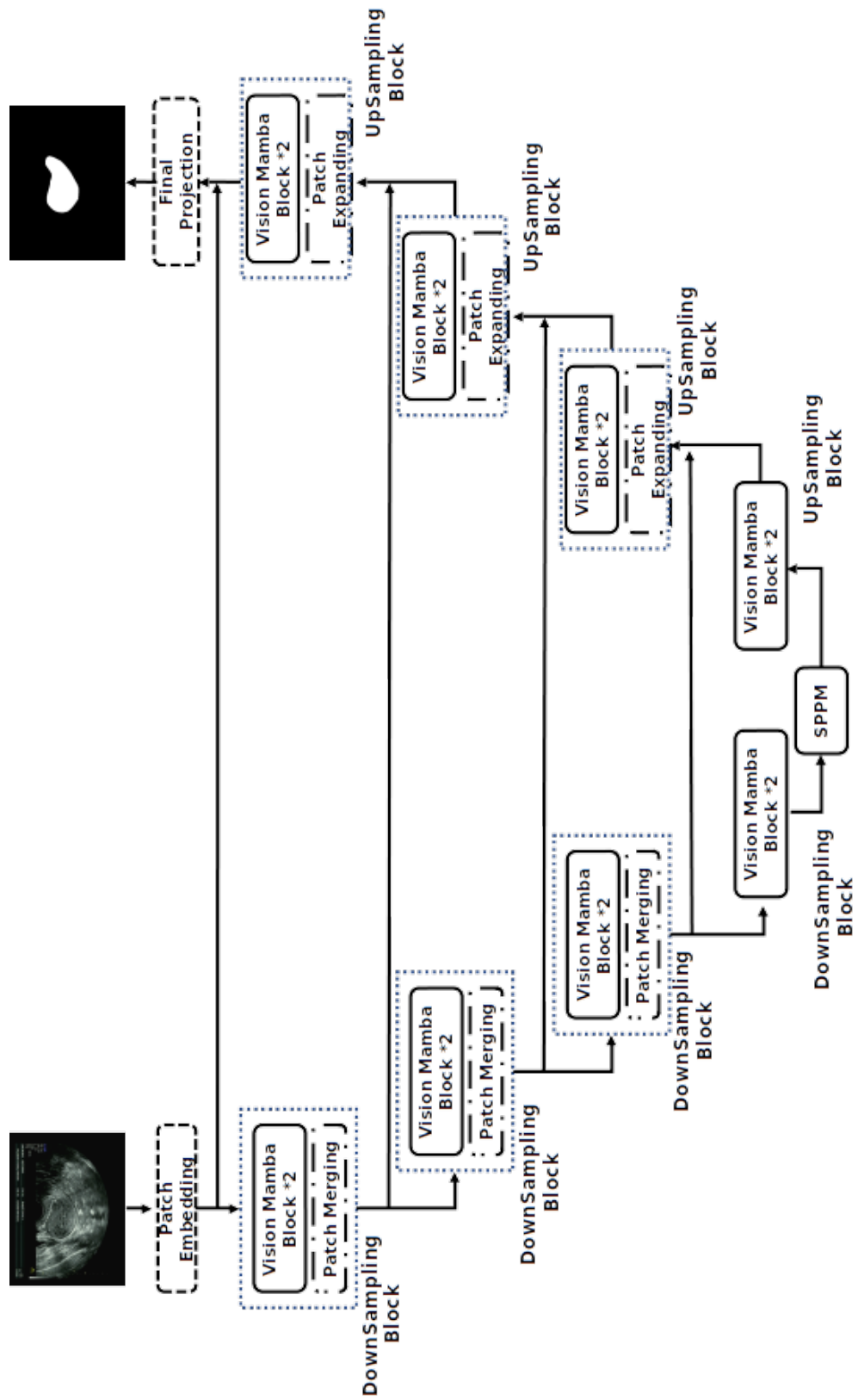


Figure 4.2: Mamba-UNet network model structure

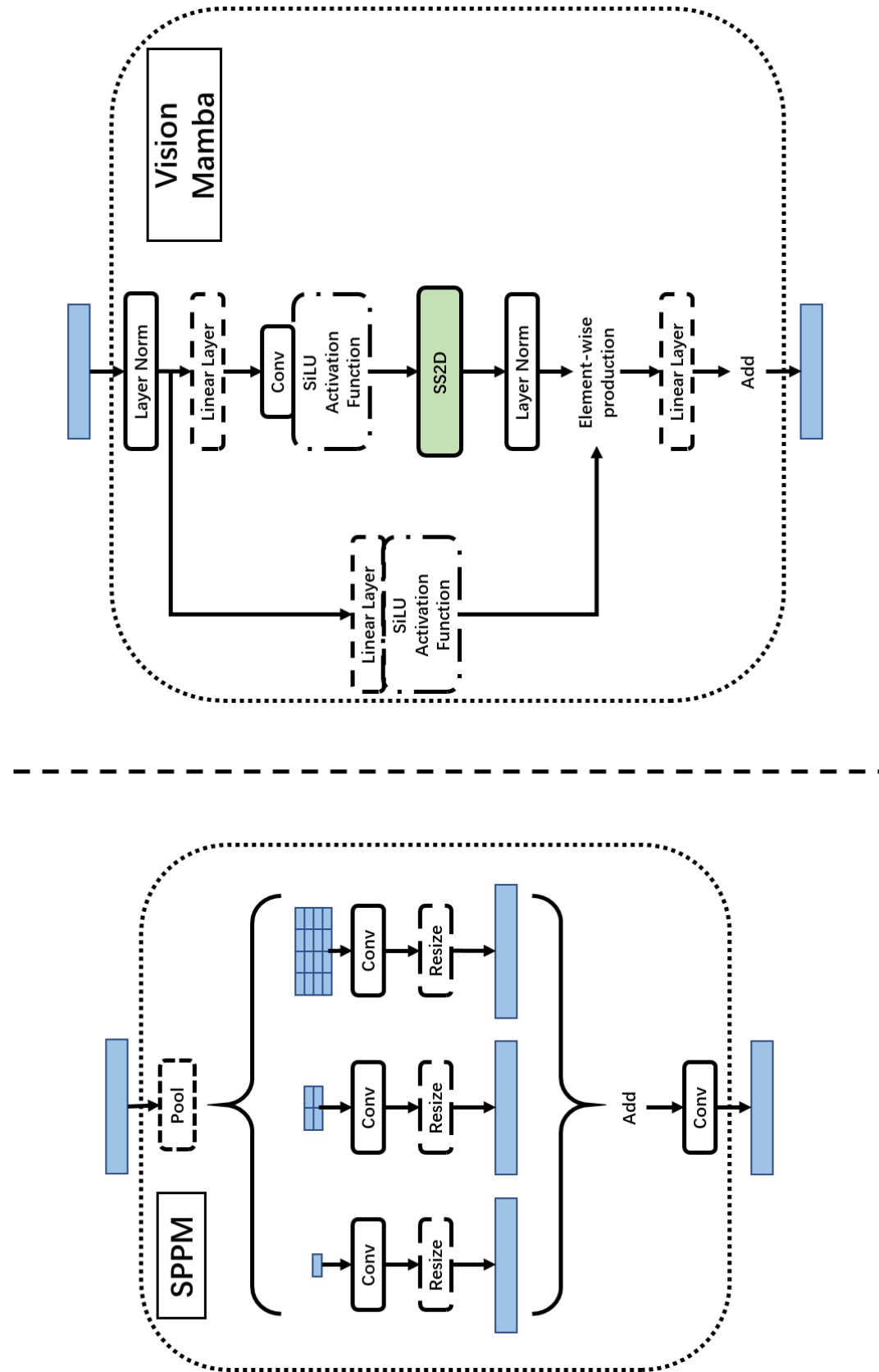


Figure 4.3: Simple pyramid pooling module (SPPM) and Vision Mamba (VM) block

grate the VM [63] module based on the U-Net architecture [1]. This module enhances location-aware visual recognition through location embedding, making the model more robust in dense prediction tasks. The aim of VM is to bring the cutting-edge state space model (SSM), known as Mamba [79], into the realm of computer vision.

The Mamba block is added to better segment the perimetrium of the uterus. The Mamba block has the following characteristics:

- **Bidirectional modeling capability:** Mamba uses a bidirectional SSM, which can analyze both forward and backward directions at the same time to model the data. This bidirectional modeling capability can better capture global contextual information and enhance the model's performance.
- **Position-aware:** Mamba introduces position embeddings, which can provide spatial information perception for visual recognition tasks. This makes Mamba more robust in dense prediction tasks.
- **Efficient computing and memory complexity:** Compared with other SSM-based models, Mamba has higher computing efficiency and lower memory usage. It can save computing resources when processing high-resolution images and enables direct sequential visual representation learning without relying on 2D prior knowledge.

Secondly, we incorporate a simple pyramid pooling module (SPPM) [33]. In Figure 4.3, SPPM begins by integrating input features through the pyramid pooling module, which incorporates three global average pooling operations with bin sizes of 1×1 , 2×2 , and 4×4 . Subsequently, the resulting features undergo convolution and upsampling operations. The convolution operation employs a kernel size of 1×1 , resulting in an output channel smaller than the input channel. Following this, the features are upsampled, and another convolution operation is applied to refine them. In contrast to the original pyramid pooling module (PPM) [35], SPPM reduces the number of intermediate and output channels, eliminates the shortcut, and replaces the concatenation operation with addition. Consequently, SPPM exhibits enhanced efficiency and is better suited for real-time models.

The configuration of the system development platform is Ubuntu 18.04, an I12th Gen Intel® Core™ i7-12700F×20 CPU, and an Nvidia Geforce RTX 3090 GPU.

4.3.4 Experiments

Our model was rigorously compared with the traditional supervised learning model U-Net and the original BCP-Net architecture by benchmarking our approach against these established methods to evaluate its effectiveness in semi-supervised segmentation of transvaginal ultrasound uterine images. And verify its superiority in terms of segmentation accuracy and efficiency. Specifically, we analyzed the segmentation results obtained by our model versus those of the semi-supervised models U-Net and BCP-Net on various evaluation metrics including Dice coefficient (Dice), Jaccard index (Jaccard), average surface distance (ASD), and Hausdorff_95 (HD_95) [80]. HD_95 is a metric for measuring the distance between two-point sets. It is based on the Hausdorff distance but is more robust, especially to outliers. Specifically, the Hausdorff distance measures the maximum of the minimum distances between two sets, HD_95 takes the 95th percentile of these minimum distances rather than the absolute maximum. This reduces the impact

of a few outliers and provides a more stable and robust distance metric. This method is commonly used in medical image processing and computer vision tasks, especially when evaluating the performance of segmentation algorithms. Zhang et al. discussed the robustness and reduced impact of outliers when using the HD_95 metric for evaluating medical image segmentation [81]. Their research demonstrated that using the 95th percentile effectively minimizes the influence of extreme outliers while maintaining a reliable measure of segmentation accuracy. Additionally, Kamnitsas et al. highlighted the practical application of HD_95 distance in their automated brain tumor segmentation study, emphasizing its advantage in reducing the influence of outliers and providing a more stable and robust distance metric [82].

In equations 1-4, A and B represent the predicted and ground truth images, respectively, and a and b represent the pixel points in A and B . d_{AB} : The maximum distance from each pixel in the predicted mask to the nearest target pixel in the ground truth image. d_{BA} : The maximum distance from each pixel in the ground truth image to the nearest target pixel in the predicted mask. d_H : Hausdorff distance.

$$Dice = \frac{2(A \cap B)}{A + B} \quad (4.1)$$

$$Jaccard = \frac{(A \cap B)}{A \cup B} \quad (4.2)$$

$$ASD(A, B) = \sum_{a \in A} \min_{b \in B} d(a, b) / |A| \quad (4.3)$$

$$\begin{aligned} Hausdorff_{95} &= d_H(A, B) \\ &= \max \{ \text{percentile}_{95} d_{AB}, \text{percentile}_{95} d_{BA} \} \\ &= \max \left\{ \max_{a \in A} \text{percentile}_{95} \min_{b \in B} d(a, b), \max_{b \in B} \text{percentile}_{95} \min_{a \in A} d(a, b) \right\} \end{aligned} \quad (4.4)$$

4.4 Results

The experimental results show that the proposed model outperforms BCP-Net and U-Net models in terms of segmentation accuracy. Using unlabeled data in semi-supervised training greatly improves the model's ability to generalize to unseen data, especially in regions with limited labeled samples, and obtain more accurate segmentation results.

As presented in Table 4.1, Dice and Jaccard measure the overlap between predicted segmentation and ground truth. The Dice and Jaccard values of the BCP-Mamba model are 0.8655 and 0.7762, respectively, which are higher than those of the other models. The Dice of the BCP-Net and U-Net models are 0.8072 and 0.8463, respectively, Jaccard is 0.6859 and 0.7401 respectively. Likewise, ASD and HD_95, which quantify the average and maximum differences between segmentation boundaries, respectively, are significantly lower for the BCP-Mamba model, indicating closer proximity to ground-truth annotations. The most obvious difference is with the ASD measure. The ASD values for both the BCP-Net and U-Net models are approximately 20 higher than those for BCP-Mamba. These experimental findings demonstrate the superior performance

of the proposed model compared to both the BCP-Net and U-Net architectures in terms of segmentation accuracy. Despite the increased complexity of the BCP-Mamba model, the proposed achieved a prediction speed of 73 frames per second on our device, which is comparable to the 71 and 76 frames per second achieved by the U-Net and BCP-Net models, respectively.

In the comparison plot of results (Figure 5.3), each network has three distinct sections, indicated by colors representing the ground truth (green), Predicted Results (red), and the part of overlap between the ground truth and Predicted Results (yellow). Notably, the analysis shows that the BCP-Mamba model displays the widest area of overlap between the ground truth and predicted outcomes (yellow) compared to that obtained with the BCP-Net and U-Net architectures. This observation suggests that the BCP-Mamba model achieves more agreement with the ground truth annotations, indicating higher segmentation accuracy and consistency with the underlying anatomical structures in transvaginal ultrasound uterine images.

These findings are consistent with the quantitative assessment metrics discussed earlier, suggesting the improved performance of the proposed BCP-Mamba model in accurately segmenting the plasma membrane layer.

These results underscore the significant advancements achieved by leveraging semi-supervised training with unlabeled data, enhancing the model's ability to capture complex anatomical structures and nuances present in transvaginal ultrasound uterine images. Such improvements hold promise for advancing the diagnostic accuracy and treatment planning in uterine pathology, contributing to enhanced patient care and clinical outcomes.

4.5 Discussion

The segmentation of the perimetrium in transvaginal ultrasound uterine images holds significant clinical importance in the diagnosis and treatment of various uterine pathologies. The accurate delineation of this anatomical structure enables clinicians to assess the integrity of the uterine wall and identify abnormalities such as fibroids, adenomyosis, and endometrial disorders. Our semi-supervised approach leverages both labeled and unlabeled data to improve the segmentation accuracy, particularly in regions with limited labeled samples. By effectively integrating information from unlabeled data, our model achieves a more precise delineation of the plasma membrane layer, which is crucial for diagnosing uterine pathologies.

The utilization of semi-supervised learning techniques in medical image segmentation offers notable advantages, particularly in scenarios where labeled data are scarce or expensive to obtain. Our experimental results show that by utilizing both labeled and unlabeled data, the semi-supervised approach enhances the generalization and robustness of the model, thereby improving the segmentation accuracy. However, the implementation of semi-supervised learning also poses certain challenges. One notable difficulty is the requirement for careful calibration of the hyperparameters and regularization techniques to prevent overfitting and ensure the effective integration of labeled and unlabeled data.

Our experimental results indicate that the proposed semi-supervised segmentation model significantly outperforms both the BCP-Net and U-Net models in segmentation accuracy. Specifically, the Dice coefficient, Jaccard index, ASD, and HD₉₅ met-

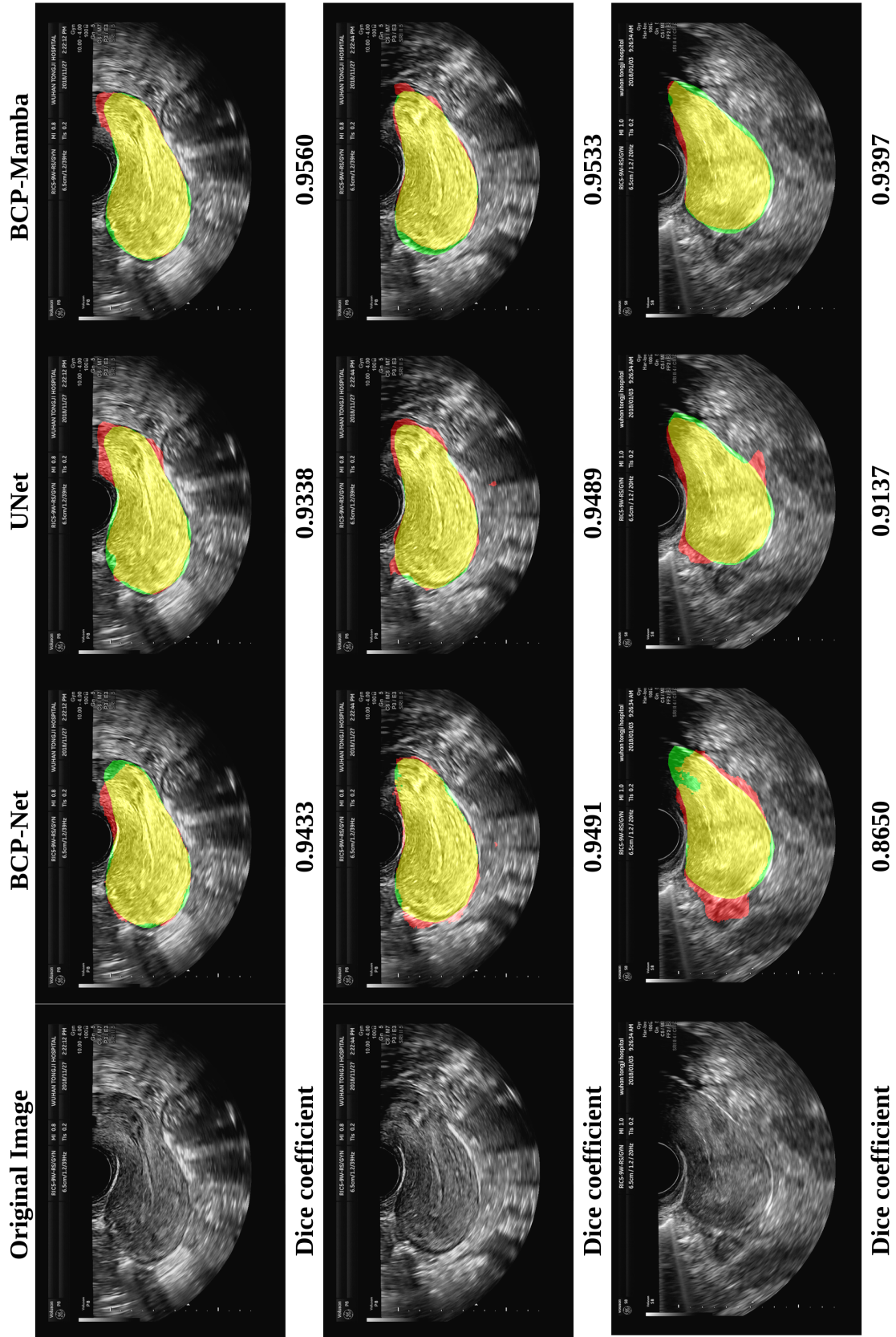


Figure 4.4: Results of the BCP-Net, UNet, BCP-Mamba models. Ground truth (green), predicted results (red), and the part of overlap between the ground truth and predicted results (yellow).

rics consistently show higher values for our model, with mean (SD) values of 0.8655 (0.0710), 0.7762 (0.1051), 40.04 (21.61), and 14.5 (8.9), respectively, highlighting its superior performance. The comparison chart (Figure 5.3) vividly illustrates this improvement, with the BCP-Mamba model showing the largest area of overlap between Groundtruth and predicted results, represented by the yellow region, indicating a higher concordance with Groundtruth annotations and enhanced segmentation precision.

