Deep Learning Ultrasound Image Analysis: From Classification to Segmentation with Limited Data

Bahareh Behboodi

A Thesis
in
The Department
of
Electrical and Computer Engineering

Presented in Partial Fulfillment of the Requirements
For the Degree of
Doctor of Philosophy (Electrical and Computer Engineering) at
Concordia University
Montréal, Québec, Canada

October 2024

CONCORDIA UNIVERSITY
SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: **Bahareh Behboodi**

Entitled:

Deep Learning Ultrasound Image Analysis: From Classification to Segmentation with Limited Data

and submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY (Electrical and Computer Engineering)

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

_____ Chair
Dr. Nematollah Shiri

_____ External Examiner
Dr. Jochen Lang

_____ Arm's Length Examiner
Dr. Arash Mohammadi

_____ Examiner
Dr. Wei-Ping Zhu

_____ Examiner
Dr. Maria Amer

_____ Thesis Supervisor
Dr. Hassan Rivaz

Approved _____
Dr. Jun Cai, Graduate Program Director

_____

Date of Defence _____
Dr. Mourad Debbabi, Dean, Engineering and Computer Science

Abstract

Deep Learning Ultrasound Image Analysis: From Classification to Segmentation with Limited Data

Bahareh Behboodi, Ph.D.

Concordia University, 2024

Ultrasound (US) is one of the most widely used imaging modalities in diagnostic and surgical settings due to its affordability, safety, and non-invasive nature. However, US images are prone to speckle noise, leading to low resolution and making clinical interpretation challenging. Recently, researchers have applied state-of-the-art deep learning (DL) algorithms from the field of computer vision to the clinical domain. These algorithms require extensive annotated data to achieve meaningful results. In clinical US imaging, however, there are limited available datasets because annotating US images is time-consuming and requires expert radiologists. Additionally, many hospitals restrict data sharing due to patient privacy policies, further limiting the development of DL algorithms for clinical US images. To address these limitations, this thesis focuses on developing innovative DL algorithms capable of performing with small datasets. Specifically, in Chapter 2, we use simulated US images as an alternative dataset to pre-train a breast tumor segmentation model. We further explore how network design complexity affects segmentation performance with limited data. In Chapter 3, we leverage 2D planes from 3D uterus US scans to develop a segmentation model using data from only 10 cervical cancer patients. In Chapter 4, we create a compact segmentation network with just 0.82 million parameters, applying knowledge distillation to transfer knowledge from a well-trained teacher model with 96 million parameters. This approach is ideal for portable US devices, where computational and memory-efficient models are required at the bedside. In Chapter 5, we introduce a novel approach to breast lesion classification by incorporating background as an additional class, improving the detection of invasive ductal carcinomas. In Chapter 6, we develop a framework for detecting quadriceps muscle thickness in US images, an important biomarker for frailty assessment. This framework also provides activation maps, highlighting the model's focus on either the muscle body or bone surface. The availability of well-annotated datasets for DL model development has been a significant challenge in this thesis. To address this gap, in our final chapter, Chapter 7, we present a publicly available, expert-annotated dataset of intra-operative US images for brain tumor resection—the first of its kind, verified by two expert surgeons. Finally, in Chapter 8, we summarize our findings with concluding remarks and potential future works.

Acknowledgments

First and foremost, I would like to express my deepest gratitude to my supervisor, Prof. Hassan Rivaz, for his invaluable guidance, unwavering support, and continuous encouragement throughout this journey. His expertise, insight, and patience have been instrumental in shaping this work, and I am truly grateful for his mentorship. I would also like to thank my labmates, who have been a wonderful source of support and collaboration. The discussions we shared made the lab a truly enjoyable place to work.

I also wish to extend my sincere appreciation to the members of my thesis committee, Dr. Nematollah Shiri, Dr. Jochen Lang, Dr. Arash Mohammadi, Dr. Wei-Ping Zhu, Dr. Maria Amer. Your valuable feedback, insightful comments, and constructive criticism have significantly contributed to the improvement of this work. Thank you for your time, effort, and dedication.

I am especially thankful to Dr. Rupert Brooks, who provided substantial help and offered valuable insights into the finer details of this research. Your support and expertise were crucial to the success of this thesis, and I am deeply appreciative of your contributions.

I am deeply thankful for the endless love and support of my family. Your unwavering belief in me has been my greatest source of strength. You have always been there to offer a kind word, a listening ear, and endless encouragement, and for that, I am forever grateful.

Special thanks to my friend Hamze, whose friendship and support have meant the world to me. Your encouragement and companionship have been a constant source of motivation, and I am truly fortunate to have you by my side.

I would like to extend my heartfelt appreciation to my dear friend Hanie. Your kindness, understanding, and unwavering support have been a pillar of strength for me throughout this journey. Thank you for always being there for me, both in times of joy and challenge.

Lastly, I want to express my deep gratitude to my friend Mozhi (Rooti). Her motivational support was incredibly helpful during the challenging time of my PhD life, and I am truly thankful for her unwavering encouragement.

Contents

List of Figures

List of Tables

List of Abbreviations

ANN Artificial Neural Network

AUC Area Under Curve

BCE Binary Cross Entropy

BN Batch Normalization

CFS Clinical Frailty Score

CNN Convolutional Neural Network

DL Deep Learning

DSC Dice Similarity Coefficient

FCN Fully Convolutional Network

FI Frailty Index

iMR Intra-operative MRI

iUS Intra-operative Ultrasound

ML Machine Learning

NLP Natural Language Processing

QMT Quadriceps Muscle Thickness

RF Radio-frequency

RNN Recurrent Neural Network

SNR Signal-to-noise Ratio

US Ultrasound

ViT Vision Transformer

Chapter 1

Introduction

This chapter begins with a brief overview of ultrasound (US) imaging, highlighting its applications in the medical field. It then explores the applications of US image analysis using deep learning (DL) algorithms, outlines the associated challenges, and discusses the motivations and objectives that drive this research. Following that, a roadmap of the thesis is provided, offering a summary of each chapter and how they contribute to the overall study. The chapter concludes by listing the publications that have resulted from the work conducted during the current Ph.D. dissertation.