Our approach benefits from integrating the Vision Mamba (VM) module and the Simple Pyramid Pooling Module (SPPM). The VM module enhances location-aware visual recognition, making the model more robust in dense prediction tasks, while the SPPM addresses inconsistencies due to varying image sizes. These enhancements, coupled with the semi-supervised learning framework, allow our model to effectively leverage unlabeled data, improving its generalization to unseen data and yielding more accurate segmentation results.

The reason why Mamba works in vision tasks is that it combines bidirectional SSM, which helps model the global visual context of the data and provides location-aware visual recognition through location embedding. The advantages of Mamba include higher computational and memory efficiency, which makes it suitable for processing high-resolution images, and it performs well on ImageNet classification tasks. In addition, Mamba is more efficient, with lower GPU memory footprint and inference time, which enables it to directly perform sequential visual representation learning without relying on prior 2D information.

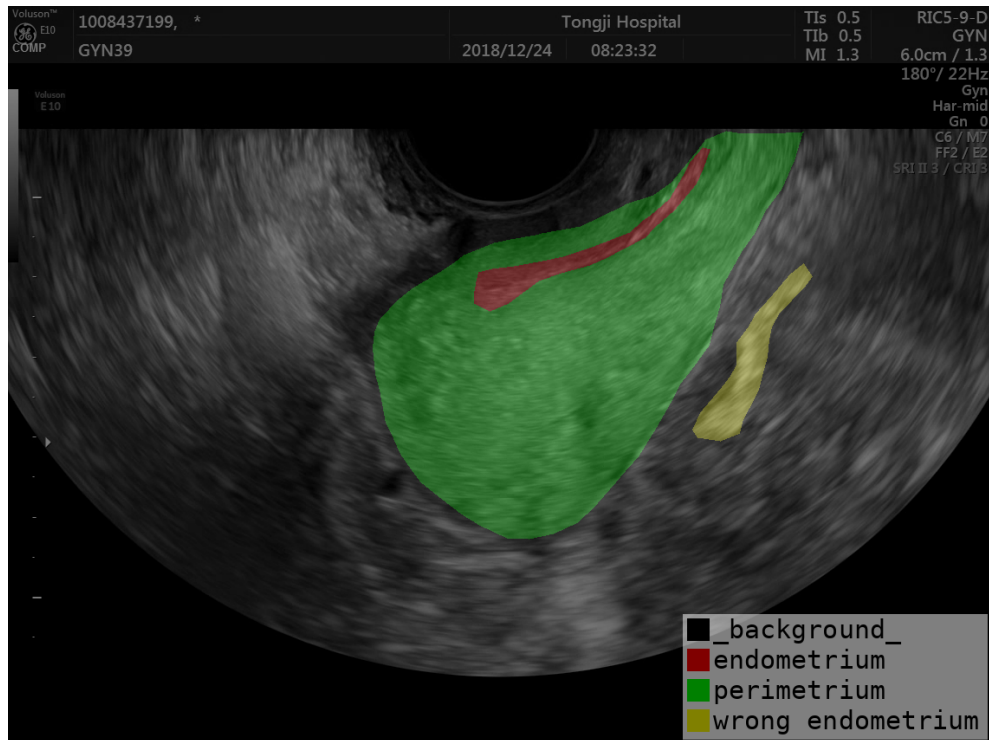


Figure 4.5: Example of endometrial prediction error

Accurate segmentation of the uterine ectoderm improves endometrial thickness measurements, crucial for diagnosing uterine diseases. For example, in the image below, the endometrium is incorrectly categorized as being outside the uterus (marked as yellow in Figure 4.5) due to the uneven echogenicity of the abnormal endometrium. The correct position of the endometrium should be shown as red in Figure 4.5. By segmenting the

perimetrium (marked as green in Figure 4.5) of the uterus first, we can eliminate external interference and improve the accuracy of endometrial segmentation. Our method helps reduce errors from uneven echoes and unclear boundaries, leading to better diagnostic precision and patient outcomes. The semi-supervised learning framework also increases efficiency by using both labeled and unlabeled data, reducing the need for extensive manual annotations.

A significant limitation is the scarcity of publicly available datasets specifically for transvaginal uterine ultrasound imaging. This scarcity restricts the training and validation of segmentation algorithms, leading to challenges in developing models that are both robust and generalizable. The size of the dataset used in our study was only from a medical center and relatively small, and the performance of the model may vary when applied to larger and more diverse datasets, which requires further validation and refinement. Moreover, the datasets that do exist often lack diversity in terms of patient demographics, clinical conditions, and imaging protocols. To address this limitation, we plan to collaborate with multiple medical centers to compile a more extensive and diverse dataset. This collaboration will involve both retrospective and prospective studies to rigorously test and validate our segmentation algorithms.

4.6 Conclusion

In this paper, we present BCP-Mamba semi-supervised model for segmenting the perimetrium. Our method leverages labeled and unlabeled data to improve the segmentation accuracy, outperforming other semi-supervised models such as U-Net and BCP-Net in experimental evaluations. The satisfactory results achieved demonstrate the potential of semi-supervised learning methods in enhancing medical image segmentation tasks, especially when labeled data is limited. In addition, ectoderm segmentation technology is critical to improving the practicality of transvaginal ultrasound imaging in clinical practice and improving patient care. To ensure the extensiveness of the model, we will verify its effectiveness on other datasets in the future.

Table 4.1: Comparison of evaluation parameters of the model.

Models	Dice(mean(SD))	Sensitivity(mean(SD))	Specificity(mean(SD))
U-Net	0.8072(0.1046)	78.74%(13.75%)	98.10%(1.27%)
BCP-Net	0.8463(0.0979)	94.39%(7.10%)	96.67%(1.71%)
BCP-Mamba	0.8655(0.0710)	89.82%(12.20%)	97.96%(1.64%)

Models	Jaccard(mean(SD))	ASD(mean(SD))	HD_95(mean(SD))
U-Net	0.6859(0.1455)	58.59(32.19)	23.0(9.8)
BCP-Net	0.7401(0.1229)	59.70(25.87)	21.3(10.9)
BCP-Mamba	0.7762(0.1051)	40.05(21.61)	14.6(8.9)

Chapter 5

Deep Learning Application for Transvaginal Uterine Ultrasound Images Processing(Semi-Supervised Segmentation, Multi-StudentNet)

5.1 Introduction

The endometrium is the innermost layer of the uterus and plays a vital role in reproductive health [83]. The endometrium undergoes cyclical changes throughout the menstrual cycle [84]. Accurate segmentation of the endometrium is critical for the evaluation of various gynecological diseases, including infertility, abnormal uterine bleeding, and endometrial cancer [85] [86]. Effective segmentation can provide insight into the structure and abnormalities of the endometrium, aiding in early diagnosis and treatment planning [58].

Transvaginal ultrasound (TVUS) is a preferred imaging modality for its high resolution and non-invasive nature, providing detailed images of the pelvic organs, including the endometrium [87]. However, the current clinical practice for endometrial segmentation relies on manual delineation by experienced clinicians. This process is labor-intensive and time-consuming, with considerable variability both between different observers and within the same observer over time [88] [89]. Such variability can lead to inconsistent assessments, adversely affecting clinical decision-making and patient outcomes [90]. Therefore, automated solutions are urgently needed to provide accurate and consistent segmentation of the endometrium, reducing the burden on clinicians and improving diagnostic reliability.

Automating the segmentation of the endometrium in ultrasound images presents unique challenges. Ultrasound images are often characterized by poor contrast, speckle noise, and anatomical variability, complicating the segmentation process. In recent years, deep learning has become a potent tool for medical image analysis, achieving notable success across various applications. R. Almajalid et al. developed a deep learning framework for breast ultrasound (BUS) image segmentation using a modified U-Net architecture. Their method, tested on 221 BUS images, achieved superior results with a DSC of 0.825 and a similarity rate of 0.698, demonstrating improved robustness and accuracy in tumor segmentation compared to existing methods [91]. Kerkeni A et al. proposed a multiscale region growing (MSRG) method for coronary artery segmentation in

2D X-ray angiograms, achieving 80% accuracy in easier cases and 70% in challenging ones with a mean precision of 82%. The MSRG method, effective with the Frangi filter, outperforms other techniques in sensitivity and robustness against noise, stenosis, and poor contrast [92]. Guo et al. developed an expanded U-Net for breast ultrasound image segmentation, achieving high accuracy with an average DSC of 90.5% and IOU of 82.7%. This method improves texture detail and edge feature retention, outperforming the general U-Net [93]. Zheng et al. proposed an Improved Cascade Mask R-CNN for thyroid nodule detection and recognition in ultrasound images, achieving an mAP of 87.1% and recognition accuracy of 98.67%. The model includes an enhanced detector, balanced L1 loss function, and soft non-maximum suppression, improving localization and accuracy to assist radiologists effectively [94].

However, there is a paucity of studies specifically addressing semi-supervised methods for endometrial segmentation. Most existing approaches rely on fully supervised learning, which requires extensive labeled datasets — a significant limitation in medical imaging where annotated data is often scarce and expensive. Semi-supervised learning can bridge this gap by effectively utilizing both labeled and unlabeled data, thereby reducing the dependency on large annotated datasets.

In this paper, we proposed Multi-StudentNet, a semi-supervised deep-learning model for endometrial segmentation in TVUS images. Our approach efficiently combines labeled and unlabeled data, significantly minimizing the requirement for extensive manual annotation while preserving accuracy. In addition, because the model uses weights from multiple models, feature sharing is achieved, thereby enhancing the robustness and accuracy of the model.

5.2 Methods

In this study, we proposed Multi-StudentNet, a semi-supervised deep learning framework for endometrial segmentation in transvaginal ultrasound (TVUS) images. Our approach effectively integrates both labeled and unlabeled data to enhance model training. The methodology encompasses three primary stages: data preprocessing, model training, and validation.

5.2.1 Dataset and preprocessing

As illustrated in Figure 5.1, a dataset, comprising 1664 endometrial images from various patients, was collected from Tongji Hospital of Huazhong University of Science and Technology. Each image was annotated by our team and subsequently verified by a professional physician to ensure accuracy. The dataset includes three types of cases: normal, polyp, and cancer. Polyp and cancer cases present greater challenges in distinguishing endometrial boundaries due to uneven echoes and blurred edges.

The dataset was divided into three subsets: training, validation, and test sets. The training set was further split into 597 labeled and 597 unlabeled images, totaling 1194 images. The validation set and test set consisted of 170 and 300 images, respectively.

During training, half of the training set was used for supervised training of the teacher model, while the remaining data facilitated semi-supervised training of the student models. This strategic division allows the model to leverage both labeled and unlabeled data, enhancing its learning capability and segmentation performance.

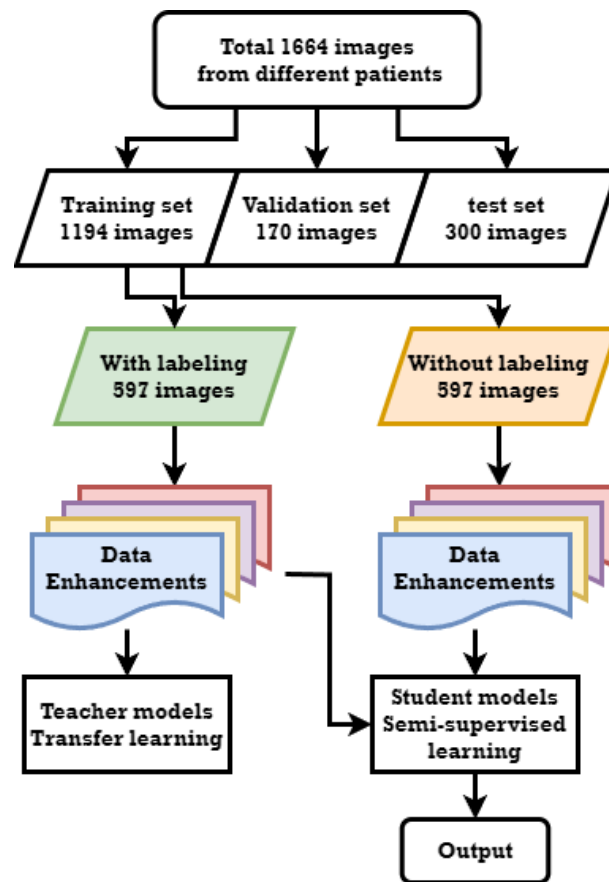


Figure 5.1: Dataset utilization flowchart for semi-supervised segmentation.

To expand the dataset, various data augmentation techniques were applied during preprocessing, including random horizontal flips, random rotations, Gaussian noise addition, and color jittering. These techniques increased the diversity of the training data, improving the model’s robustness and generalizability. Additionally, all images and masks were resized to a standard size to ensure uniformity.

5.2.2 Experimental setup

Division of the data set

The experimental setup, illustrated in Figure 5.2, was designed to implement and evaluate our semi-supervised learning framework for endometrial segmentation. The dataset was split into two sections: one designated for training the teacher models and the other for training the student models. Separate DataLoader instances were utilized for training and validation sets to ensure efficient data handling.

Teacher Models Training

Three different teacher models (DeepLabV3-ResNet50, FCN-ResNet50, and DeepLabV3-ResNet101) [95] [96] [97] were initialized with pre-trained weights. The final layers of these models were modified for the semantic segmentation task. Each teacher model underwent transfer learning using the labeled data to adapt to our dataset. The models were trained over multiple epochs using a composite loss (equation (5.1)) function that combined label smoothing cross-entropy loss (equation (5.2)) and Dice loss (equation (5.3)) [98] [99]. Label smoothing cross-entropy loss helps in handling class imbalance and preventing overfitting by smoothing the labels, while dice loss specifically enhances boundary accuracy by focusing on the overlap between predicted and true segmentation. Performance was evaluated on both training and validation sets, and the best-performing model weights were saved based on validation metrics. The loss function formulas used are as follows.

The total loss is a weighted sum of the label smoothing cross-entropy loss L_{LS} and Dice loss L_{Dice} :

$$\mathcal{L}_{\text{Total}} = \alpha \mathcal{L}_{\text{LS}} + \beta \mathcal{L}_{\text{Dice}} \quad (5.1)$$

where α and β are hyperparameters that control the contribution of each loss component.

The label smoothing cross-entropy loss L_{LS} is defined as:

$$\mathcal{L}_{\text{LS}} = - \sum_{i=1}^K \left[(1 - \epsilon) y_i + \frac{\epsilon}{K} \right] \log q_i \quad (5.2)$$

where y_i is the true label, q_i is the predicted probability, K is the number of classes, and ϵ is the smoothing parameter.

The Dice loss L_{Dice} is defined as:

$$\mathcal{L}_{\text{Dice}} = 1 - \frac{2 \sum_{i=1}^N p_i g_i}{\sum_{i=1}^N p_i + \sum_{i=1}^N g_i} \quad (5.3)$$

where p_i is the predicted binary mask and g_i is the ground truth binary mask.

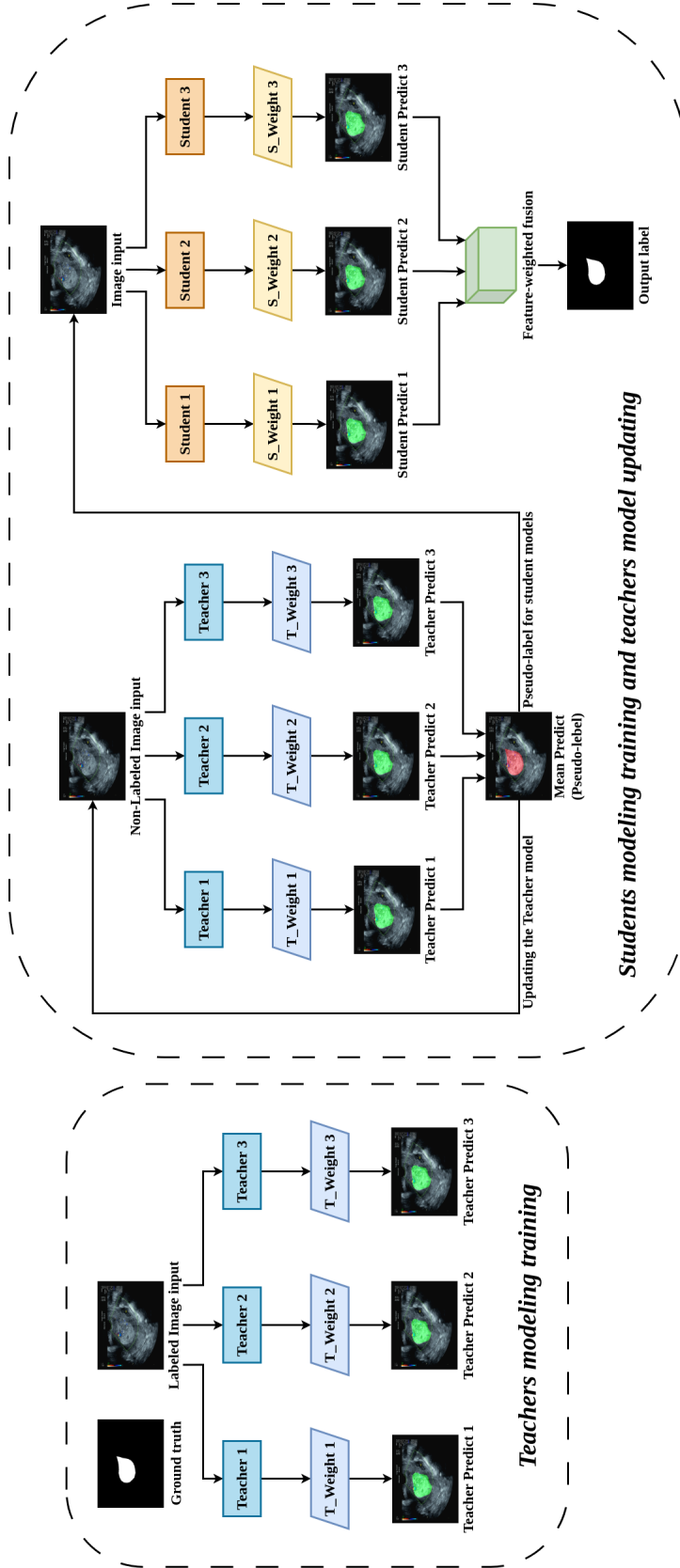


Figure 5.2: Multi-StudentNet Framework for Endometrial Segmentation.

Student Models Training

After training the teacher models, we initialized the student models with pre-trained weights and adapted their final layers for the semantic segmentation task. The training employed a semi-supervised approach where pseudo-labels were generated by averaging the outputs of the teacher models on the unlabeled data. These pseudo-labels, along with the labeled data, were used to train the student models. To ensure accuracy, pseudo-labels were periodically updated throughout the training process. Additionally, pseudo-labels were used to further fine-tune the teacher models, enhancing their ability to generate more accurate labels. The training process incorporated both labeled and pseudo-labeled data, with models optimized using the composite loss (5.1) function over several epochs.

Evaluation Indicators

To comprehensively assess the performance of the proposed semi-supervised segmentation model, we used a series of evaluation metrics, including the DSC, Specificity, Sensitivity and precision with standard deviation(SD).

Their formulas are as follows:

$$DSC = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (5.4)$$

$$Precision = \frac{TP}{FP + TP} \quad (5.5)$$

$$Sensitivity = \frac{TP}{FN + TP} \quad (5.6)$$

$$Specificity = \frac{TN}{FP + TN} \quad (5.7)$$

(TP: True positive, FP: false positive, TN: true negative, FN: false negative)

DSC is commonly employed to assess the similarity between two samples, with values ranging from 0 to 1. It is frequently utilized in the field of medical imaging for the purpose of image segmentation. In this context, a DSC value of 1 signifies the optimal segmentation outcome, while a value of 0 indicates the poorest result. [36].

5.3 Result

Table 5.1: Comparison of evaluation indexes(SD) of the model.

Model	DSC(SD)	Specificity(SD)	Sensitivity(SD)	Precision(SD)
Mean of Supervised Teacher Model	0.8166(0.00006)	99.57%(0.0014%)	82.26%(0.03%)	82.26%(0.03%)
Mean-Teacher DeeplabV3-Resnet50	0.7851(0.0386)	99.34%(0.47%)	82.16%(10.21%)	77.16%(10.21%)
Mean-Teacher FCN-Resnet50	0.7951(0.0411)	98.97%(0.64%)	86.19%(9.78%)	77.19%(9.78%)
Mean-Teacher DeeplabV3-Resnet101	0.7368(0.0580)	97.78%(1.07%)	90.86%(8.44%)	66.86%(8.44%)
Mean-Teacher Multi-Student	0.8075(0.0418)	99.06%(0.62%)	87.09%(8.78%)	77.09%(8.78%)

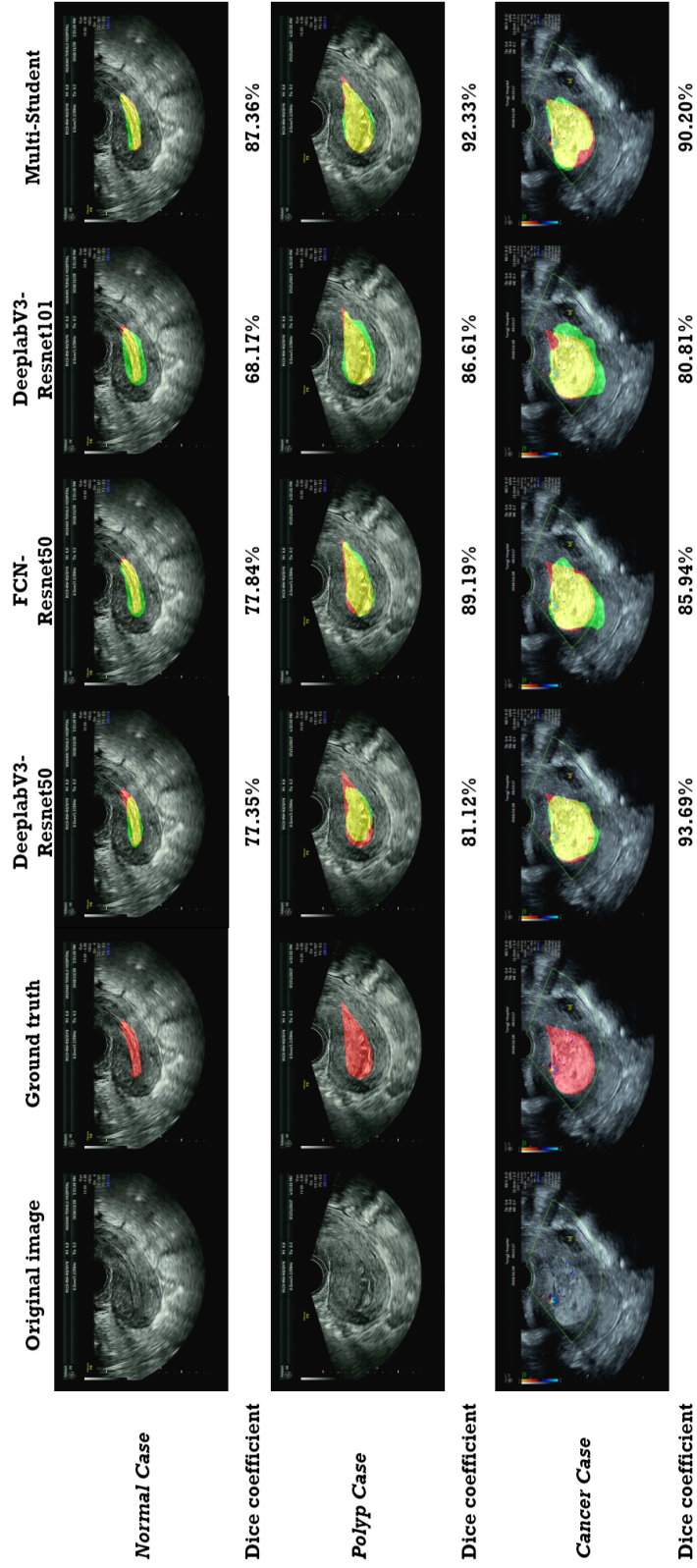


Figure 5.3: Example of results for the TVUS data in different models.

The proposed semi-supervised segmentation model, Multi-StudentNet, was rigorously evaluated using a dataset of transvaginal ultrasound images. The performance of the model was assessed using several evaluation metrics, including the DSC, specificity, sensitivity and precision. The evaluation results are summarized in Table 6.2 and visually represented in Figure 5.3.

Figure 5.3 illustrates the segmentation performance of the student models and their integrated output across different cases, including normal, polyp, and cancer conditions. Each row of images presents the original ultrasound image, ground truth, predictions from student models Mean-Teacher DeeplabV3-Resnet50 (MTDR50), Mean-Teacher FCN-Resnet50 (MTFR50), and Mean-Teacher DeeplabV3-Resnet101 (MTDR101), and the integrated result from all student models. The DSCs for each model and the integrated output are also provided.

In the normal case, the integrated student model achieved a DSC of 87.36%, outperforming individual student models MTDR50 (77.35%), MTFR50 (77.84%), and MTDR101 (68.17%). For the polyp case, the integrated student model achieved a DSC of 92.33%, which is higher than the individual student models MTDR50 (81.12%), MTFR50 (89.19%), and MTDR101 (86.61%). In the cancer case, the integrated student model achieved a DSC of 90.20%, demonstrating superior performance compared to individual student models MTFR50 (85.94%), and MTDR101 (80.81%).

In contrast to the teacher model with fully supervised learning, the multi-student model achieves a close level of proximity when trained with unlabeled data.

The overall performance of Multi-StudentNet demonstrated a high DSC of 0.81, indicating a strong overlap between the predicted segmentation and the ground truth. The model maintained high specificity (0.99) across all samples, reflecting its ability to correctly identify negative cases and minimize false positives. Sensitivity was 0.87, demonstrating the model's robustness in capturing true positives, which is crucial for clinical diagnostics. The slightly lower precision (0.77) suggests an area for improvement in reducing false positives.

5.4 Discussion

The evaluation results demonstrate that our semi-supervised learning framework for endometrial segmentation in transvaginal ultrasound images achieves high accuracy and reliability. The overall DSC of 0.81 indicates strong agreement between the predicted segmentation and the ground truth, which is crucial for clinical applications where precise delineation of the endometrium is necessary. The exceptionally high specificity (0.99) across all models suggests that the framework is effective in identifying negative cases, thereby minimizing false positives that can lead to unnecessary clinical interventions.

The sensitivity metrics at 0.87 for the overall model, reflect the framework's robustness in detecting true positives. These metrics are particularly important in clinical settings where missing a true positive could have significant implications for patient care. The lower precision of 0.77 indicates that there is still room for improvement in reducing false positives, which would further enhance the model's clinical applicability by reducing the potential for overdiagnosis.

When examining the individual student models, we observed slight differences in performance. Models S1 and S2 showed similar performance, with DSCs of 0.79 and

0.80, respectively. Both models demonstrated high specificity and strong sensitivity, indicating their reliability. Model S3 had the lowest DSC (0.74) but the highest sensitivity (0.91). This indicates that while Model S3 was effective at detecting true positives, it produced more false positives, as reflected by its lower precision (0.63).

These changes show the importance of model selection and tuning in a semi-supervised learning framework. The performance differences between the student models suggest that further optimization, such as fine-tuning model parameters or incorporating additional unlabeled data, could improve overall performance. In addition, ensemble methods that combine the strengths of multiple models could be explored to achieve more balanced and robust results.

In addition, our approach also has important clinical implications. Accurate and consistent endometrial segmentation is essential for the diagnosis and management of various gynecological diseases, including infertility, abnormal uterine bleeding, and endometrial cancer. By reducing the reliance on manual demarcation, our semi-supervised framework not only improves efficiency but also minimizes inter- and intra-observer variability, leading to more consistent and reliable diagnosis.

However, our study still has several limitations. The size and diversity of the dataset may affect the generalizability of the model. Our study used a specific set of transvaginal ultrasound images, and further validation on larger and more diverse datasets is needed to confirm the robustness of the model in different populations and clinical settings. In addition, although semi-supervised methods reduce the need for extensive manual annotation, they still require a large amount of labeled data for initial training. Future work could explore fully unsupervised methods or active learning strategies to further reduce the need for manual labeling.

In summary, our semi-supervised deep learning framework for endometrial segmentation showed promising results with high accuracy and reliability. This method addresses many limitations of current segmentation techniques and provides a reliable and clinically relevant solution. Future research should focus on further optimizing the model, validating its performance on different datasets, and exploring advanced learning strategies to enhance its clinical applicability.

5.5 Conclusion

This paper presents Multi-StudentNet, a semi-supervised deep-learning framework for endometrial segmentation in transvaginal ultrasound images. By effectively integrating labeled and unlabeled data through multiple student models, our approach significantly reduces the need for extensive manual annotations while maintaining high accuracy, demonstrated by a DSC of 0.81 and specificity of 0.99. This framework addresses the limitations of manual segmentation methods, enhancing diagnostic precision and workflow efficiency in clinical settings. The use of multiple student models not only improves learning capability but also offers potential for optimization through ensemble techniques.

Future work will focus on expanding the dataset and employing advanced learning strategies to further validate and enhance the robustness of Multi-StudentNet. Additionally, applying this model to different types of medical imaging data will test its generalizability and adaptability, ensuring its broader applicability in various clinical scenarios.

Chapter 6

Deep Learning Application for Transvaginal Ultrasound Video Images Processing(Keyframe extraction)

6.1 Introduction

The endometrium, the lining of the uterus, is crucial for women's reproductive health. It undergoes monthly changes in sync with menstrual cycles, influenced by the body's reproductive hormones, oestrogen, and progesterone [100]. Endometrial measurement is an indispensable part of gynaecological ultrasonography, which is the first-line diagnostic tool for endometrial diseases [101] [102]. Endometrial thickness (ET) is the maximum measurement of endometrial layers on a long-axis transvaginal view of the uterus to depict an endometrial echo, including anterior-posterior thickness [103]. However, the proficiency of ultrasound scanning varies among practitioners and there is a lack of standardized methods to assess scanning quality [104].

Transvaginal ultrasound (TVUS) is a diagnostic imaging modality widely used to evaluate uterine structures [105]. Although TVUS is safe and effective, it suffers from challenges such as image noise, operator dependence, and variability in diagnostic accuracy [105] [106]. The measurement of ET involves the assessment of morphological and physiological indicators, including ET, stage, and volume, as well as several factors, such as endometrial peristaltic waves and blood flow [103]. Sonographers typically select representative frames for ET measurements based on visual indicators, which is a time-consuming and subjective process, especially for less-experienced practitioners [107]. Therefore, we developed an automated system called EndoUSScan to identify key frames from ultrasound videos, which could significantly improve the efficiency and accuracy of ET measurement.

The emergence of deep learning (DL) has revolutionized medical image analysis, although its application to TVUS imaging remains relatively in a preliminary stage because of limited data [108]. While DL has been successfully applied to tasks like tumour segmentation in brain and colon images, challenges persist because of the scarcity of annotated medical images for training convolutional neural networks (CNNs) [109]. To address these challenges, innovative approaches, such as the boundary-weighted domain adaptive neural network (BOWDA-Net) for prostate magnetic resonance imaging segmentation, have been proposed [103]. Moreover, new DL models like DoubleU-Net have enhanced segmentation performance across various medical image datasets [110].

In recent years, some studies have explored the use of DL for measuring ET in static ultrasound images. For instance, Hu et al. developed a fully automated method using U-Net for endometrial segmentation and achieved promising results on test datasets [111]. Similarly, in our previous study, a two-step method demonstrated accurate ET measurement using TVUS images [58]. This method used the maximum inscribed circle technique for precise ET calculation. However, both methods relied on the manual selection of images for ET measurement, which introduces subjectivity and repetition into the process.

In this study, we propose a keyframe detection system called EndoUSScan for automated keyframe extraction from ultrasound videos to facilitate ET measurement. Our method aims to improve diagnostic accuracy and facilitate training for inexperienced sonographers by reducing their subjectivity and standardizing ET measurements.

6.2 Related work

Keyframe detection plays a crucial role in facilitating indexing and classification related to ultrasound imaging. Recent advancements in computer-aided detection systems have leveraged ultrasound images for neural recognition, showing promising potential for cost-effective implementation [112]. Furthermore, studies indicate that keyframe detection can significantly enhance cancer detection rates in breast ultrasound imaging for a substantial portion of detected lesions [113].

Various approaches have been proposed for keyframe detection in ultrasound (US) videos. Ciompi et al. focused on detecting keyframes in US videos based on morphological features [114], while Baumgartner et al. employed convolutional neural networks (CNNs) to classify keyframes and determine their alignment with the 12 standard planes of fetal ultrasound imaging [115]. Similarly, Stoean et al. undertook the classification of four keyframes specific to the fetal heart. Pu et al. employed a combination of CNN features extracted from US sequences and optical flow, followed by recurrent neural network (RNN) fusion for standard plane identification [116]. These methodologies are tailored to fetal US videos, targeting various anatomical structures characterized by distinct appearances.

Beyond fetal ultrasound, keyframe detection holds promise for enhancing the analysis of inner anatomical structures through content-based image processing and analysis. In particular, keyframe detection algorithms have been instrumental in facilitating the 3D reconstruction of anatomical structures, thereby augmenting detection capabilities in medical imaging [113]. However, despite these advancements, keyframe extraction techniques for uterine ultrasound videos remain nascent and lack standardization. Therefore, it is crucial to combine clinical knowledge and deep learning to further guide keyframe extraction for transvaginal uterine ultrasound.

6.3 Methods

6.3.1 Study Design and Ethics

A prospective observational study was conducted to develop and evaluate our novel methodology for automated key frame detection in transvaginal uterine ultrasound images. The primary objective was to assess the effectiveness and reliability of the pro-

posed method in accurately identifying key frames for endometrial thickness measurement.

Ethical approval for this retrospective study was obtained from the Institutional Review Board of Tongji Hospital, Huazhong University of Science and Technology.