1.1 Ultrasound imaging

US imaging is a non-invasive diagnosis methodology since it utilizes low-energy US waves in order to capture tissue characterizations. A US examination involves stirring sound waves, typically within the frequency range of 500 kHz to over 50 MHz, from piezoelectric sources toward body tissues. Due to the varying echogenicity of tissues, some of these waves are attenuated and reflected back to the source. These reflected waves, known as channel data or backscattered signals, are captured for further processing. During the image formation process, beamforming of the channel data generates radio-frequency (RF) data. However, RF data is not directly suitable for visualization due to its very high-frequency content. Consequently, the envelope of the RF data is extracted. The envelope data has a wide dynamic range, it is then compressed using a logarithmic algorithm to produce a US B-mode image that can be displayed on US devices. This compression step results in the loss of all phase information. RF data contains substantially more information than both envelope data and B-mode images; however, its high-frequency content makes it difficult to

visualize. As a result, B-mode images are more commonly used in medical applications.

The brightness of organs in a B-mode image is determined by the intensity of the reflected signals. Tissues with higher echogenicity (hyperechoic) appear as brighter areas, while those with lower echogenicity (hypoechoic) are displayed in shades of gray. Anechoic tissues, which do not reflect sound waves, appear completely dark [82]. In diagnostic US imaging, expert radiologists rely on the brightness in B-mode images to identify abnormal lesions. However, the limited resolution of B-mode images can make this visual detection process time-consuming for radiologists. Despite this limitation, US remains one of the best modalities for real-time examinations due to its portability and cost-effectiveness, where timely diagnosis and treatment planning are critical for patient care. However, while US offers many advantages, B-mode images are often contaminated with speckle noise and suffer from a low signal-to-noise ratio (SNR). These factors introduce challenges to US image processing.

US image processing involves the usage of computational techniques to enhance, analyze, and interpret US images for diagnosis and monitoring purposes. Scientists have increasingly turned to automatic image processing tools to address challenges introduced by speckle noise and low signal-to-noise ratio of US images. Speckle noise reduction and image enhancement techniques are the common methods to increase visibility while maintaining the integrity of fine details, which is crucial for diagnosing. These methods usually serve as pre-processing steps that can obscure important details and hinder accurate diagnosis [151]. Various traditional despeckling methods, including spatial domain techniques, anisotropic diffusion filtering, and transform domain methods have demonstrated significant improvements in image clarity while preserving essential features [171, 259]. Other methods to extract more information from US to enhance its interpretability include elastography [85] and quantitative ultrasound [110].

Image segmentation, edge detection, classification, registration, etc., are techniques that can aid clinicians in interpreting US images. Such techniques are critical steps in accurately analyzing US images that can aid in the early detection of anomalies and timely interventions. In US image segmentation, the aim is to identify the whole body of the target tissue [168], while in US edge detection, the aim is to identify only the boundaries of an organ [223, 254]. Both techniques can assist in monitoring the development of the organ, such as fetal growth, tumor progression, and other changes. In US image classification, the aim is only to distinguish abnormal tissue from normal. Image registration techniques are essential in image-guided therapies, where accurate alignment of images from different modalities is

required [261]. Among the most commonly used approaches in the abovementioned image processing techniques, machine learning (ML) and DL algorithms are the most commonly used algorithms, which offer significant potential to enhance the accuracy and efficiency of US image interpretation.

1.2 DL in medical US imaging

While some researchers are employing state-of-the-art ML and DL algorithms to enhance the quality and resolution of US images, others are focusing on overcoming the challenges in US image processing and interpretation to accelerate its use in clinical diagnosis. The primary aim of the current Ph.D. dissertation is to explore the latter category of studies. Typically, US applications are found in computer vision-related tasks, such as segmentation, classification, regression, object detection, localization, and more. In traditional ML-based methods for US image analysis, techniques such as clustering, threshold-based models, feature engineering, and other classical approaches were commonly used [246]. While these methods led to improvements, they were rarely deployed in real clinical trials due to insufficient accuracy. However, with the advancements in DL algorithms, US image processing techniques are now being applied in real clinical settings and are increasingly used in real-life scenarios.

DL algorithms are powerful tools for automatic analysis. They use non-linear mapping functions to represent complex relationships between input and output spaces. Some of the most prominent DL architectures in US include artificial neural networks (ANN) and convolutional neural networks (CNN). Recently, architectures from the natural language processing (NLP) field have also been adopted in US.

ANNs, particularly those with fully connected layers, are designed to model complex relationships within data. Fully connected layers, where each neuron is connected to every neuron in the previous layer, enable the network to learn complex patterns and interactions. These layers facilitate the transformation of input data through a series of weighted connections and activation functions, allowing the network to capture and represent high-level features. ANNs with fully connected layers are versatile and effective for tasks such as classification and regression.

CNNs, in particular, are widely used in image analysis and often consist of multiple convolutional layers, along with pooling, normalization, and fully connected layers. The convolutional layers mimic traditional feature extraction methods in image and signal processing, where features are extracted by applying filters. These layers automatically learn

to capture important patterns in the data, making CNNs especially effective for tasks like segmentation and classification. CNN layers often resemble the manual feature engineering step in traditional ML-based methods. Consider a DL algorithm where the output of each layer (l) (i.e., feature map) is the input of the next layer ($l+1$). Given an input, the feature map is formulated as:

$$x_n^{l+1} = f\Big(\sum_{m=1}^{M} W_{nm}^{l} * x_m^{l} + b_n^{l+1}\Big) \tag{1.1}$$

where x_n^{l+1} is the n^{th} feature map of $(l+1)^{th}$ layer, W_{nm}^{l} represents the 2D convolution kernel from the m^{th} feature map of l^{th} layer which is trained through feed-forward and back-propagation steps. b_n^{l+1} is the bias in $(l+1)^{th}$ layer. $f(.)$ is the nonlinear activation function applied to the sum of convolution operations (i.e., $*$).

NLP models, on the other hand, are designed to analyze and understand human language through DL techniques. They use embedding layers to convert words into dense vectors that capture semantic meanings. Recurrent layers, such as LSTMs and GRUs [44], handle sequential dependencies in text, while attention mechanisms, including those in transformers, enable the model to focus on relevant parts of the input. Transformers, like BERT [52] and GPT [183], process entire sequences simultaneously to capture long-range context effectively. These models excel in tasks such as text classification, sentiment analysis, and machine translation, making them crucial for applications across various fields, including healthcare and customer service.

A DL-based algorithm may consist of various combinations of the abovementioned models. In DL-based segmentation techniques, the CNN layers follow the fully convolutional networks (FCN) scheme [140] to create an output with the same size as the ground-truth masks. Fully convolutional networks comprise three main components: encoder (or feature extractor), bottle-neck, and decoder. The most common networks for semantic segmentation are UNet [188], V-Net [153], etc. For classification tasks, there are two main sections: feature extractor and classifier. The classifier is assigned to fully connected layers (i.e., ANNs) where all the neurons in this layer are connected with neurons in the previous layer. The number of neurons in the last layer of fully connected layers is equal to the number of classes. The most common clasisfications networks to name are VGG [140], ResNet [89], MobileNet [192], EfficientNet [220], etc. In addition to traditional CNN architectures, Vision Transformers (ViTs) [58] have emerged as a powerful alternative for image analysis. ViTs utilize self-attention mechanisms to capture relationships between different parts of the image, making them particularly effective for tasks that require understanding contextual information across various

regions. In the context of US imaging, ViTs can be beneficial for understanding complex spatial relationships.

1.3 Thesis statement

Research in US image processing can be broadly divided into two main categories. The first focuses on image reconstruction [75, 78, 79] and image enhancement [200, 201], which improve the visualization of fine details in US images. The second category involves studies on US image interpretation, which support more accurate diagnosis and effective treatment planning. Within this scope, our work specifically concentrates on segmentation, classification, and regression algorithms, which can aid clinicians to better use US images and improve clinical decision-making.

The segmentation of US images presents more challenges compared to other techniques, but it offers a wider range of applications. For instance, it is particularly valuable in guided surgeries. Additionally, it enables tracking and monitoring changes in organ geometry during treatments, such as tumor growth or reduction, as well as changes in tissue size. This is why our primary focus has been on segmentation applications. However, we also engage in the classification and regression of US images. From a development perspective, classification is quite similar to segmentation. In classification, we assign labels to entire images, while in segmentation, labels are applied at the pixel level. Classification is especially useful for detecting abnormalities during patient assessments and for diagnostic purposes. Regression analysis, similar to classification, assigns a value to each image. However, while classification produces integer labels, regression provides continuous values, allowing for a more nuanced interpretation of the data. Despite the remarkable performance of DL-based techniques in segmentation, classification, and regression applications, validating them on US images may be problematic for two main reasons. First, for any new dataset, especially for applications that require real-time segmentation, the performance of the state-of-the-art DL-based techniques adopted from computer vision studies must be rigorously evaluated. Second, the scarcity of publicly accessible sources with high-quality ground truth information makes the validation process challenging. To this end, the current thesis is motivated by the development of novel DL-based methods to accomplish the following overall objectives:

1. Novel segmentation and classification techniques that are effective with limited data
2. Introduction of publicly available US dataset with manual annotations to help the development of future DL methods.

It is important to note that our work spans multiple problems and methodologies related to DL in US. Therefore, we provided a general literature review in this introduction, while more focused literature reviews are provided at the beginning of each chapter. A general view of each proposed method is provided in Sec. 1.4.

1.4 Roadmap of the thesis

The remainder of the thesis is organized as follows. Chapter 2 introduces a novel approach that leverages simulated US images as a potential training dataset for tumor segmentation tasks, particularly in situations where annotated data is scarce. This chapter further explores the effectiveness of pre-training on simulated data and investigates how various factors in network design, such as architecture choices and hyperparameters, influence segmentation performance. By examining the interplay between pre-training data and network structure, this chapter aims to provide insights into optimizing DL models for US image segmentation when facing limited annotated datasets. Chapter 3 focuses on the development of DL methods for the automatic segmentation of the uterus in 3D US images. Given the limitation of available 3D scans, this chapter proposes an innovative approach by creating 2D models that effectively utilize the available data. Chapter 4 delves into optimizing knowledge distillation (KD) and teacher-student training techniques to achieve efficient breast US image segmentation. In this chapter, smaller neural networks (i.e. student) are trained using KD, enabling them to perform similar to the well-trained network (i.e. teacher) even with a limited number of training images. The chapter provides insights into the methodology and highlights the performance improvements achieved through this process, making it suitable for resource-constrained environments. It is worth noting that the proposed student network achieves comparable results to that of teacher network with only 0.82 trainable parameters. Chapter 5 introduces a novel approach to breast cancer classification in US images by leveraging DL with a multi-task learning framework. This method enhances classification accuracy by handling a greater number of classes. Chapter 6 presents DeepSarc-US, a DL framework designed for assessing sarcopenia from US images. This chapter integrates advanced techniques in segmentation and classification to address the challenge of muscle thickness measurement, framing it as a regression problem. Finally, Chapter 7 presents a comprehensive dataset of manual annotations for intra-operative brain tumor US images. By providing open access to these annotations, the chapter aims to facilitate further research and development in the field of brain tumor imaging. In clinical US imaging, the availability of datasets is limited,

as annotating US images is time-intensive and requires skilled radiologists. Moreover, data sharing is often restricted by hospitals due to patient privacy regulations, posing additional challenges for advancing deep learning algorithms in this field. Consequently, this thesis focuses on developing methodologies across various US applications.

1.5 List of publications

The list of published/in preparation journal and conference papers is given below:

- **Journal papers:**

    1. Behboodi B, Brooks R, Rivaz H, "Optimizing Knowledge Distillation for Efficient Breast Ultrasound Image Segmentation: Insights and Performance Enhancement", Artificial Intelligence in Health. 2024 (accepted-under publication)
    2. Behboodi B, Obrand J, Afilalo J, Rivaz H. DeepSarc-US: A Deep Learning Framework for Assessing Sarcopenia Using Ultrasound Images. Applied Sciences. 2024; 14(15):6726.
    3. Behboodi, B., Carton, X., Chabanas, M., Solheim, O., R. Munkvold, B. K., Rivaz, H., Xiao, Y., Reinertsen, I. "Open access segmentations of intraoperative brain tumor ultrasound images", Medical Physics. 2024.
    https://doi.org/10.1002/mp.17317

- **Conference papers:**

    1. Behboodi B, Rivaz H, "Ultrasound segmentation using U-Net: learning from simulated data and testing on real data", 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE 2019.
    2. Behboodi B, Fortin M, J. Belasso C, Brooks R, Rivaz H, "Receptive Field Size as a Key Design Parameter for Ultrasound Image Segmentation with U-Net", 2020 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE 2020.
    3. Behboodi B, Amiri M, Brooks R, Rivaz H, "Breast Lesion Segmentation in Ultrasound Images with Limited Annotated Data", 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). IEEE 2020.