6.3.2 Dataset Collection and Processing

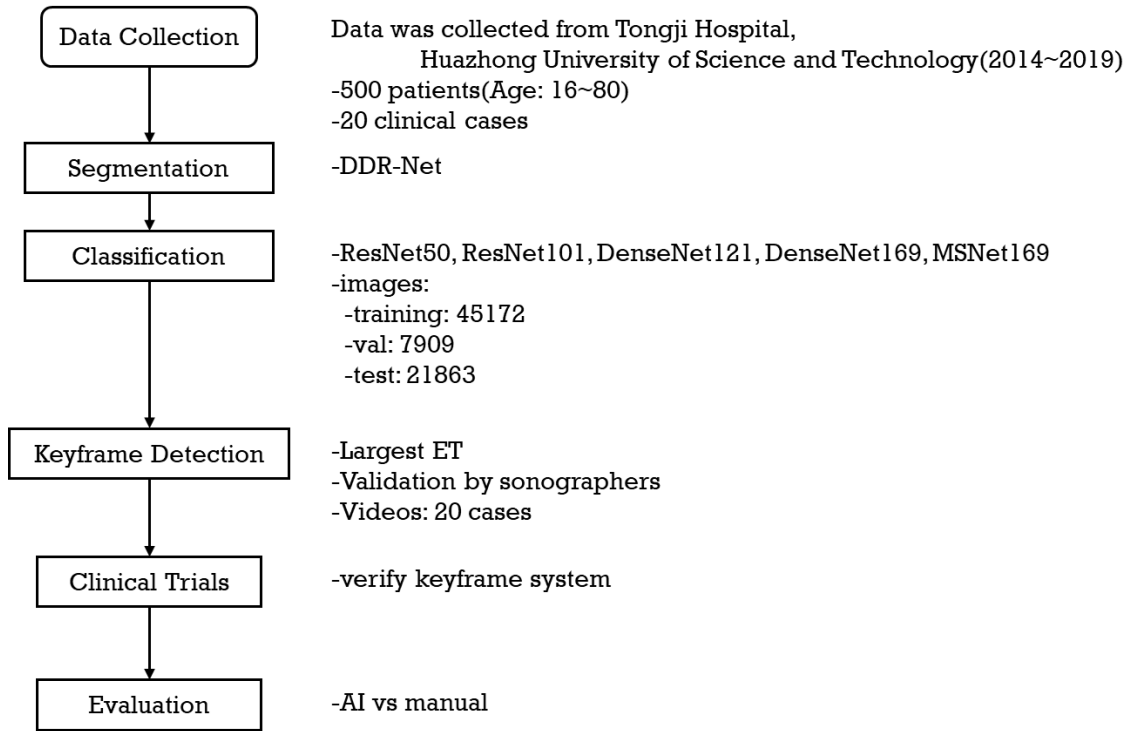


Figure 6.1: Absolute endometrial thickness errors measured by the proposed method

A comprehensive dataset of uterine ultrasound videos was collected from Tongji Hospital, Huazhong University of Science and Technology. As shown in Figure 6.1, the dataset contains 349 videos (107,054 images) representing a variety of patient demographics and clinical scenarios. Patients ranged in age from 16 to 80 years, ensuring a broad representation of the cohort population. Data collection occurred between 2014 and 2019 to capture potential temporal changes in ultrasound image features. The ultrasound examinations were performed using a GE Voluson E10 ultrasound machine, renowned for its advanced imaging capabilities and high-resolution transducers. To ensure consistency and standardization, the ultrasound examinations were conducted by a team of experienced sonographers following a standardized protocol. The acquisition parameters, such as the transducer frequency, gain, and depth settings, were optimized to achieve optimal image quality and clarity. In addition, to improve the quality of the images, and to be able to ensure the smooth running of the experiments, we have done the following processing of the dataset: (1) Preprocessing: the image is converted from Dicom data to PNG image and resampled to PNG image according to Tag:(0028,0004) Photometric Interpretation of Dicom file. We receive Dicom data with two Photometric interpretation values, RGB and YBR FULL 422. (2) Image widening. Including deflation, random rotation, random flip, contrast enhancement, and so on (3) Deep learning

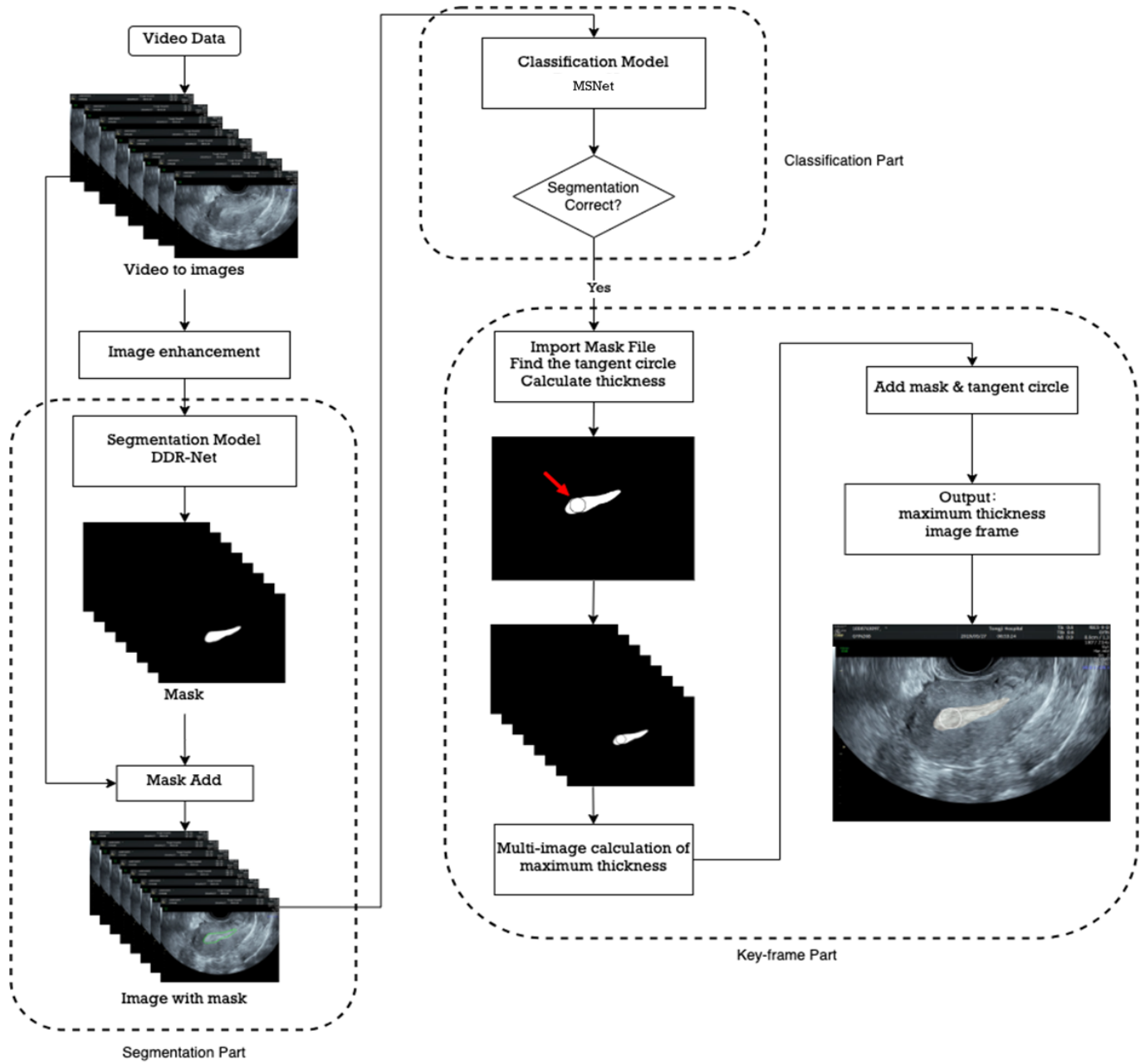


Figure 6.2: Methods of processing

dataset cross-validation, randomly split the case-based dataset into train, validate, and test.

6.3.3 Segmentation

Data

1,050 TUVS images from patients at the ages of 16–80 years were collected in 2014–2018 at the Gynecology and Obstetrics Division, Tongji Hospital, Huazhong University of Science and Technology (HUST). These data were collected by TUVS devices of GE Voluson E1 (GE Corp., United States). the original data’s resolutions and sizes differed depending on the parameters used during the pelvic examination. We rescaled the images to 852×1136 to normalize the input size and lower the processing cost. The data were then randomly split into a training set (840 images from 302 cases) and a test set (210 images from 68 cases). We also guaranteed that the images in the training set are from different cases in the test one. This study has been approved by the Institutional Review Board of Tongji Hospital, HUST.

Models

As shown in Table 6.1, the models are combinations of different backbones and segmentation architectures. Backbones were selected from Resnet-50, Vanilla CNN, Vanilla Mini CNN, and VGG-16. These backbones were all pre-trained by over one million images in ImageNet [117]. Segnet and U-Net are semantic segmentation models. DDRNets are dual resolution networks with deep high resolution representation for high resolution images.

Quantitative evaluations show that Segnet is more efficient in terms of time and memory usage in comparison to other architectures [34]. Segnet is an open source project for image segmentation developed by a team at the University of Cambridge. The implementation of image segmentation consists of a convolutional neural network that has two main components: encoder and decoder. The encoder itself is actually a sequence of convolutional networks. The encoder network consists of a convolutional layer, a pooling layer and a BatchNormalization layer. Decoder networks map the objects to specific pixel points. The decoder up-samples the encoded feature images and then convolves the up-sampled images to refine the geometry of the objects and compensate the loss of details caused by the shrinking of the objects by the pooling layer in the encoder [118].

U-Net, initially a convolutional neural network for 2D image segmentation, won the ISBI 2015 Cell Tracking Challenge and the Dental Caries Detection Challenge, respectively [1]. U-Net is a U-shaped network that was proposed to solve the problem of medical image segmentation. The structure is to convolve and pool the images first. U-Net includes 4 pooling layers. If the size of the original image is 224×224 , the feature maps would become 112×112 , 56×56 , 28×28 , and 14×14 with four different sizes. After four times of upsampling, we can get a prediction result of 224×224 with the same size as the input image. The data of medical images are small and difficult to obtain, and the amount of data may be only a few hundred or even less than 100. As a result, overfitting is common if large networks are used. Therefore, the lightweight U-Net is advantageous in case of small training data [1].

DDRNets start with a backbone and then split into two parallel deep branches with different resolutions [2]. One branch generates relatively high-resolution feature maps, capturing detailed spatial information necessary for accurate boundary delineation. The other branch extracts rich semantic information through multiple downsampling operations, enabling the network to understand the global context of the image. This dual-branch structure ensures that the model simultaneously captures fine-grained details and broader semantic information, addressing the inherent trade-off between spatial resolution and semantic depth.

To facilitate effective communication between the two branches, DDRNets employ multiple bilateral connections, which allow high-resolution features to inform the semantic branch and vice versa. This bidirectional information fusion improves the model's ability to integrate multi-scale features, enhancing its performance on tasks requiring precise localization and contextual understanding. These connections act as "bridges," enabling the network to leverage complementary strengths of both branches while maintaining computational efficiency.

In addition to this innovative dual-branch design, DDRNets incorporate a new module called DAPPM (Dual-Resolution Adaptive Pyramid Pooling Module). DAPPM is specifically designed to handle low-resolution feature maps. It extracts multi-scale contextual information through pyramid pooling and merges them in a cascade fashion, enabling the network to effectively integrate both local and global context. By focusing on low-resolution inputs, DAPPM enhances the network's ability to capture global dependencies, a critical feature for improving semantic segmentation accuracy.

Before training on the semantic segmentation dataset, the dual-resolution network is pre-trained on ImageNet, leveraging transfer learning to initialize the model with robust feature representations. This pre-training significantly accelerates convergence and improves generalization. Unlike the original ResNet, which uses a 7×7 convolutional layer in the input stem, DDRNets replace this layer with two consecutive 3×3 convolutional layers. This modification reduces computational overhead while preserving spatial information, making the network more efficient for high-resolution image processing.

The remaining basic blocks of ResNet are utilized to construct the backbone and the subsequent dual branches, ensuring a balance between computational efficiency and expressive power. To extend the output dimensions, a bottleneck block is added at the end of each branch, further enhancing the network's representational capacity. Importantly, all modules, including the backbone and branches, are pre-trained on ImageNet, with the exception of the segmentation header and the DAPPM module. This selective pre-training ensures that the most critical components of the network are optimized for general feature extraction, while task-specific modules are fine-tuned during the segmentation training phase.

This carefully designed architecture not only achieves state-of-the-art performance on semantic segmentation tasks but also demonstrates remarkable adaptability to various input resolutions and dataset characteristics. By combining high-resolution feature preservation, semantic richness, and multi-scale contextual understanding, DDRNets exemplify a robust and versatile approach to semantic segmentation.

Fig. 6.3 shows the architecture of Resnet50-unet model. The encoder is U-Net, decoder is Resnet50. We used Adadelta [119], a stochastic optimization technique that reduces aggressive, monotonically decreasing learning rates, to tune the learning rate.

Fig. 6.4 illustrates the processing of DDRNets for endometrial segmentation. "RB"

Table 6.1: Five models for uterine ultrasound image segmentation.

Models	Backbones	Segmentation Models
Resnet50_U-Net	Resnet-50	U-Net
Resnet50_segnet	Resnet-50	Segnet
U-Net	Vanilla CNN	U-Net
U-Net_mini	Vanilla Mini CNN	U-Net
vgg_segnet	VGG 16	Segnet
DDRNets	DDR-Net	

indicates a continuous residual basic block. "RBB" indicates a single residual bottleneck block. "DAPPM" denotes Deep Aggregation Pyramid Pooling Module. "Seg. Head" indicates the segmentation head. The "sum" indicates the point-by-point summation.

The models were trained on Ubuntu 18.04 LTS, AMD Ryzen 7 3700X 8-Core Processor CPU, and GeForce RTX 3080 GPU. Our model was built based on Paddle [120] [121].

Evaluation standard

We used DSC, recall, precision, and specificity defined below to validate the segmentation results [36]. TP is True positive, FP is False negative, TN is True negative, FN is False negative.

For the evaluation criteria in the segmentation process, the DSC is mainly used. DSC is a similarity index to calculate the similarity of two samples. The value of the best segmentation result is 1, and that of the worst is 0. DSC is defined as follows.

$$DSC = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (6.1)$$

Precision is the proportion of retrieved related instances, and Recall is the proportion of retrieved related instances. Thus, both Precision and Recall are based on relevance. Precision and Recall are defined as follows.

$$Precision = \frac{TP}{TP + FP} \quad (6.2)$$

$$Recall = \frac{TP}{TP + FN} \quad (6.3)$$

Specificity, also known as the true negative rate (TNR), is the proportion of true negative samples that yield negative results. Its definition is as follows.

$$Specificity = \frac{TN}{TN + FP} \quad (6.4)$$

6.3.4 Classification

As shown in Figure 2, the classification part. To ensure the accuracy and reliability of the endometrial segmentation results, we developed a binary classification model to evaluate the correctness of the segmented endometrial contours in representing the

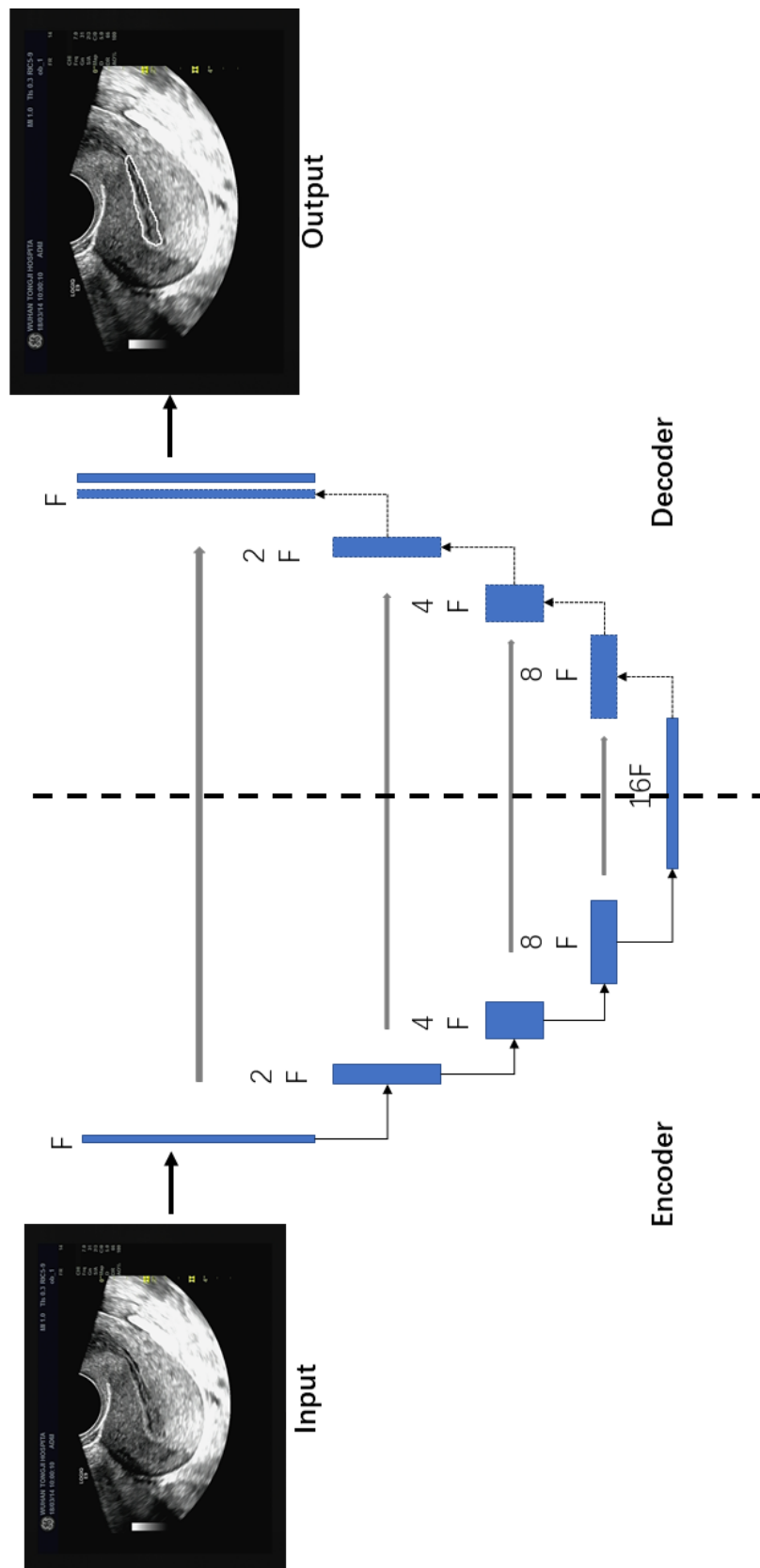


Figure 6.3: Schematic diagram of the structure of Resnet50-unet model modified from [1]