4. Behboodi B, Rivaz H, Lalondrelle S, Harris E, "Automatic 3D Ultrasound Segmentation of Uterus Using Deep Learning", 2021 IEEE International Ultrasonics Symposium (IUS). IEEE 2021.

5. Behboodi B, Rasaee H, K. Z. Tehrani A, Rivaz H, "Deep classification of breast cancer in ultrasound images: more classes, better results with multi-task learning", Proc. SPIE, Medical Imaging 2021: Ultrasonic Imaging and Tomography.

We also made contributions to the following papers, which are not included in this thesis:

1. Belasso C. J., Behboodi B, Benali H, Boily M, Rivaz H, Fortin M, "LUMINOUS database: lumbar multifidus muscle segmentation from ultrasound images", BMC Musculoskeletal Disorders, 21, 1-11. 2020. [23]

2. Amiri M, Brooks R, Behboodi B, Rivaz H, "Two-stage ultrasound image segmentation using U-Net and test time augmentation". International journal of computer assisted radiology and surgery, 15, 981-988. 2020.[4]

Chapter 2

Effects of Pre-training Data and Network Design on Segmentation of Ultrasound Images

This chapter is based on our published papers in [14, 15, 16].

2.1 Background

In medical applications, image segmentation has been used in diagnosis, image-guided interventions, pre-surgical planning, etc [260]. As discussed in Chapter 1, US, as a non-invasive diagnosis methodology, utilizes low-energy US waves in order to capture tissue characterizations. However, due to data acquisition limitations and speckle noise, an experienced radiologist is always required for interpreting US images with high complexity and ambiguity [167]. Traditional machine learning and recent deep learning methodologies have been adopted for US segmentation analysis. Despite the remarkable success of CNNs, their achievements rely on a large number of training images. However, in medical tasks, preparing such large datasets is expensive in terms of cost and time. To cope with this limitation in medical tasks, one group of studies is taking advantage of pre-trained networks and augmentation methodologies [229, 263]. Another group of studies is proposing new CNN architectures such as fully convolutional network (FCN) [140], U-Net [188], SegNet [9] and V-Net [154]. There are also several proposed networks built upon U-Net such as Deep Residual U-Net [263], Attention U-Net [169], Deeply-supervised CNN [264] and Inception U-Net [182], to name a few.

In 2018, Kumar *et al.* [120] proposed a Multi U-Net algorithm for real-time and automatic segmentation of breast masses, in which they achieved a Dice score of 0.82. The proposed algorithm originated from the U-Net structure proposed by Ronneberger *et al.* [188]. A similar adaptation of U-Net was introduced in Carton *et al.* [33] for the automatic segmentation of the resection cavity. Yap *et al.* [250] exploited three different well-known CNN-based structures for breast lesion detection and compared the results with conventional segmentation methods. A 3-D CNN-based structure with additional post-processing steps was explained in Chiang *et al.* for breast tumor detection [42]. Anas *et al.* [5] employed a novel approach based on recurrent neural network (RNN) for ultrasound segmentation during prostate biopsy incorporating MRI. They have investigated the time series of ultrasound images. Another methodology in prostate segmentation has been proposed by Wang *et al.* [233]. They investigated prostate segmentation in transrectal ultrasound using a CNN-based approach named deep attentional features (DAF). SUMNet was exploited by Nandamuri *et al.* [164] in order to segment thyroid and intravascular segmentation in 3D ultrasound volumes. The structure was built upon FCN [140]. Isensee *et al.* [108] proposed nnU-Net, a guideline in architectural design and hyperparameter tuning that performs well on a wide range of datasets. Peng *et al.* [175] showed that the receptive field plays an important role in segmentation tasks. They proposed a Global Convolutional Network wherein a large kernel size is adopted in the architecture design.

2.2 Problem Statement

Despite the ever-growing body of literature proposing new CNN-based algorithms for medical segmentation tasks, few are successful in re-applying to different medical datasets. A growing number of researchers are focusing on the applications of recently developed deep learning methodologies on ultrasound segmentation. However, the current results are not sufficiently accurate, robust, and generalizable enough for clinical trials. In addition, training deep learning architectures requires large amounts of data, which, in the case of ultrasound images, is expensive in terms of data acquisition and interpretation. Moreover, although deep learning algorithms usually perform well, their performance depends on the input image type, either natural or medical images. Therefore, the proper structure should be investigated more specifically. To this end, in the current chapter, two main avenues are exploited:

- *Avenue_1: Importance of Data for Pre-Training.*
  The focus is to find the most suitable data for pre-training a well-known segmentation

architecture in order to enhance segmentation results.

- *Avenue_2: Importance of Network Design.*
  The goal is to propose a strategy showing that the receptive field is a key parameter in designing a segmentation architecture.

2.3 Methodology

In this section, the proposed two avenues are elaborated in more detail.

2.3.1 Avenue_1: Importance of Data for Pre-Training

In deep learning approaches, the improvement in results directly depends on the number of training data. Therefore, such techniques perform better if they have a larger amount of training data. In medical images, especially in US images, annotating enough number of training data is expensive. To this end, in two phases, we explore the most suitable data for pre-training a segmentation network for US images especially if limited annotated US images are available.

Proposed Strategy: *Phase_1*

In *Phase_1* of the proposed strategy, the use of simulated data for pre-training the segmentation network is explored [14]. Then, the pre-trained network is tested on tissue-mimicking phantom data. As previously discussed, annotation of US images is an expensive task in terms of cost and time. On the other hand, the performance of deep learning algorithms directly depends on the number of training sets. Therefore, in the current phase, simulated US images are proposed as the suitable training set as the annotations already exist. In addition, the number of simulated US images is not limited. The contributions of *Phase_1* are summarized as follows:

- Can a network be trained on simulation data and tested on real data?
- Which data provides the best segmentation results, RF data, envelope data, or B-mode images?

Simulated US images are generated using the publicly available US simulation software, Field II [111, 112] which is based on MATLAB release 2018. The software is based on the physics of ultrasound waves which spread spherically. The received response of a spherical

wave emitted by a point is called spatial impulse response. The spatial impulse response varies as a function of position and formulates the ultrasound field as a function of time. The ultrasound field is emitted as soon as the virtual transducer is excited by the delta function. The parameters for the virtual transducer and phantom are summarized as outlined in Table 2.1. We randomly distribute scatterers (i.e., points with different acoustic impedances to the surrounding tissue and scatter the waves) in the virtual phantom to ensure each mm3 has, on average, 4 scatterers. The simulated images randomly consist of hyperechoic lesions (i.e., tissues with higher echogenicity) and anechoic lesions (i.e., tissues with lower echogenicity). In hyperechoic lesions, the scatterer intensities are k times larger than the background, where k is a random integer value between 1 and 10. The virtual lesions are placed between -20 and +20 mm in the lateral direction and between 30 and 90 mm in the axial direction. Lesion shapes are circles or ellipses with random sizes. The radii of circles are between 1-3 mm, and the semi-major and semi-minor axes of ellipses are between 5-9 and 1-5 mm, respectively.

Table 2.1: Virtual US transducer parameters.

Property Name	**Property Value**
Number of RF lines	50
Start depth of virtual phantom	30 *mm* from the transducer surface
Depth of virtual phantom	90 *mm* from the transducer surface
Lateral distance of virtual phantom	40 *mm* (from -20 to 20 *mm*)
Speed of sound	1540 m/s
Center frequency	3.5 MHz
Sampling frequency	100 MHz

In total, 700 images are simulated. We then split the data into training, validation, and testing data sets considering 60%, 15%, and 25% splitting factors of the total images, yielding 420, 105, and 175 images, respectively. RF, envelope, and B-mode images with the initial size of 14069×50 are then resized to 512 × 512, and mirrored to size 572 × 572. The intensity range is normalized to the range between 0-1 before feeding to the U-Net. As mentioned earlier, simulated data may consist of one hyperechoic and one anechoic lesion. Therefore, including the background, three classes should be categorized. The ground truth of a simulated image is in the size of 388×388×3.

RF data of the tissue-mimicking phantom is acquired from a CIRS Multi-Purpose Multi-Tissue ultrasound phantom with an Alpinion E-Cube system (Bothell, WA) using the L3-12H transducer at the center frequency of 10 MHz and a sampling rate of 40 MHz. Table 2.2 indicates the setup of the US transducer. Figure 2.1 represents the tissue-mimicking phantom

that has been used. The phantom data includes different types of lesions in different depths with circular shapes. In this work, a total of 6 phantom images with a depth of 40 mm are acquired from different locations of the phantom.

Table 2.2: Parameters of the Alpinion US machine.

Property Name	**Property Value**
Number of RF lines	384
Start depth of simulation data	4 *mm* from the transducer surface
Depth of simulation data	40 *mm* from the transducer surface
Lateral distance of simulation data	40 *mm* (from -20 to 20 *mm*)
Speed of sound	1540 m/s
Center frequency	10 MHz
Sampling frequency	40 MHz

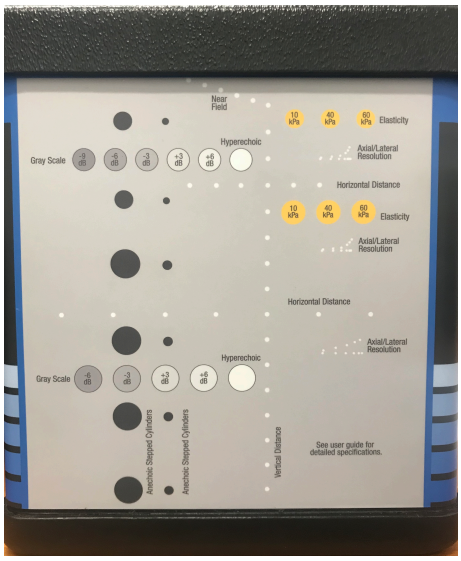


Figure 2.1: CIRS Multi-Purpose Multi-Tissue US phantom.

The U-Net architecture [188] is used for training which consists of two paths, a contracting path and an expansive path. The contracting path (left path in Fig. 2.2) consists of several repetitions of two convolution (conv) and one max-pooling (max pool) layers with the kernel size of 3×3 and 2×2, respectively. The expansive path (right path in Fig. 2.2) comprises the repetition of the concatenation of the features extracted from corresponding layers in the contracting path (copy and crop), two convolutions, and one upsampling (up-conv) layers. In this path, the kernel for the convolution layer is the same as the contracting path and 2×2

for the upsampling layer. The final layer has 1×1 convolution kernels. Activation functions in convolution layers are set to ReLU (see Eq. 2.1) except the last layer, which is set to Softmax (see Eq. 2.2). We pose the segmentation problem as a pixel-wise classification that leads to a three-class classification for our dataset. The last layers of our architecture are three 1×1 convolution layers where the loss function is set to categorical cross-entropy. The learning rate, optimizer, and weights initializer are set to $1e-5$, Adam [116], and He-normal [88], respectively. For the remaining parameters, we follow the initial parameters proposed in [188]. U-Net is trained for 100 epochs on simulated RF, envelope, and B-mode images of solely the simulation data, yielding three different trained weights. Subsequently, the trained weights are used to test on simulated (different from the training simulation set) and phantom data, yielding predicted segmentation masks. In order to fit the data in the memory, the batch size is set to 8. The codes for implementing U-Net are scripted on Python (version 3.6) using Keras with Tensorflow backend. A Titan Xp NVIDIA GPU with 12 GB of memory on Ubuntu 16.04 LTS is used for training and testing.