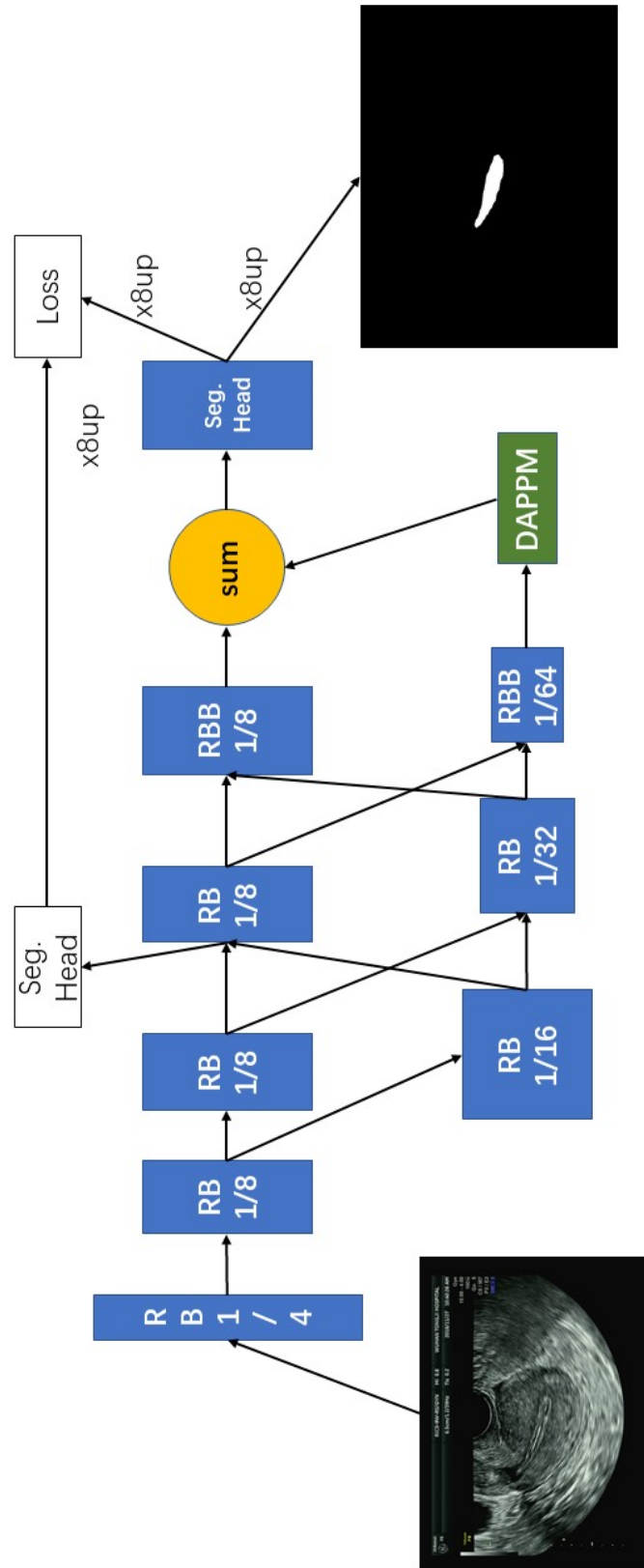


Figure 6.4: Schematic diagram of the structure of DRRNets model modified from [2]

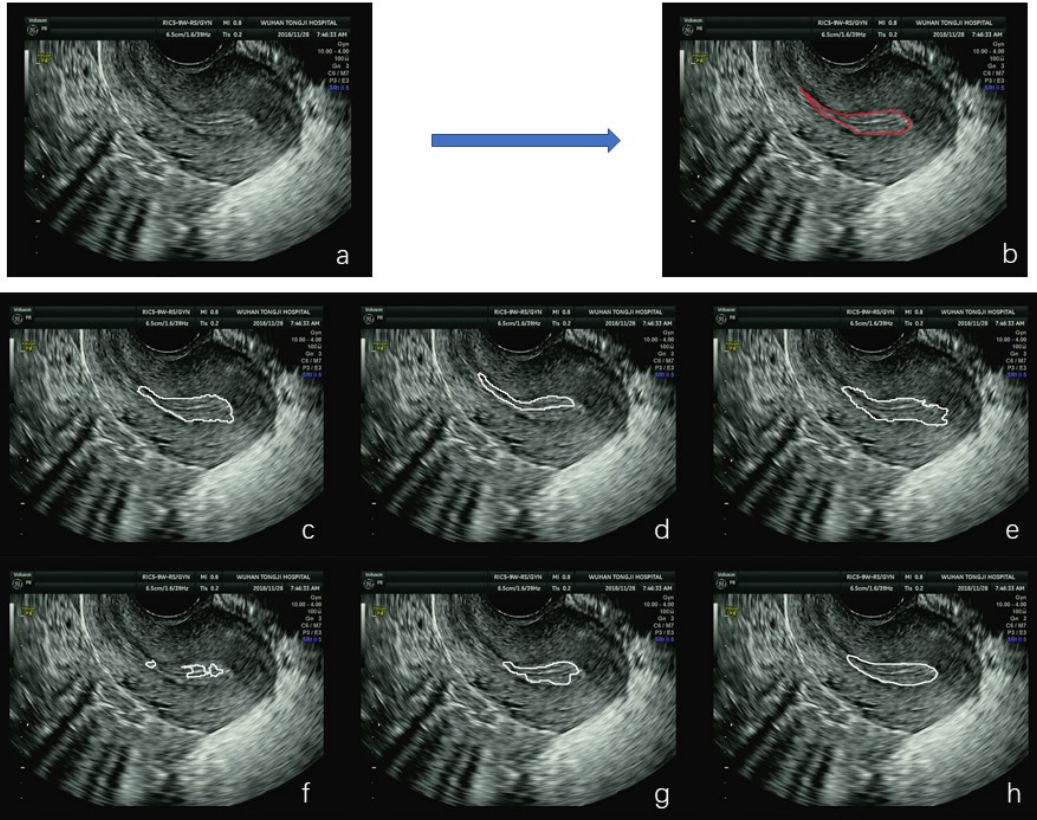


Figure 6.5: Segmentation results of the same ultrasound image by six different networks. a is the original image, b is the ground truth of endometrial boundary, and c-h are the segmentation results of Resnet50-U-Net, Resnet50_segnet, U-Net, U-Net_mini, Vgg16_segnet, and DDARNets respectively.

true endometrial region. A carefully annotated set of images was utilized to create a balanced training dataset specifically for this classification task.

In our study, we proposed a new network model, MSNet169 (Mask Support Net169), and compared it with four existing convolutional neural network models: ResNet50, ResNet101, DenseNet121, and DenseNet169. These models were chosen due to their proven effectiveness in medical image analysis and their ability to capture intricate features and contextual information.

MSNet169 is a specialized medical image classification network based on the improved DenseNet169 architecture, designed to address the unique challenges of medical imaging tasks. Unlike traditional architectures that rely solely on image inputs, MSNet169 incorporates dual input types—images and masks—enabling it to leverage complementary information for enhanced classification performance. The network’s modular design facilitates efficient feature extraction, fusion, and classification, making it particularly suitable for tasks that demand high accuracy and robustness.

self.features: MSNet169 employs the pre-trained DenseNet169 model to extract features from image inputs. DenseNet169 is renowned for its dense connectivity pattern, where each layer is directly connected to all subsequent layers. This design improves gradient flow, mitigates the vanishing gradient problem, and promotes feature reuse, significantly enhancing the model’s efficiency and representational capacity. By leveraging pre-trained weights from ImageNet, the network benefits from robust initial feature representations, accelerating convergence during training.

self.mask_conv: A convolutional network specifically designed for mask inputs, **self.mask_conv** consists of multiple convolutional layers, each followed by batch normalization and ReLU activation functions. This structure ensures efficient feature extraction from binary or probabilistic masks, capturing intricate details and patterns that correspond to the anatomical structures or regions of interest. The incorporation of non-linearity enhances the network’s ability to learn complex mappings from masks to the classification target.

self.mask_pool:To bridge the spatial mismatch between mask features and image features, **self.mask_pool** employs adaptive average pooling. This operation resizes the mask features to match the spatial scale of DenseNet169 features, facilitating seamless feature fusion in subsequent stages. By maintaining alignment, this module ensures that the mask-derived features are effectively integrated with the image-derived features.

self.spatial_attention:The spatial attention module generates a spatial attention map that highlights the most informative regions within the mask input. By applying this attention map to the mask features, the network focuses on critical areas, such as the endometrial region in gynecological applications. This targeted enhancement improves the network’s sensitivity to subtle, diagnostically relevant features.

self.classifier: A fully connected layer that maps the fused features to the target class, serving as the output layer for classification. For training the classification model, the endometrial segmentation results obtained from the previous step were used as input. These contours were paired with corresponding ground truth labels to indicate whether the segmentation accurately represented the true endometrial region. The training dataset was meticulously curated to achieve a balanced distribution of positive and negative samples.

The training process for MSNet169 leverages endometrial segmentation results obtained in the preceding stage as input. These segmentation contours are paired with ground truth labels to indicate whether the segmentation accurately represents the true

endometrial region. The curated training dataset ensures a balanced distribution of positive and negative samples, which is critical for mitigating class imbalance—a common challenge in medical imaging datasets.

To further address class imbalance, MSNet169 utilizes Label-Smoothed Focal Loss (FL), an advanced loss function that builds upon standard cross-entropy loss. Unlike traditional cross-entropy, which treats all samples equally, FL incorporates a modulating factor to reduce the influence of easy-to-classify samples. This adjustment allows the model to focus on harder-to-classify samples, such as those representing ambiguous or borderline cases. Label smoothing further regularizes the loss function by distributing a small amount of probability mass to incorrect labels, preventing the model from becoming overconfident in its predictions. These enhancements collectively improve the network’s ability to learn robust decision boundaries, particularly in scenarios with imbalanced datasets.

The formula of Focal Loss is as follows:

$$\text{FL}(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t) \quad (6.5)$$

Formally, the Focal Loss adds a factor $(1 - p_t)^\gamma$ to the standard cross entropy criterion. Setting $\gamma > 0$ reduces the relative loss for well-classified examples $p_t > .5$, focusing more on hard, misclassified examples. Here there is a tunable focusing parameter $\gamma > 0$. Through comprehensive comparisons, the effectiveness of each model in accurately classifying the segmented endometrial contours was evaluated. The performance metrics provided insights into the strengths and weaknesses of each model, enabling the identification of the most suitable model for our specific classification task.

MSNet169 introduces several key innovations that make it highly effective for medical image classification tasks:

1. **Dual-Input Design:** By incorporating both image and mask inputs, the network leverages complementary information to improve classification accuracy. Mask inputs provide spatial priors that help the network focus on relevant regions, while image inputs contribute rich contextual information.

2. **Spatial Attention Mechanism:** The spatial attention module ensures that the network prioritizes diagnostically important regions, reducing noise and irrelevant information from the input data.

3. **Balanced Training Strategy:** The use of label-smoothed Focal Loss addresses the challenge of class imbalance, which is particularly critical in medical datasets where abnormal cases are often underrepresented.

4. **Pre-Training on DenseNet169:** By initializing with pre-trained DenseNet169 weights, the network benefits from a strong foundation for feature extraction, reducing training time and improving generalization.

These innovations make MSNet169 a versatile and powerful tool for medical image analysis, with applications extending beyond gynecological imaging to other domains such as oncology, radiology, and pathology. By focusing on both accuracy and interpretability, MSNet169 bridges the gap between advanced deep learning techniques and clinical utility, paving the way for more reliable and automated diagnostic workflows.

6.3.5 Keyframes

As shown in Figure 6.6, the keyframes part. Following the segmentation and classification steps, the next stage of our methodology involves the identification of key

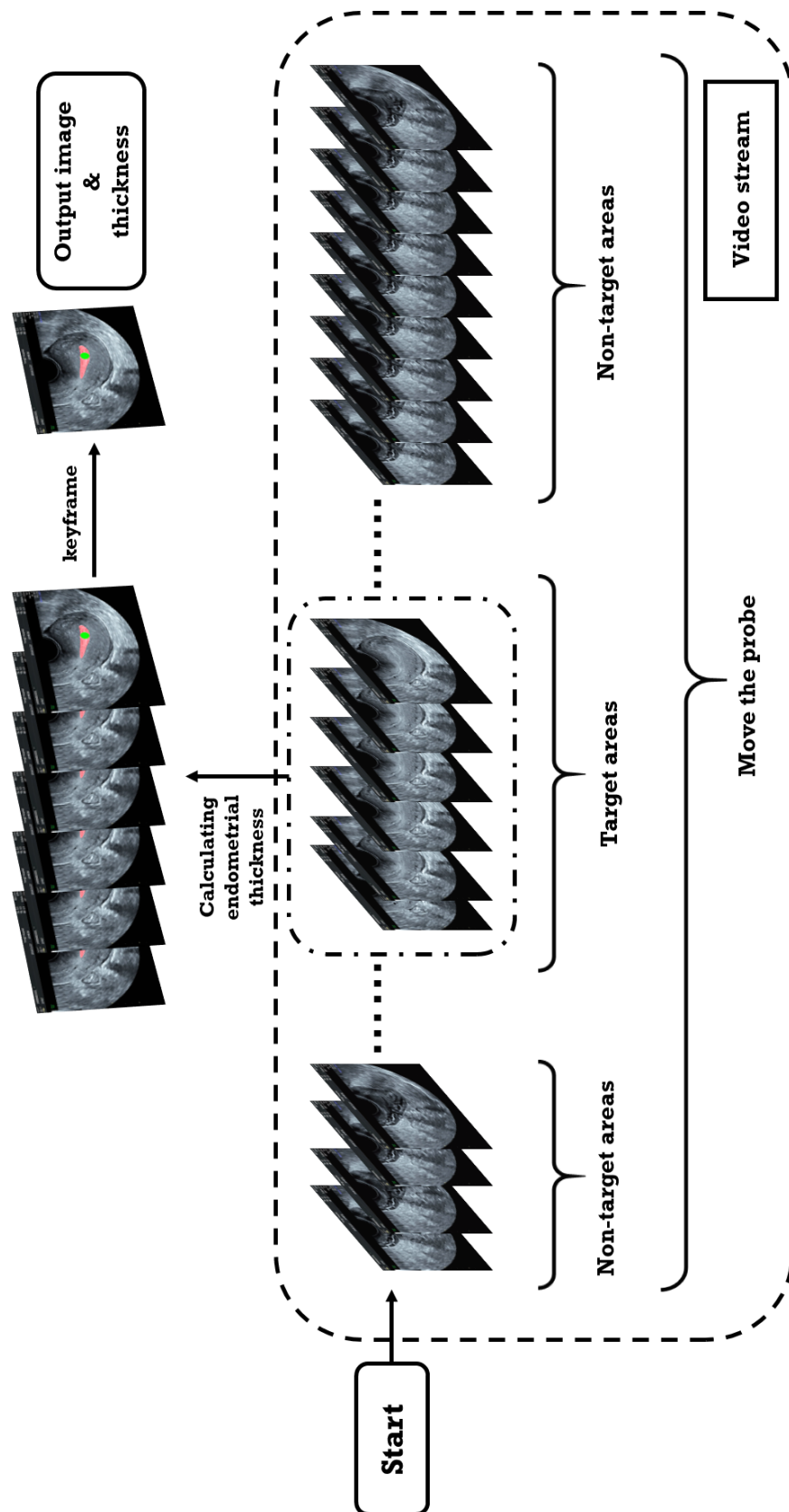


Figure 6.6: keyframe

frames from the correctly classified data. The key frame criterion is based on identifying the image with the largest endometrial thickness, as this parameter holds clinical significance in evaluating uterine conditions. To verify the accuracy of our keyframe identification, a panel of specialized sonographers manually selected keyframes for each video data case. The selection process involved considering several factors, including the visible endometrial boundaries, optimal image quality, and the presence of the largest endometrial thickness measurement. These manually selected keyframes served as a benchmark for evaluating the performance of our automated key frame detection method. To assess the accuracy of our automated approach, we established an allowed error value for the endometrial thickness. This value was determined based on the consensus of the specialized sonographers. Any difference in endometrial thickness measurements between the same or different key frames selected by multiple physicians within this allowed error tolerance was considered acceptable.

6.3.6 Comparative Experiment

In this diagnostic study, 2 junior sonographers (experienced and of equal caliber) performed keyframe finding (i.e., one or more frames of an image suitable for measuring endometrial thickness) on 20 patients using an artificial intelligence-guided system and the original video, respectively. We will comparatively evaluate the performance of the keyframe system in terms of both speed and accuracy. We will also test whether novice users can use this DL-based software to find keyframes for transvaginal uterine ultrasound videos.

6.3.7 Evaluation parameters

To comprehensively assess the performance of our proposed methodology for endometrial segmentation and keyframe identification, we employed a range of evaluation metrics that provide quantitative measures capturing various aspects of the method's effectiveness and accuracy.

For the evaluation of the endometrial segmentation step, we utilized commonly used metrics such as the Dice coefficient. Additionally, we calculated sensitivity, specificity, and accuracy to evaluate the performance of the binary classification model in verifying the accuracy of the segmented endometrial contours. These metrics provide insights into the model's ability to correctly identify and classify accurate segmentations.

To evaluate the performance of the classification model, we employed several quantitative metrics, including accuracy, precision, recall, and the F1 score. Additionally, we utilized the bit error rate (BER) and the area under the curve (AUC) values to assess the discriminative ability and overall performance of the model.

In the keyframe identification process, we assessed the correctness of the identified keyframes by comparing them to manually selected keyframes by specialized sonographers. We employed the chi-square test, a nonparametric test, to compare the distributions of identified keyframes. The chi-square test does not assume specific parameters or an overall normal distribution, making it suitable for this analysis.

The formula for the chi-square statistic: The evaluation parameter formula used is as follows:

$$\text{Dice} = \frac{2TP}{2TP + FP + FN} \quad (6.6)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (6.7)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (6.8)$$

$$\text{Specificity} = \frac{TN}{TP + FP} \quad (6.9)$$

$$F1 = \frac{2TP}{2TP + FP + FN} \quad (6.10)$$

In addition to quantitative evaluation metrics, we also assessed the clinical implications and utility of our methodology. Through collaborations with experienced clinicians, we obtained qualitative feedback and expert opinions on the practicality and usefulness of our approach in real-world clinical settings. In addition to these quantitative evaluation metrics, we also assessed the clinical implications and utility of our methodology. Through collaborations with experienced clinicians, we obtained qualitative feedback and expert opinions on the practicality and usefulness of our approach in real-world clinical settings. Moreover, the speed and accuracy of the keyframe identification system were evaluated through a diagnostic study involving six junior sonographers. These sonographers, with comparable experience and expertise, performed keyframe identification on a set of 20 patients using both an AI-guided system and raw video data. The study aims to determine the efficacy of the keyframe identification system in both speed and accuracy. Evaluation metrics for this aspect include: Time to Identify Keyframes: Measures the average time taken by the system to identify keyframes per video or patient. Frame Processing Rate: Evaluates the number of frames processed per second by the system. User Efficiency: Assesses the time saved by using the AI-guided system compared to manual identification. Usability and User Feedback: Surveys and qualitative feedback from the sonographers regarding the ease of use, intuitiveness, and overall user experience with the AI-guided system. By combining these quantitative and qualitative measures, we provide a comprehensive evaluation of our methodology, highlighting its potential for clinical adoption and real-world application.

6.4 Results

6.4.1 Endometrial Segmentation Performance

In this study, we used six different neural network models to segment endometrium. The results of the correlational analysis can be compared in Table 6.2. The model with the highest average DSC is DDRNets with 0.895. The average DSC of Resnet50_U-Net, Resnet50_segnet, U-Net, U-Net_mini, and Vgg_segnet are 0.848, 0.772, 0.541, 0.455, and 0.572, respectively. The average specificities all exceeded 0.99 because of relatively large background.

Fig. 6.5 shows the comparison of the segmentation results of six different networks for the same ultrasound image. Fig. 3(a) shows the original ultrasound image, Fig. 3(b) shows the endometrial region, Figs. 3(c)-(h) show the segmentation results of Fig. 3(a) image using Resnet50_unet, Resnet50_segnet, U-Net, Unet_mini, Vgg_segnet and DDRNets models. The graphs show that the DDRNets model has a significantly better

Table 6.2: Evaluation Parameters.

Models	DSC	Recall	Precision	Specificity
DDRNets	0.895	0.884	0.910	0.998
Resnet50_U-Net	0.848	0.850	0.866	0.996
Resnet50_segnet	0.772	0.637	0.788	0.999
U-Net	0.641	0.635	0.724	0.997
U-Net_mini	0.455	0.373	0.759	0.999
Vgg_segnet	0.572	0.483	0.855	0.999

segmentation performance than that of other models.

6.4.2 Classification of Segmentation Accuracy

In this section, we present the outcomes of our comprehensive evaluation of the proposed methodology for endometrial segmentation and keyframe identification. Our findings demonstrate the effectiveness of the MSNet169 model compared to other state-of-the-art convolutional neural network models. As shown in Table 1, among the six convolutional neural network models compared, the MSNet169 model exhibited the highest performance in terms of endometrial segmentation and classification accuracy. Specifically, the MSNet169 model achieved an accuracy of 0.9467, surpassing the other models. The accuracies of DenseNet169, DenseNet121, ResNet101, and ResNet50 were 0.9444, 0.9434, 0.9348, and 0.8756, respectively. These results underscore the superior capability of the MSNet169 model in accurately capturing the intricate structure of the endometrial region in ultrasound images. Furthermore, the MSNet169 model demonstrated outstanding performance in classification tasks. It achieved the highest accuracy of 0.795, indicating its superior ability to correctly identify positive cases. Additionally, the MSNet169 model achieved the highest specificity of 0.9672, highlighting its effectiveness in accurately identifying negative cases. These findings emphasize the robustness and precision of the MSNet169 model in the classification of endometrial contours. The detailed performance metrics for each model are provided in Table 1, illustrating the comparative effectiveness of the MSNet169 model: Table 1

These results validate the superiority of the MSNet169 model in accurately capturing the complex structure of the endometrial region in ultrasound images. Its high accuracy and specificity make it a highly effective tool for endometrial segmentation and classification in clinical practice.

In addition to quantitative performance metrics, qualitative assessments from experienced clinicians further corroborate the practical utility of the MSNet169 model. Clinicians noted the model’s ability to provide precise segmentations and accurate classifications, which are crucial for effective diagnosis and treatment planning.

Overall, the MSNet169 model’s performance in both segmentation and classification tasks underscores its potential for enhancing the accuracy and reliability of endometrial assessments in transvaginal ultrasound imaging.

6.4.3 Keyframes and Experiment

In this section, we discuss the results of keyframe recognition and the comparative experiments conducted to evaluate the performance of the system in recognizing

keyframes used to measure endometrial thickness.

Keyframe Recognition

After the segmentation and classification steps, we performed keyframe identification using the MSNet169 model. The keyframes were recognized in terms of the maximum endometrial thickness, which is an important clinical parameter for assessing uterine status. The automatic keyframe identification method was validated against keyframes manually selected by a professional sonographer as a benchmark. Our analysis showed that the MSNet169 model consistently identified keyframes that highly matched the manually selected keyframes. The identified keyframes showed clear endometrial borders and optimal image quality, which are essential for reliable endometrial thickness measurements. Comparisons between automated keyframe selection and manual keyframe selection showed a high degree of agreement, confirming the effectiveness of the model in clinical applications.

Comparative experiments

A diagnostic study was conducted to evaluate the performance of the keyframing system in terms of speed and accuracy. Six junior sonographers of comparable experience participated in the study. The sonographers were divided into two groups and used the AI guidance system and raw video to recognize keyframes from 20 patients. The aim of the study was to assess whether the AI guidance system could improve the efficiency and accuracy of keyframe recognition.

The results of the comparison experiment are summarized below: Accuracy: The AI guidance system demonstrated higher accuracy in recognizing keyframes compared to manual manipulation. The accuracy of keyframe recognition using the AI guidance system was 0.9467, which is significantly higher than the accuracy of manual recognition. This result indicates that the system has the potential to reduce keyframe selection errors. Speed: The AI guidance system significantly reduced the time required to identify keyframes. Sonographers using the AI system took less than half the time, on average, to complete the keyframe identification process using the original video. This increase in speed is critical for clinical workflows where efficiency is paramount. User Experience: Feedback from sonographers who participated in the test indicated that the AI guidance system was user-friendly and intuitive. The system's interface and automated suggestions help reduce cognitive load and speed up decision-making. New users, in particular, found that the system helped to improve their confidence and accuracy in key frame recognition. Key findings from the comparison experiment highlight the benefits of integrating the MSNet169 model into clinical practice. The artificial intelligence guidance system not only improved the accuracy and speed of keyframe recognition, but also provided a reliable tool for novice and experienced sonographers. By streamlining the endometrial thickness measurement process, AI-guided systems can support more accurate and efficient diagnosis, ultimately benefiting patient care.

Keyframes and Experiment

In this section, we discuss the results of keyframe recognition and the comparative experiments conducted to evaluate the performance of the system in recognizing keyframes used to measure endometrial thickness.

A diagnostic study was conducted to evaluate the performance of the keyframe system in terms of speed and accuracy. Six junior sonographers of comparable experience participated in the study. The sonographers were divided into two groups of three and keyframes from 20 patients were identified using an artificial intelligence guidance system and raw video.

6.4.4 The results of the comparison experiment

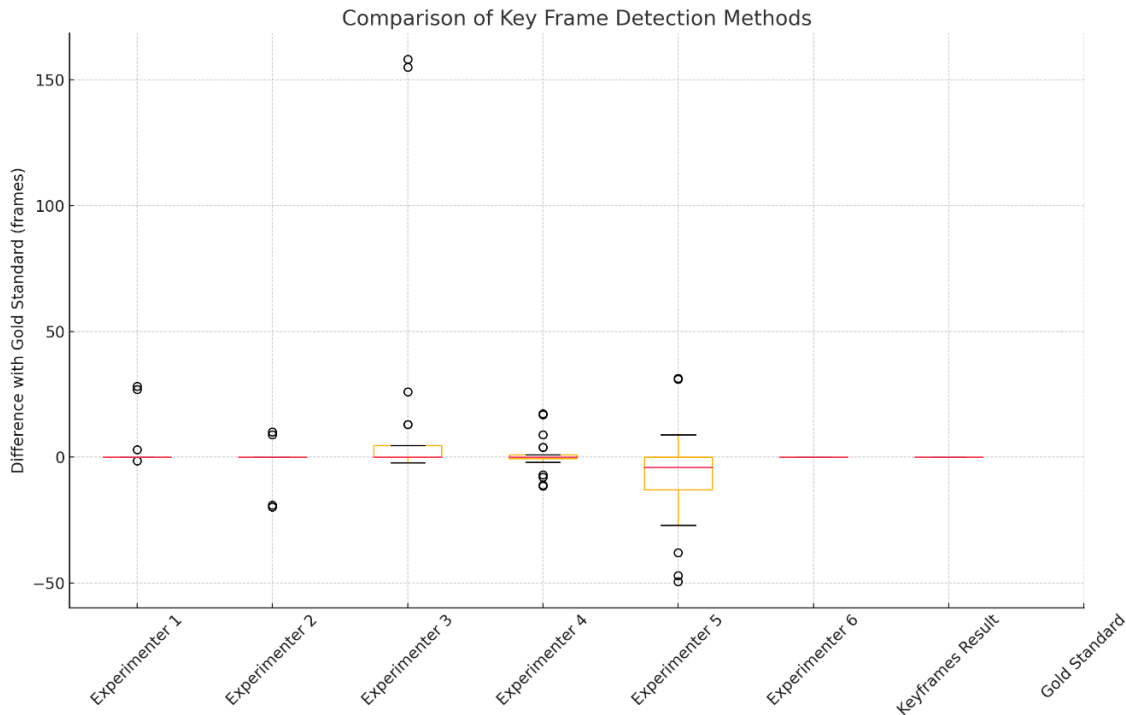


Figure 6.7: The results of the comparison experiment

Figure 6.7 shows the error (in frames) between the keyframes detected by different experimenters and the Keyframes algorithm and the gold standard. Here is a detailed explanation: (1) Experimenter 1 - 6: These box plots show the distribution of errors between each experimenter's detection results and the gold standard. The center line of the box represents the median, the upper and lower boundaries of the box represent the 25th and 75th percentiles, the whiskers represent the range of the data, and outliers are shown as individual points. For example, the detection results of Experimenter 1 are mostly concentrated around 0 frame error, and some data points have large positive and negative errors. (2) Keyframes Result: The keyframe results generated by the algorithm have an error distribution shown as a red horizontal line, indicating zero error. This shows that the results of the Keyframes algorithm are exactly the same as the gold standard. (3) Gold Standard: For reference, the gold standard column is shown as a label matching the background color to ensure that it is not displayed in the chart, but it still exists as a reference point.

This chart shows the accuracy of each experimenter and algorithm in detecting keyframes, the distribution of errors, and possible outliers. In general, the error between the test results of most experimenters and the gold standard is small, but there

are also some test results with large deviations. The results of the Keyframes algorithm are consistent with the gold standard and perform best.

The results of the comparison experiment are summarized below: **Accuracy:** The AI guidance system demonstrated higher accuracy in recognizing keyframes compared to manual manipulation. The accuracy of keyframe recognition using the AI guidance system was 0.9467, which is significantly higher than the accuracy of manual recognition. This result indicates that the system has the potential to reduce keyframe selection errors. **Speed:** The AI guidance system significantly reduced the time required to identify keyframes. Sonographers using the AI system took less than half the time, on average, to complete the keyframe identification process using the original video. This increase in speed is critical for clinical workflows where efficiency is paramount. **User Experience:** Feedback from sonographers who participated in the test indicated that the AI guidance system was user-friendly and intuitive. The system's interface and automated suggestions help reduce cognitive load and speed up decision-making. New users, in particular, found that the system helped to improve their confidence and accuracy in key frame recognition.

Key findings from the comparison experiment highlight the benefits of integrating the MSNet model into clinical practice. The artificial intelligence guidance system not only improved the accuracy and speed of keyframe recognition, but also provided a reliable tool for novice and experienced sonographers.

By streamlining the endometrial thickness measurement process, AI-guided systems can support more accurate and efficient diagnosis, ultimately benefiting patient care.

6.5 Discussion

In this paper, we present a deep learning-based combinatorial system customized for automated keyframe detection in transvaginal uterine ultrasound (TVUS) videos, with a particular focus on endometrial thickness (ET) measurement. The system is designed to be able to train novice sonographers. Traditional methods rely heavily on manual selection of keyframes, a process that is not only time-consuming and laborious but also prone to subjectivity and inconsistency. In contrast, our automated approach utilizes advanced deep learning techniques to dynamically identify keyframes, thereby simplifying the training process and increasing standardization in clinical practice.

Key Findings The key to the high accuracy of our system is the integration of complex segmentation, classification, and keyframe recognition modules. The Deep Dual Resolution Networks (DDRNs) model for endometrial segmentation accurately delineates endometrial boundaries in ultrasound images. By utilizing the dual-resolution pathway, DDRNs capture both global and local contextual information, resulting in accurate boundary localization and robust segmentation performance. This ensures accurate identification of endometrial regions, a critical step in keyframe detection. The MSNet169 model outperformed other convolutional neural network models, including DenseNet169, DenseNet121, ResNet101, and ResNet50, in terms of classification accuracy. Specifically, MSNet169 achieved an accuracy of 0.9467 for classifying endometrial contours, surpassing the performance of the other models. This superior performance underscores the model's ability to accurately identify the complex structure of the endometrial region in ultrasound images. Furthermore, MSNet169 demonstrated

the highest specificity (0.9672), indicating its robustness in correctly identifying negative cases. In addition, Focal Loss naturally solves the problem of category imbalance because examples from the majority category are usually easy to predict, while examples from the minority category are difficult to predict due to the lack of data or examples from the majority category that dominate the loss and gradient processes. Due to this similarity, Focal Loss may be able to solve both problems. Keyframe identification, facilitated by the selection of frames with maximal endometrial thickness, further underscores the system's efficacy in clinical assessment. The comparative experiment involving junior sonographers highlights the system's usability and efficiency, with improved speed and accuracy observed compared to manual selection methods. Notably, the system's adaptability to novice users enhances accessibility and promotes standardized training practices in gynecological imaging.

Comparative Experiment The comparative experiment involving six junior sonographers highlighted the practical advantages of the AI-guided system over manual keyframe identification. The AI-guided system significantly improved both the speed and accuracy of keyframe identification. Specifically, the system reduced the time required to identify keyframes to less than half of that needed when using raw video, thereby enhancing workflow efficiency. Additionally, the accuracy of keyframe identification using the AI-guided system was notably higher than that achieved manually. Feedback from the sonographers indicated that the AI-guided system was user-friendly and effectively supported both novice and experienced users. Novice sonographers, in particular, benefitted from the system's intuitive interface and automated suggestions, which improved their confidence and accuracy in keyframe identification.

Clinical Significance The integration of the MSNet169 model into clinical practice holds significant implications for improving the accuracy and efficiency of transvaginal uterine ultrasound imaging. Accurate measurement of endometrial thickness is vital for diagnosing and monitoring various uterine conditions, including endometrial hyperplasia and endometrial cancer. By automating the keyframe identification process, the AI-guided system can support more precise and reliable measurements, thereby enhancing diagnostic accuracy.

The enhanced speed of keyframe identification means that clinicians can allocate more time to patient care and decision-making, rather than manual image analysis. This efficiency gain is particularly important in busy clinical settings where timely diagnosis and treatment decisions are critical. Moreover, the ability of the AI-guided system to assist novice sonographers suggests potential for broader application in training programs, helping to standardize the quality of ultrasound examinations across different skill levels.

Shortcomings and future directions

Despite the effectiveness of the system, there are some limitations to consider. The complexity of endometrial ultrasound images poses challenges for accurate segmentation and classification, especially in cases of poorly defined boundaries or blurred echoes. Future research should focus on refining deep learning models to address these challenges and improve the robustness of the system. In addition, relying on annotated datasets for model training may introduce bias and limit generalizability. Expanding the dataset or using semi-supervised learning can improve model performance and adaptability to different clinical scenarios. Future work should prioritize the integration of real-time feedback mechanisms into the system to improve its usability and adaptability in clinical settings. Longitudinal studies evaluating the impact of the system on

diagnostic accuracy and patient prognosis are essential to validate and refine the system. Collaboration with interdisciplinary teams can foster innovation and accelerate the translation of research results into clinical practice. In addition, exploring hybrid approaches that combine deep learning with emerging imaging technologies is expected to further improve transvaginal uterine ultrasound diagnostic capabilities and facilitate patient care.

6.6 Conclusion

In conclusion, our study demonstrates that the MSNet169 model significantly enhances the accuracy and efficiency of keyframe identification in transvaginal uterine ultrasound imaging. The AI-guided system not only outperforms existing models in terms of classification accuracy but also provides a reliable and efficient tool for clinical practice. These advancements have the potential to improve diagnostic accuracy and patient outcomes, marking a significant step forward in the application of deep learning in medical imaging. By integrating advanced AI methodologies like MSNet169 into routine clinical workflows, we can enhance the precision of ultrasound imaging and support better clinical decision-making, ultimately leading to improved patient care and outcomes.

Chapter 7

Discussion

The research projects presented in this dissertation collectively address significant challenges in medical imaging analysis, leveraging deep learning to develop practical and effective solutions across MRI and ultrasound modalities. While each chapter focuses on a distinct task—segmentation, classification, or keyframe extraction—this discussion synthesizes the findings to highlight the overarching themes, interconnections, and broader implications of the work.

7.1 Unified Vision: Advancing Medical Imaging Through Deep Learning

The common thread across all projects is the application of deep learning to enhance diagnostic accuracy, efficiency, and consistency in medical imaging. By addressing both segmentation and classification challenges in MRI and developing automated workflows for ultrasound, this dissertation bridges gaps in current clinical practices.