$$f(z) = \begin{cases} 0, & \text{if } z \leq 0 \\ z & \text{if } z \geq 0 \end{cases} \tag{2.1}$$

$$f(z_i) = \frac{\exp(z_i)}{\sum_j \exp(z_j)} \tag{2.2}$$

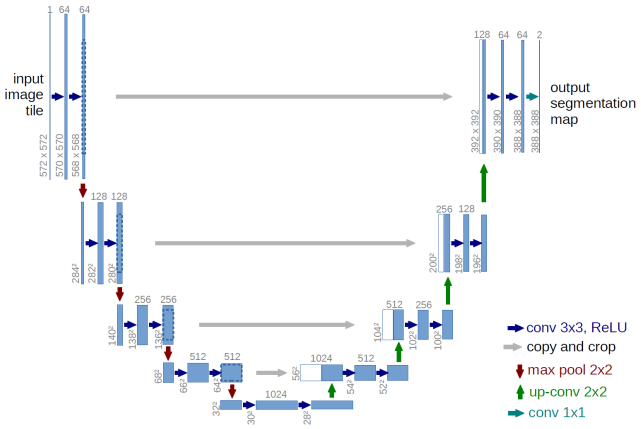


Figure 2.2: The U-Net architecture (used from [188]).

To evaluate the performance of the network, here we use two different metrics to compare the predicted mask with the ground truth mask, DSC (see Eq. 2.3) and F_2-score (see Eq. 2.4). For the simulation data, the images are simulated based on the predefined information of the location of the lesions, which is considered as the ground truth. However, for the

phantom data, the ground truth is manually obtained using the ImageJ software [194].

$$DSC = \frac{2|P \cap R|}{|P| + |R|}, \quad (P : prediction, R : groundtruth) \tag{2.3}$$

$$F_2 = \frac{5TP}{5TP + 4FN + FP}, (TP : TruePositive, FP : FalsePositive, FN : FalseNegative) \tag{2.4}$$

Proposed Strategy: *Phase_2*

In *Phase_1*, we discussed the application of simulated US images for pre-training the U-Net architecture. However, we only evaluated the pre-trained network only on a tissue-mimicking phantom. As a complementary study for *Phase_1*, we further study the application of pre-training U-Net on simulated US images for testing on *in vivo* real US images especially when limited annotations are available. To this end, three different steps are proposed as our workflow summarized in Fig. 2.3. In the first path, U-Net is trained using only 15% of the *in vivo* data. In the second path, U-Net is first pre-trained on the simulated data and then fine-tuned using the same 15% of the *in vivo* data. The third path is similar to the second path with the difference that natural images are used for pre-training. More details are provided in the following paragraphs.

The *in vivo* data includes 163 breast B-mode US images with lesions where the mean image size is 760 × 570. The breast lesions of interest are generally hypoechoic (i.e. tissues with lower echogenicity), that is, darker than surrounding tissue. Only 15% of the total number of *in vivo* images is used as training and validation sets and the remaining 85% is set as the testing set. The total number of training images selected is 4 times larger than the total number of validation images yielding 19, 5, and 139 images for training, validation, and testing sets, respectively. The same simulation data discussed in *Phase_1* is used. The natural images are publicly available at [237]. The dataset consists of 10000 images of salient objects with their annotations. In our work, the dataset was split into training, validation, and testing sets with splitting factors of 60%, 15%, and 25% of the total number of images, yielding 6000, 2500, and 1500 images, respectively. As the architecture, the same U-Net that was previously explained in *Phase_1* is used. The details of the training scheme are summarized in Table 2.3.

In the first path, the U-Net structure with the above-mentioned parameters was trained on *in vivo* images from scratch using 19 and 5 images as training and validation sets, re-

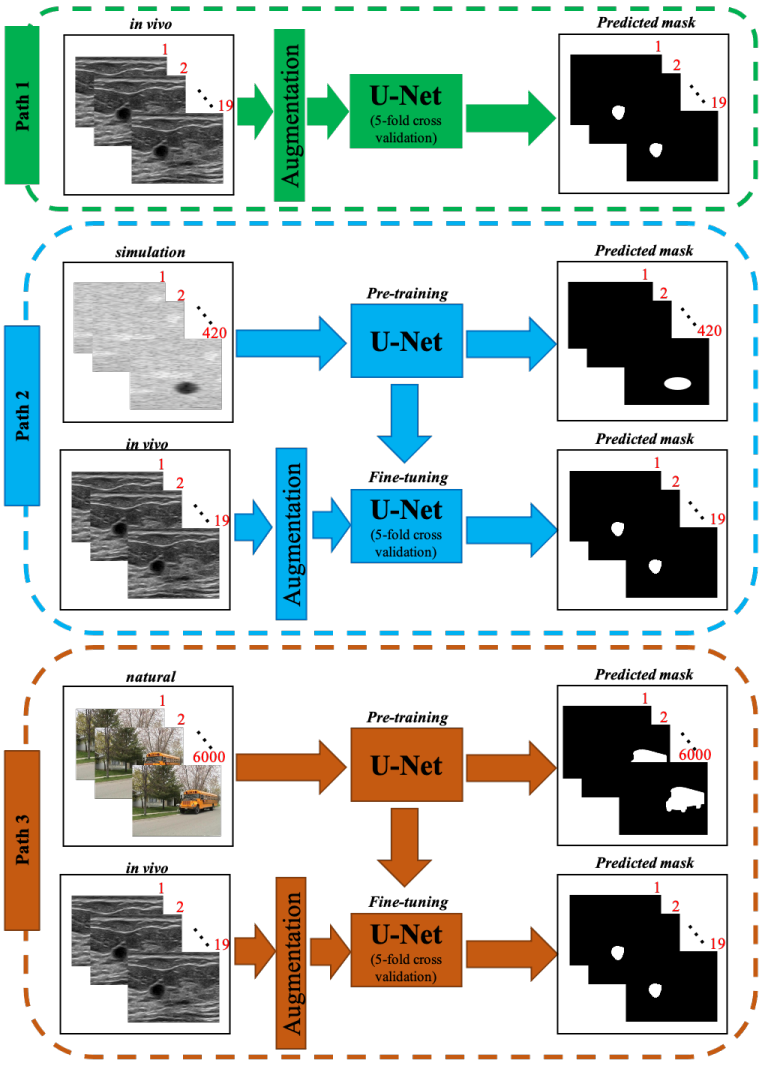


Figure 2.3: Proposed workflow for *Phase_2* when limited annotated *in vivo* data is available.

spectively, and was tested on 139 images. We call this trained network as *Pt_invivo*. Due to the small number of training data, we used 5-fold cross-validation to prevent variation in performance. Prior to each optimization iteration, we performed "on-the-fly" augmentation by applying random height-shift, width-shift, and zooming.