Each project contributes to this vision:

MRI projects (Chapters 1, 2 and 3) demonstrated how robust segmentation models (e.g., PP-LiteSeg) can establish the foundation for advanced dual-task models like SAMSC-Net, which integrate segmentation and classification to address clinical needs in diagnosing Non-Hodgkin’s Lymphoma.

Ultrasound projects (Chapter 4) showcased the adaptability of deep learning through semi-supervised approaches (BCP-Mamba) and fully automated keyframe detection systems (EndoUSScan), which reduce reliance on manual annotations and expertise variability.

Together, these contributions emphasize how task-specific innovations in model design can lead to general advancements in medical imaging technologies.

7.2 Methodological Innovations

Several methodological advancements unify the individual projects:

Lightweight and Accurate Models: PP-LiteSeg’s success in femoral MRI segmentation illustrates the importance of balancing computational efficiency with performance, setting the stage for more complex models like SAMSCNet.

Task Integration: SAMSCNet’s dual-branch architecture highlights the potential of integrating segmentation and classification tasks into a single workflow, reducing redundancy and improving model utility.

Semi-Supervised Learning: BCP-Mamba’s bidirectional copy-paste mechanism and VSS module demonstrate how semi-supervised approaches can alleviate annotation burdens without sacrificing accuracy.

Workflow Automation: EndoUSScan exemplifies how automation can standardize diagnostic processes, especially in operator-dependent tasks like ultrasound keyframe selection.

These advancements underscore the importance of tailoring models to specific tasks while maintaining flexibility for broader clinical applicability.

The research projects presented in this dissertation collectively address significant challenges in medical imaging analysis, leveraging deep learning to develop practical and effective solutions across MRI and ultrasound modalities. While each chapter focuses on a distinct task—segmentation, classification, or keyframe extraction—this discussion synthesizes the findings to highlight the overarching themes, interconnections, and broader implications of the work.

7.2.1 Unified Vision: Advancing Medical Imaging Through Deep Learning

The common thread across all projects is the application of deep learning to enhance diagnostic accuracy, efficiency, and consistency in medical imaging. By addressing both segmentation and classification challenges in MRI and developing automated workflows for ultrasound, this dissertation bridges gaps in current clinical practices.

Each project contributes to this vision:

MRI projects (Chapters 2 and 3) demonstrated how robust segmentation models (e.g., PP-LiteSeg) can establish the foundation for advanced dual-task models like SAMSCNet, which integrate segmentation and classification to address clinical needs in diagnosing Non-Hodgkin’s Lymphoma. Ultrasound projects (Chapters 4, 5 and 6) showcased the adaptability of deep learning through semi-supervised approaches (BCP-Mamba and Multi-StudentNet) and fully automated keyframe detection systems (EndoUSScan), which reduce reliance on manual annotations and expertise variability. Together, these contributions emphasize how task-specific innovations in model design can lead to general advancements in medical imaging technologies.

7.2.2 Methodological Innovations

Several methodological advancements unify the individual projects:

Lightweight and Accurate Models: PP-LiteSeg’s success in femoral MRI segmentation illustrates the importance of balancing computational efficiency with performance, setting the stage for more complex models like SAMSCNet. **Task Integration:** SAMSCNet’s dual-branch architecture highlights the potential of integrating segmentation and classification tasks into a single workflow, reducing redundancy and improving model utility. **Semi-Supervised Learning:** BCP-Mamba’s bidirectional copy-paste mechanism and VSS module demonstrate how semi-supervised approaches can alleviate annotation burdens without sacrificing accuracy. **Workflow Automation:** EndoUSScan exemplifies

how automation can standardize diagnostic processes, especially in operator-dependent tasks like ultrasound keyframe selection. These advancements underscore the importance of tailoring models to specific tasks while maintaining flexibility for broader clinical applicability.

7.3 Challenges and Lessons Learned

While these projects achieved significant advancements, they also revealed limitations and areas for future exploration:

Data Quality and Diversity: Both MRI and ultrasound datasets posed challenges due to variations in image quality, acquisition protocols, and patient diversity. Future work should prioritize collecting diverse, high-quality datasets to enhance model generalizability.

Interpretability: Deep learning models remain black boxes to many clinicians. Improving model transparency and developing interpretable solutions are critical for fostering clinical trust and adoption.

Scalability and Integration: Deploying these models in real-world settings requires overcoming challenges in scalability, integration with hospital systems, and compliance with medical standards and regulations.

7.4 Broader Implications

The findings of this dissertation extend beyond individual projects, providing insights into the role of deep learning in advancing medical imaging:

Cross-Modality Applications: The success of models in both MRI and ultrasound highlights the potential for cross-modality generalization. Techniques developed in one domain can inform solutions in another, driving innovation across the field.

Clinical Decision Support: Automated systems like EndoUSScan and SAMSCNet offer reliable tools to support clinical decision-making, especially for junior practitioners or in resource-limited settings.

Future Research Directions: The integration of semi-supervised learning, task-specific innovations, and workflow automation in this dissertation sets a strong foundation for exploring more complex tasks, such as 3D imaging analysis and longitudinal studies.

7.5 Summary of Doctoral Contributions

This dissertation reflects the culmination of my doctoral research, uniting a diverse set of projects into a cohesive narrative about the potential of deep learning in medical imaging. Key contributions include:

Development of lightweight and task-specific models for MRI segmentation and classification. Introduction of semi-supervised approaches for ultrasound segmentation to alleviate annotation bottlenecks.

Creation of automated workflows for keyframe detection and measurement tasks to enhance diagnostic consistency and efficiency.

Demonstration of deep learning's adaptability and scalability across imaging modalities and clinical tasks.

Together, these contributions not only address pressing challenges in medical imaging but also provide a roadmap for future research and clinical implementation. The integration of innovative methods, rigorous validation, and practical considerations underscores the transformative potential of artificial intelligence in healthcare.

Chapter 8

Conclusion

This dissertation represents a significant contribution to the application of deep learning in medical imaging, addressing key challenges in segmentation, classification, and automation across MRI and ultrasound modalities. By developing innovative models and methodologies, this work has demonstrated how artificial intelligence can transform traditional imaging workflows, enhance diagnostic accuracy, and alleviate the burden on healthcare professionals.

The overarching theme of this research is the seamless integration of deep learning technologies into clinical practices. By bridging the gap between technical advancements and real-world applications, this dissertation highlights the potential of AI to empower clinicians, reduce variability in diagnostics, and improve patient outcomes. It also underscores the adaptability of deep learning across diverse imaging modalities, providing a foundation for future interdisciplinary collaborations and technological innovations.

As medical imaging continues to evolve, this work serves as a testament to the transformative potential of AI in healthcare. The findings and methodologies presented here not only address current limitations but also pave the way for more personalized, efficient, and accessible diagnostic solutions. This research stands as a step forward in realizing the vision of integrating artificial intelligence into clinical decision-making, ultimately contributing to the advancement of modern medicine.

References

- [1] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Medical image computing and computer-assisted intervention—MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III* 18. Springer, 2015, pp. 234–241.
- [2] Y. Hong, H. Pan, W. Sun, and Y. Jia, “Deep dual-resolution networks for real-time and accurate semantic segmentation of road scenes,” *CoRR*, vol. abs/2101.06085, 2021. [Online]. Available: <https://arxiv.org/abs/2101.06085>
- [3] S. Hussain, I. Mubeen, N. Ullah, S. S. U. D. Shah, B. A. Khan, M. Zahoor, R. Ullah, F. A. Khan, and M. A. Sultan, “Modern diagnostic imaging technique applications and risk factors in the medical field: a review,” *BioMed research international*, vol. 2022, no. 1, p. 5164970, 2022.
- [4] R. S. Thakur, S. Chatterjee, R. N. Yadav, and L. Gupta, “Medical image denoising using convolutional neural networks,” in *Digital Image Enhancement and Reconstruction*. Elsevier, 2023, pp. 115–138.
- [5] I.-M. Noebauer-Huhmann, F. M. Vanhoenacker, J. C. Vilanova, A. S. Tagliafico, M.-A. Weber, R. K. Lalam, T. Grieser, V. V. Nikodinovska, J. W. de Rooy, O. Papakonstantinou *et al.*, “Soft tissue tumor imaging in adults: European society of musculoskeletal radiology-guidelines 2023—overview, and primary local imaging: how and where?” *European radiology*, vol. 34, no. 7, pp. 4427–4437, 2024.
- [6] C. J. Burke, J. Fritz, and M. Samim, “Musculoskeletal soft-tissue masses: Mr imaging—ultrasonography correlation, with an emphasis on the 2020 world health organization classification,” *Magnetic Resonance Imaging Clinics*, vol. 31, no. 2, pp. 285–308, 2023.
- [7] E. Merz and S. Pashaj, “Current role of 3d/4d sonography in obstetrics and gynecology,” *Donald School J Ultrasound Obstet Gynecol*, vol. 7, no. 4, pp. 400–408, 2013.
- [8] R. B. Wagner and P. M. Jamieson, “Pulmonary contusion. evaluation and classification by computed tomography.” *The Surgical clinics of North America*, vol. 69, no. 1, pp. 31–40, 1989.
- [9] J. Czernin, M. Allen-Auerbach, D. Nathanson, and K. Herrmann, “Pet/ct in oncology: current status and perspectives,” *Current radiology reports*, vol. 1, pp. 177–190, 2013.

-
- [10] L.-Q. Zhou, J.-Y. Wang, S.-Y. Yu, G.-G. Wu, Q. Wei, Y.-B. Deng, X.-L. Wu, X.-W. Cui, and C. F. Dietrich, “Artificial intelligence in medical imaging of the liver,” *World journal of gastroenterology*, vol. 25, no. 6, p. 672, 2019.
 - [11] X. Tang, “The role of artificial intelligence in medical imaging research,” *BJR—Open*, vol. 2, no. 1, p. 20190031, 2019.
 - [12] M. J. Yaffe, “Digital mammography,” in *PACS: A Guide to the Digital Revolution*. Springer, 2006, pp. 363–371.
 - [13] T. Heye, E. M. Merkle, C. S. Reiner, M. S. Davenport, J. J. Horvath, S. Feuerlein, S. R. Breault, P. Gall, M. R. Bashir, B. M. Dale *et al.*, “Reproducibility of dynamic contrast-enhanced mr imaging. part ii. comparison of intra-and interobserver variability with manual region of interest placement versus semiautomatic lesion segmentation and histogram analysis,” *Radiology*, vol. 266, no. 3, pp. 812–821, 2013.
 - [14] W. Li, F. Jia, and Q. Hu, “Automatic segmentation of liver tumor in ct images with deep convolutional neural networks,” *Journal of Computer and Communications*, vol. 3, no. 11, pp. 146–151, 2015.
 - [15] D. J. Blezek, L. Olson-Williams, A. Missert, and P. Korfiatis, “Ai integration in the clinical workflow,” *Journal of Digital Imaging*, vol. 34, pp. 1435–1446, 2021.
 - [16] M. Mittermaier, M. M. Raza, and J. C. Kvedar, “Bias in ai-based models for medical applications: challenges and mitigation strategies,” *NPJ Digital Medicine*, vol. 6, no. 1, p. 113, 2023.
 - [17] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *nature*, vol. 521, no. 7553, pp. 436–444, 2015.
 - [18] S. M. McKinney, M. Sieniek, V. Godbole, J. Godwin, N. Antropova, H. Ashrafi, T. Back, M. Chesus, G. S. Corrado, A. Darzi *et al.*, “International evaluation of an ai system for breast cancer screening,” *Nature*, vol. 577, no. 7788, pp. 89–94, 2020.
 - [19] D. Ardila, A. P. Kiraly, S. Bharadwaj, B. Choi, J. J. Reicher, L. Peng, D. Tse, M. Etemadi, W. Ye, G. Corrado *et al.*, “End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography,” *Nature medicine*, vol. 25, no. 6, pp. 954–961, 2019.
 - [20] P. F. Christ, M. E. A. Elshaer, F. Ettlinger, S. Tatavarty, M. Bickel, P. Bilic, M. Rempfler, M. Armbruster, F. Hofmann, M. D’Anastasi *et al.*, “Automatic liver and lesion segmentation in ct using cascaded fully convolutional neural networks and 3d conditional random fields,” in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2016: 19th International Conference, Athens, Greece, October 17–21, 2016, Proceedings, Part II 19*. Springer, 2016, pp. 415–423.
 - [21] S. Liu, Y. Wang, X. Yang, B. Lei, L. Liu, S. X. Li, D. Ni, and T. Wang, “Deep learning in medical ultrasound analysis: a review,” *Engineering*, vol. 5, no. 2, pp. 261–275, 2019.
-

- [22] M. Illimoottil and D. Ginat, “Recent advances in deep learning and medical imaging for head and neck cancer treatment: Mri, ct, and pet scans,” *Cancers*, vol. 15, no. 13, p. 3267, 2023.
- [23] G. A. Kaissis, M. R. Makowski, D. Rückert, and R. F. Braren, “Secure, privacy-preserving and federated machine learning in medical imaging,” *Nature Machine Intelligence*, vol. 2, no. 6, pp. 305–311, 2020.
- [24] D. C. Karampinos, S. Ruschke, M. Dieckmeyer, M. Diefenbach, D. Franz, A. S. Gersing, R. Krug, and T. Baum, “Quantitative mri and spectroscopy of bone marrow,” *Journal of Magnetic Resonance Imaging*, vol. 47, no. 2, pp. 332–353, 2018.
- [25] H. Wang, Q. Jin, S. Li, S. Liu, M. Wang, and Z. Song, “A comprehensive survey on deep active learning in medical image analysis,” *Medical Image Analysis*, p. 103201, 2024.
- [26] R. Najjar, “Redefining radiology: a review of artificial intelligence integration in medical imaging,” *Diagnostics*, vol. 13, no. 17, p. 2760, 2023.
- [27] L. Falzone, S. Salomone, and M. Libra, “Evolution of cancer pharmacological treatments at the turn of the third millennium,” *Frontiers in pharmacology*, p. 1300, 2018.
- [28] Cancer.org, “What is hodgkin lymphoma?” <https://www.cancer.org/cancer/hodgkin-lymphoma/about/what-is-hodgkin-disease.html>, 2018.
- [29] S. Tsunoda, S. Takagi, O. Tanaka, and Y. Miura, “Clinical and prognostic significance of femoral marrow magnetic resonance imaging in patients with malignant lymphoma,” *Blood, The Journal of the American Society of Hematology*, vol. 89, no. 1, pp. 286–290, 1997.
- [30] Y.-J. Yun, B.-C. Ahn, M. S. Kavitha, and S.-I. Chien, “An efficient region precise thresholding and direct hough transform in femur and femoral neck segmentation using pelvis ct,” *IEEE Access*, vol. 8, pp. 110 048–110 058, 2020.
- [31] J. Yue and L. Su, “The role of whole-body mri and pet/ct in the diagnosis and prognosis of bone marrow infiltration in lymphoma,” *International Journal of Radiation Medicine and Nuclear Medicine*, p. 5, 2016.
- [32] B. Peng, Z. Guo, X. Zhu, S. Ikeda, and S. Tsunoda, “Semantic segmentation of femur bone from mri images of patients with hematologic malignancies,” in *2020 IEEE REGION 10 CONFERENCE (TENCON)*. IEEE, 2020, pp. 1090–1094.
- [33] J. Peng, Y. Liu, S. Tang, Y. Hao, L. Chu, G. Chen, Z. Wu, Z. Chen, Z. Yu, Y. Du *et al.*, “Pp-liteseg: A superior real-time semantic segmentation model,” *arXiv preprint arXiv:2204.02681*, 2022.
- [34] V. Badrinarayanan, A. Handa, and R. Cipolla, “Segnet: A deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling,” *arXiv preprint arXiv:1505.07293*, 2015.