In the second path, U-Net was first pre-trained using 420 and 105 simulation images as its training and validation sets, respectively. Similar to the first avenue, the U-Net was initialized using parameters mentioned in Table 2.3. For simplicity, we refer to the trained U-Net with simulated data as *Pt_sim*. Afterward, the contraction path of *Pt_sim* was fine-tuned on *in vivo* training and validation sets based on parameters in Table 2.3 except that weights were initialized using the *Pt_sim* weights. We call the fined-tuned network as *Ft_sim_invivo* which was tested on *in vivo* test set. 5-fold cross-validation and "on-the-fly" augmentation

Table 2.3: U-Net parameters.

Parameter	Value
Activation function (except last layer)	ReLU [162]
Activation function (last layer)	Softmax
Loss function	Dice score
Optimizer	Adam [116]
Learning rate	10^{-5}
No. of epochs	150
Batch size	8
Weight initializer	He-normal [88]
Kernel-regularizer	L2-norm

was used for fine-tuning our *Ft_sim_invivo* network.

In the third path, similar to the second path described above, U-Net was first pre-trained and then fine-tuned on *in vivo*. However, for pre-training the network we used 6000 and 2500 natural images as training and validation sets, respectively. For simplicity, the pre-trained U-Net with natural images is referred to as *Pt_nat* and the fine-tuned network using *Pt_nat* is referred to as *Ft_nat_invivo*. 5-fold cross-validation and "on-the-fly" augmentation were used in the fine-tuning step.

In all three paths explained in *Phase_2*, we use DSC score for evaluation. In the evaluation step, the predicted masks which are the output of the last layer (i.e. Softmax layer) are first binarized using the ***argmax*** function and then compared with the ground truth masks. It is worth mentioning that in *Phase_1* three types of simulation data are used, RF, envelope, and B-mode images. However, as only the B-mode images of *in vivo* data are available, we trained U-Net solely on the B-mode of simulation data.

2.3.2 Avenue_2: Importance of Network Design

As explained in 2.2, most of the deep learning algorithms fail in re-applying to different medical datasets. Isensee *et al.* [108] proposed a guideline for designing and tuning hyperparameters of a segmentation architectural based on U-Net. Although their proposed strategy performs well in various datasets, some of their architectural choices are relatively unjustified, and their system tends to propose ensembles of networks. Therefore, their proposed scheme leads to very computationally expensive model designs. The receptive field is proposed by Peng *et al.* [175] as an important parameter in designing segmentation architectures. They proposed Global Convolutional Network wherein a large kernel size is adopted in the architecture design. U-Net by itself is a powerful architecture for medical image segmentation but

its proper design is a key element. Inspired by Isensee *et al.* and Peng *et al.*, in *Avenue_2*, we aim to explore the key factors in designing U-Net such that the sufficient receptive field is covered.

Theoretical receptive field (TRF) [141] and effective receptive field (ERF) [145] are the main factors that we consider in our proposed Strategy. Sub-sampling and dilated convolutions are the two essential parameters that directly affect the size of the receptive field [145]. Therefore, our key contribution in this section is summarized in answering the following two questions:

- How do dilated convolutions and pooling layers affect the ERF?
- How can we control the ERF with a low computational complexity?

We demonstrate how ERF and TRF can affect performance using different segmentation tasks with different sizes of the target masks in US images. We focus our experimental setup on two main phases. In *Phase_1* and *Phase_2* we explore our first and second questions, respectively. A total of 16 U-Net-based networks are proposed. In each network, a *Conv* block consists of two repetitive sets of 3×3 convolutions, batch normalization, and LeakyReLU (see 2.5) [242] activation function. An *MPool* block comprises a max-pooling layer with the default kernel of 2 unless otherwise indicated. An *Up_Conv* block includes an up-sampling with the same kernel size of its corresponding pooling layer in the contraction path, followed by a *Conv* block. Figure 2.4 shows the details of our proposed networks. The number of trainable parameters in all proposed networks is adjusted in the same range to prevent the impact of the number of parameters in training. The mean, standard deviation, minimum, and maximum number of parameters are 550070.8, 25368.6, 520138, and 594890, respectively.

$$f(z) = \begin{cases} \alpha\, z, & \text{if } z \leq 0 \\ z & \text{if } z \geq 0 \end{cases}, \quad (0 < \alpha < 1) \tag{2.5}$$

Proposed Strategy: *Phase_1*

In *Phase_1*, we aim to investigate the effects of dilated convolutions and pooling layers. To this end, we design a total of 12 U-Net-based networks. We denote 6 networks as *Dilation_nets* and 6 networks as *Pooling_nets* for our analysis on dilated convolutions and pooling layers, respectively.