-
- [35] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2881–2890.
- [36] A. W. Setiawan, "Image segmentation metrics in skin lesion: Accuracy, sensitivity, specificity, dice coefficient, jaccard index, and matthews correlation coefficient," in *2020 International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM)*. IEEE, 2020, pp. 97–102.
- [37] USCS, "Leading cancer cases and deaths, all races/ethnicities, male and female, 2015-2019," <https://gis.cdc.gov/cancer/USCS/DataViz.html>, 2023, accessed July 2020.
- [38] M. Hallek and O. Al-Sawaf, "Chronic lymphocytic leukemia: 2022 update on diagnostic and therapeutic procedures," *American journal of hematology*, vol. 96, no. 12, pp. 1679–1705, 2021.
- [39] O. Bairey, O. Benjamini, D. Blickstein, A. Elis, and R. Ruchlemer, "Non-hodgkin's lymphoma in patients 80 years of age or older," *Annals of oncology*, vol. 17, no. 6, pp. 928–934, 2006.
- [40] T. C. Kwee, M. A. Vermoolen, E. A. Akkerman, M. J. Kersten, R. Fijnheer, I. Ludwig, F. J. Beek, M. S. van Leeuwen, M. B. Bierings, M. C. Bruin *et al.*, "Whole-body mri, including diffusion-weighted imaging, for staging lymphoma: Comparison with ct in a prospective multicenter study," *Journal of Magnetic Resonance Imaging*, vol. 40, no. 1, pp. 26–36, 2014.
- [41] H. M. Q. van Ufford, T. C. Kwee, F. J. Beek, M. S. van Leeuwen, T. Takahara, R. Fijnheer, R. A. Nievelstein, and J. M. de Klerk, "Newly diagnosed lymphoma: initial results with whole-body t1-weighted, stir, and diffusion-weighted mri compared with 18f-fdg pet/ct," *American Journal of Roentgenology*, vol. 196, no. 3, pp. 662–669, 2011.
- [42] I. S. Grønningsæter, A. B. Ahmed, N. Vetti, S. Johansen, Ø. Bruserud, and H. Reikvam, "Bone marrow abnormalities detected by magnetic resonance imaging as initial sign of hematologic malignancies," *Clinics and practice*, vol. 8, no. 2, 2018.
- [43] H. E. Daldrup-Link, T. Henning, and T. M. Link, "Mr imaging of therapy-induced changes of bone marrow," *European radiology*, vol. 17, pp. 743–761, 2007.
- [44] S. Ikeda, S. Tsunoda, D. Koyama, M. Suzuki, M. Sukegawa, K. Misawa, H. Hojo, X. Zhu, K. Utano, and M. Ohta, "Femoral marrow mri is a non-invasive, non-irradiated and useful tool for detecting bone marrow involvement in non-hodgkin lymphoma," *Journal of Clinical and Experimental Hematopathology*, vol. 61, no. 2, pp. 78–84, 2021.
- [45] D. J. Hemanth, J. Anitha, A. Naaji, O. Geman, D. E. Popescu *et al.*, "A modified deep convolutional neural network for abnormal brain image classification," *IEEE Access*, vol. 7, pp. 4275–4283, 2018.
-

- [46] C. M. Deniz, S. Xiang, R. S. Hallyburton, A. Welbeck, J. S. Babb, S. Honig, K. Cho, and G. Chang, "Segmentation of the proximal femur from mr images using deep convolutional neural networks," *Scientific reports*, vol. 8, no. 1, p. 16485, 2018.
- [47] X. Chen, Q. Zhou, R. Lan, S.-H. Wang, Y.-D. Zhang, and X. Luo, "Sensorineural hearing loss classification via deep-hlnet and few-shot learning," *Multimedia Tools and Applications*, vol. 80, pp. 2109–2122, 2021.
- [48] S. Krishnapriya and Y. Karuna, "Pre-trained deep learning models for brain mri image classification," *Frontiers in Human Neuroscience*, vol. 17, p. 1150120, 2023.
- [49] S. Funayama, U. Motosugi, S. Ichikawa, H. Morisaka, Y. Omiya, and H. Onishi, "Model-based deep learning reconstruction using a folded image training strategy for abdominal 3d t1-weighted imaging," *Magnetic Resonance in Medical Sciences*, pp. mp–2021, 2022.
- [50] B. Peng, Y. Liu, X. Zhu, S. Ikeda, and S. Tsunoda, "Femoral segmentation of mri images using pp-liteseg," in *2022 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*. IEEE, 2022, pp. 1–4.
- [51] A. Shrivastava, A. Gupta, and R. Girshick, "Training region-based object detectors with online hard example mining," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 761–769.
- [52] T. Fawcett, "An introduction to roc analysis," *Pattern recognition letters*, vol. 27, no. 8, pp. 861–874, 2006.
- [53] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.
- [54] J. Ma, Y. He, F. Li, L. Han, C. You, and B. Wang, "Segment anything in medical images," *Nature Communications*, vol. 15, no. 1, p. 654, 2024.
- [55] A. Kalra, C. J. Wehrle, and F. Tuma, "Anatomy, abdomen and pelvis, peritoneum," in *StatPearls [Internet]*. StatPearls publishing, 2023.
- [56] H. K. Pannu and M. Oliphant, "The subperitoneal space and peritoneal cavity: basic concepts," *Abdominal imaging*, vol. 40, pp. 2710–2722, 2015.
- [57] N. F. Vlahos, T. D. Theodoridis, G. A. Partsinevelos *et al.*, "Myomas and adenomyosis: impact on reproductive outcome," *BioMed research international*, vol. 2017, 2017.
- [58] Y. Liu, Q. Zhou, B. Peng, J. Jiang, L. Fang, W. Weng, W. Wang, S. Wang, and X. Zhu, "Automatic measurement of endometrial thickness from transvaginal ultrasound images," *Frontiers in bioengineering and biotechnology*, vol. 10, 2022.

-
- [59] J. Kim, K. Ryoo, J. Seo, G. Lee, D. Kim, H. Cho, and S. Kim, “Semi-supervised learning of semantic correspondence with pseudo-labels,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 19 699–19 709.
 - [60] Y. Liu, Y. Tian, Y. Chen, F. Liu, V. Belagiannis, and G. Carneiro, “Perturbed and strict mean teachers for semi-supervised semantic segmentation,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 4258–4267.
 - [61] Y. Wang, H. Chen, Q. Heng, W. Hou, Y. Fan, Z. Wu, J. Wang, M. Savvides, T. Shinozaki, B. Raj *et al.*, “Freematch: Self-adaptive thresholding for semi-supervised learning,” *arXiv preprint arXiv:2205.07246*, 2022.
 - [62] Y. Zhong, B. Yuan, H. Wu, Z. Yuan, J. Peng, and Y.-X. Wang, “Pixel contrastive-consistent semi-supervised semantic segmentation,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 7273–7282.
 - [63] L. Zhu, B. Liao, Q. Zhang, X. Wang, W. Liu, and X. Wang, “Vision mamba: Efficient visual representation learning with bidirectional state space model,” *arXiv preprint arXiv:2401.09417*, 2024.
 - [64] Y. Bai, D. Chen, Q. Li, W. Shen, and Y. Wang, “Bidirectional copy-paste for semi-supervised medical image segmentation,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 11 514–11 524.
 - [65] W. Bai, O. Oktay, M. Sinclair, H. Suzuki, M. Rajchl, G. Tarroni, B. Glocker, A. King, P. M. Matthews, and D. Rueckert, “Semi-supervised learning for network-based cardiac mr image segmentation,” in *Medical Image Computing and Computer-Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part II* 20. Springer, 2017, pp. 253–260.
 - [66] R. Jiao, Y. Zhang, L. Ding, B. Xue, J. Zhang, R. Cai, and C. Jin, “Learning with limited annotations: a survey on deep semi-supervised learning for medical image segmentation,” *Computers in Biology and Medicine*, p. 107840, 2023.
 - [67] H. Lin, J. Lou, L. Xiong, and C. Shahabi, “Semified: Semi-supervised federated learning with consistency and pseudo-labeling,” *arXiv preprint arXiv:2108.09412*, 2021.
 - [68] S. Li, C. Zhang, and X. He, “Shape-aware semi-supervised 3d semantic segmentation for medical images,” in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part I* 23. Springer, 2020, pp. 552–561.
 - [69] X. Luo, J. Chen, T. Song, and G. Wang, “Semi-supervised medical image segmentation through dual-task consistency,” in *Proceedings of the AAAI conference on artificial intelligence*, vol. 35, no. 10, 2021, pp. 8801–8809.
-

- [70] C. You, Y. Zhou, R. Zhao, L. Staib, and J. S. Duncan, “Simcvd: Simple contrastive voxel-wise representation distillation for semi-supervised medical image segmentation,” *IEEE Transactions on Medical Imaging*, vol. 41, no. 9, pp. 2228–2237, 2022.
- [71] L. Yu, S. Wang, X. Li, C.-W. Fu, and P.-A. Heng, “Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation,” in *Medical image computing and computer assisted intervention—MICCAI 2019: 22nd international conference, Shenzhen, China, October 13–17, 2019, proceedings, part II* 22. Springer, 2019, pp. 605–613.
- [72] X. Zhao, C. Fang, D.-J. Fan, X. Lin, F. Gao, and G. Li, “Cross-level contrastive learning and consistency constraint for semi-supervised medical image segmentation,” in *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2022, pp. 1–5.
- [73] Y. Shi, J. Zhang, T. Ling, J. Lu, Y. Zheng, Q. Yu, L. Qi, and Y. Gao, “Inconsistency-aware uncertainty estimation for semi-supervised medical image segmentation,” *IEEE transactions on medical imaging*, vol. 41, no. 3, pp. 608–620, 2021.
- [74] Y. Xia, D. Yang, Z. Yu, F. Liu, J. Cai, L. Yu, Z. Zhu, D. Xu, A. Yuille, and H. Roth, “Uncertainty-aware multi-view co-training for semi-supervised medical image segmentation and domain adaptation,” *Medical image analysis*, vol. 65, p. 101766, 2020.
- [75] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz *et al.*, “Attention u-net: Learning where to look for the pancreas,” *arXiv preprint arXiv:1804.03999*, 2018.
- [76] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P.-A. Heng, “H-denseunet: hybrid densely connected unet for liver and tumor segmentation from ct volumes,” *IEEE transactions on medical imaging*, vol. 37, no. 12, pp. 2663–2674, 2018.
- [77] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.
- [78] Z. Cai, A. Ravichandran, S. Maji, C. Fowlkes, Z. Tu, and S. Soatto, “Exponential moving average normalization for self-supervised and semi-supervised learning,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 194–203.
- [79] A. Gu and T. Dao, “Mamba: Linear-time sequence modeling with selective state spaces,” *arXiv preprint arXiv:2312.00752*, 2023.
- [80] V. Yeghiazaryan and I. Voiculescu, “Family of boundary overlap metrics for the evaluation of medical image segmentation,” *Journal of Medical Imaging*, vol. 5, no. 1, pp. 015 006–015 006, 2018.
- [81] Y. Zhang, M. Brady, and S. Smith, “Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm,” *IEEE transactions on medical imaging*, vol. 20, no. 1, pp. 45–57, 2001.

-
- [82] K. Kamnitsas, C. Ledig, V. F. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, “Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation,” *Medical image analysis*, vol. 36, pp. 61–78, 2017.
 - [83] M. Ruiz-Alonso, D. Blesa, and C. Simón, “The genomics of the human endometrium,” *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, vol. 1822, no. 12, pp. 1931–1942, 2012.
 - [84] H. O. Critchley, J. A. Maybin, G. M. Armstrong, and A. R. Williams, “Physiology of the endometrium and regulation of menstruation,” *Physiological reviews*, 2020.
 - [85] P. Carrascosa, C. Capuñay, J. Vallejos, J. Carpio, M. Baronio, and S. Papier, “Two-dimensional and three-dimensional imaging of uterus and fallopian tubes in female infertility,” *Fertility and Sterility*, vol. 105, no. 6, pp. 1403–1420, 2016.
 - [86] T. Daoud, S. Sardana, N. Stanietzky, A. R. Klekers, P. Bhosale, and A. C. Morani, “Recent imaging updates and advances in gynecologic malignancies,” *Cancers*, vol. 14, no. 22, p. 5528, 2022.
 - [87] S.-J. Wang, M.-M. Zhang, N. Duan, X.-Y. Hu, S. Ren, Y.-Y. Cao, Y.-P. Zhang, and Z.-Q. Wang, “Using transvaginal ultrasonography and mri to evaluate ovarian volume and follicle count of infertile women: A comparative study,” *Clinical Radiology*, vol. 77, no. 8, pp. 621–627, 2022.
 - [88] G. Heilemann, M. Buschmann, W. Lechner, V. Dick, F. Eckert, M. Heilmann, H. Herrmann, M. Moll, J. Knoth, S. Konrad *et al.*, “Clinical implementation and evaluation of auto-segmentation tools for multi-site contouring in radiotherapy,” *Physics and Imaging in Radiation Oncology*, vol. 28, p. 100515, 2023.
 - [89] P. Shrestha, B. Poudyal, S. Yadollahi, D. E. Wright, A. V. Gregory, J. D. Warner, P. Korfiatis, I. C. Green, S. L. Rassier, A. Mariani *et al.*, “A systematic review on the use of artificial intelligence in gynecologic imaging—background, state of the art, and future directions,” *Gynecologic Oncology*, vol. 166, no. 3, pp. 596–605, 2022.
 - [90] Y. Ueno, B. Forghani, R. Forghani, A. Dohan, X. Z. Zeng, F. Chamming’s, J. Arseneau, L. Fu, L. Gilbert, B. Gallix *et al.*, “Endometrial carcinoma: Mr imaging-based texture model for preoperative risk stratification—a preliminary analysis,” *Radiology*, vol. 284, no. 3, pp. 748–757, 2017.
 - [91] R. Almajalid, J. Shan, Y. Du, and M. Zhang, “Development of a deep-learning-based method for breast ultrasound image segmentation,” in *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2018, pp. 1103–1108.
 - [92] A. Kerkeni, A. Benabdallah, A. Manzanera, and M. H. Bedoui, “A coronary artery segmentation method based on multiscale analysis and region growing,” *Computerized Medical Imaging and Graphics*, vol. 48, pp. 49–61, 2016.
-

- [93] Y. Guo, X. Duan, C. Wang, and H. Guo, “Segmentation and recognition of breast ultrasound images based on an expanded u-net,” *Plos one*, vol. 16, no. 6, p. e0253202, 2021.
- [94] Y. Zheng, L. Qin, T. Qiu, A. Zhou, P. Xu, and Z. Xue, “Automated detection and recognition of thyroid nodules in ultrasound images using improve cascade mask r-cnn,” *Multimedia Tools and Applications*, vol. 81, no. 10, pp. 13 253–13 273, 2022.
- [95] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, “Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. corr abs/1606.00915 (2016),” *arXiv preprint arXiv:1606.00915*, 2016.
- [96] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [97] J. Long, E. Shelhamer, T. Darrell, and U. Berkeley, “Fully convolutional networks for semantic segmentation. arxiv 2015,” *arXiv preprint arXiv:1411.4038*, 2014.
- [98] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.
- [99] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. Jorge Cardoso, “Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations,” in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, Proceedings 3*. Springer, 2017, pp. 240–248.
- [100] B. G. Reed and B. R. Carr, “The normal menstrual cycle and the control of ovulation,” 2015.
- [101] P. Kainz, M. Pfeiffer, and M. Urschler, “Segmentation and classification of colon glands with deep convolutional neural networks and total variation regularization,” *PeerJ*, vol. 5, p. e3874, 2017.
- [102] E. Pascoal, J. Wessels, M. Aas-Eng, M. S. Abrao, G. Condous, D. Jurkovic, M. Espada, C. Exacoustos, S. Ferrero, S. Guerriero *et al.*, “Strengths and limitations of diagnostic tools for endometriosis and relevance in diagnostic test accuracy research,” *Ultrasound in obstetrics & gynecology*, vol. 60, no. 3, pp. 309–327, 2022.
- [103] Q. Zhu, B. Du, and P. Yan, “Boundary-weighted domain adaptive neural network for prostate mr image segmentation,” *IEEE transactions on medical imaging*, vol. 39, no. 3, pp. 753–763, 2019.

-
- [104] Z. Xu and M. Niethammer, “Deepatlas: Joint semi-supervised learning of image registration and segmentation,” in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II* 22. Springer, 2019, pp. 420–429.
 - [105] X. Chen, B. M. Williams, S. R. Vallabhaneni, G. Czanner, R. Williams, and Y. Zheng, “Learning active contour models for medical image segmentation,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 11 632–11 640.
 - [106] J. Zhou, W. Fu, W. Hu, Z. Sun, T. He, and Z. Zhang, “Challenges and advances in analyzing tls 1.3-encrypted traffic: A comprehensive survey,” *Electronics*, vol. 13, no. 20, p. 4000, 2024.
 - [107] A. M. Eskicioglu and P. S. Fisher, “Image quality measures and their performance,” *IEEE Transactions on communications*, vol. 43, no. 12, pp. 2959–2965, 1995.
 - [108] S. Sedai, B. Antony, R. Rai, K. Jones, H. Ishikawa, J. Schuman, W. Gadi, and R. Garnavi, “Uncertainty guided semi-supervised segmentation of retinal layers in oct images,” in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part I* 22. Springer, 2019, pp. 282–290.
 - [109] M. B. Sanjeevannavar, N. R. Banapurmath, V. D. Kumar, A. M. Sajjan, I. A. Badruddin, C. Vadlamudi, S. Krishnappa, S. Kamangar, R. U. Baig, and T. Y. Khan, “Machine learning prediction and optimization of performance and emissions characteristics of ic engine,” *Sustainability*, vol. 15, no. 18, p. 13825, 2023.
 - [110] B. Yang, R. Zhang, H. Peng, C. Guo, X. Luo, J. Wang, and X. Long, “Slp-net: An efficient lightweight network for segmentation of skin lesions,” *Biomedical Signal Processing and Control*, vol. 101, p. 107242, 2025.
 - [111] M. Puttagunta and S. Ravi, “Medical image analysis based on deep learning approach,” *Multimedia tools and applications*, vol. 80, no. 16, pp. 24 365–24 398, 2021.
 - [112] H. Kwon, S. H. Oh, M.-G. Kim, Y. Kim, G. Jung, H.-J. Lee, S.-Y. Kim, and H.-M. Bae, “Enhancing breast cancer detection through advanced ai-driven ultrasound technology: A comprehensive evaluation of vis-bus,” *Diagnostics*, vol. 14, no. 17, p. 1867, 2024.
 - [113] W. Du, L. Zhang, E. Suh, D. Lin, C. Marcus, L. Ozkan, A. Ahuja, S. Fernandez, I. I. Shuvo, D. Sadat *et al.*, “Conformable ultrasound breast patch for deep tissue scanning and imaging,” *Science Advances*, vol. 9, no. 30, p. eadh5325, 2023.
 - [114] M. Sinclair, C. F. Baumgartner, J. Matthew, W. Bai, J. C. Martinez, Y. Li, S. Smith, C. L. Knight, B. Kainz, J. Hajnal *et al.*, “Human-level performance on automatic head biometrics in fetal ultrasound using fully convolutional neural networks,” in *2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*. IEEE, 2018, pp. 714–717.
-

- [115] C. F. Baumgartner, K. Kamnitsas, J. Matthew, T. P. Fletcher, S. Smith, L. M. Koch, B. Kainz, and D. Rueckert, “Sononet: real-time detection and localisation of fetal standard scan planes in freehand ultrasound,” *IEEE transactions on medical imaging*, vol. 36, no. 11, pp. 2204–2215, 2017.
- [116] B. Pu, K. Li, S. Li, and N. Zhu, “Automatic fetal ultrasound standard plane recognition based on deep learning and iiot,” *IEEE Transactions on Industrial Informatics*, vol. 17, no. 11, pp. 7771–7780, 2021.
- [117] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.
- [118] V. Badrinarayanan, A. Kendall, and R. Cipolla, “Segnet: A deep convolutional encoder-decoder architecture for image segmentation,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.
- [119] M. D. Zeiler, “Adadelata: an adaptive learning rate method,” *arXiv preprint arXiv:1212.5701*, 2012.
- [120] Y. Liu, L. Chu, G. Chen, Z. Wu, Z. Chen, B. Lai, and Y. Hao, “Paddleseg: A high-efficient development toolkit for image segmentation,” 2021.
- [121] P. Contributors, “Paddleseg, end-to-end image segmentation kit based on paddlepaddle,” <https://github.com/PaddlePaddle/PaddleSeg>, 2019.