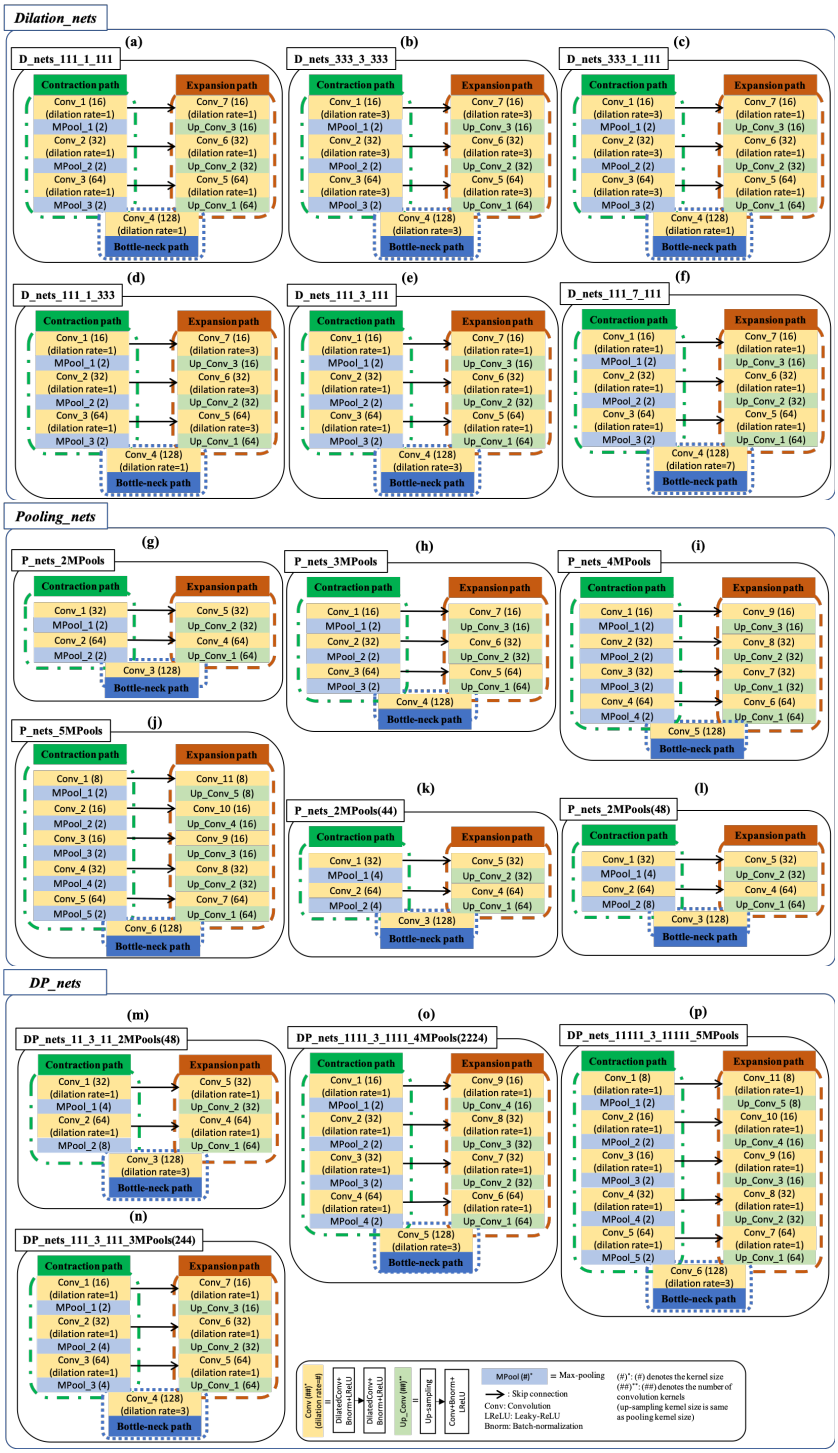


Figure 2.4: Details of proposed architectures. *Dilation_nets*: (a)-(f), *Pooling_nets*: (g)-(l), and *DP_nets*: (m)-(p).

Dilation_nets: The depth of the network (and hence the number of pooling layers) is fixed such that all the networks have only three repetitive sets of *Conv* block followed by an *MPool*

block in their contraction path. Consequently, the expansion path consists of three repetitive sets of the *Up_Conv* block followed by *Conv* block (see Fig. 2.4(a-f)). We investigate the ERF and TRF of the networks when the dilation rate is changed through contraction, bottleneck, and expansion paths. We name these networks as *D_nets_###_#_###* (see Fig. 2.4). For example, *D_nets_111_3_111* means that this network has three *MPool* block (***111***_3_111) in its contraction path, three up-sampling (111_3_***111***) in its expansion path, and all the dilation rates are set to 1 except the bottle-neck which is set to 3 (111_***3***_111).

Pooling_nets: The dilation rates of all *Conv* blocks are set to 1, with the depth set as the main difference. For example in Fig. 2.4, *P_nets_2MPool* represents a network with two repetitive sets of *Conv* block followed by a *MPool* block in their contraction path. As a result, the expansion path contains two repetitive sets of *Up_Conv* with *Conv* blocks (see Fig. 2.4 (g-l)). In the *P_nets_2MPool(84)* network, *(84)* means that the pooling kernel size of the first and second pooling layers is set to 8 and 4, respectively.

Proposed Strategy: *Phase_2*

To further comprehend the impact of the effective receptive field on the network's performance, we combine the impact of dilated convolutions and pooling layers in *Phase_2*. A total of four U-Net-based networks are designed in this phase. In Fig. 2.4, the network *DP_nets_11_3_11 _2MPool(48)* has 2 pooling layers with the kernel sizes of 4 and 8, and the dilation rate in its bottleneck is set to 3. Two different datasets are used in both *Phase_1* and *Phase_2*. The publicly available database of 163 images of breast US B-mode images [250]. In addition, 407 US images of lumbar multifidus muscle were collected at the PERFORM center with ethics approval. An expert then manually annotated the images to create segmentation masks. For simplicity, in the rest of this section, we refer to the first and second datasets as breast and muscle datasets, respectively.

All the proposed networks of *Phase_1* and *Phase_2* are trained using the following configuration. He-normal [88] weight initialization approach with L2-norm regularizer is used. The activation functions are set to LeakyReLU [242] (except the last layer) (see Eq. 2.1) and Softmax (only for the last layer) (see Eq. 2.2). The Adam [116] optimizer is optimizing the Dice loss function while the learning rate is adjusted using the cyclical learning rate approach [208] for adjusting the learning rate. The Dice loss function is used as defined in Eq. 2.6.

$$DSC\_loss = 1 - DSC \quad \text{(see Eq. 2.3)} \tag{2.6}$$

where G and P are ground truth and predicted masks, respectively, with the ϵ of 0.001 to prevent division by zero. Training is over 1000 epochs for sufficient convergence. A batch size of 4 is used for all networks except those that have 2 max-pooling layers in which the batch size is set to 2 due to lack of memory. All datasets are split into training, validation, and test sets with the splitting factor of 64%, 16%, and 20%, respectively. Images are mirrored to have a unique size of 800×800. For the evaluation, similar to *Avenue_1*, we evaluate the predicted masks with their ground truths using the DSC score. First, the output of the last layer, the Softmax layer, is thresholded with a factor of 0.5 in order to have binary predicted masks. We further investigate the difference between ERF and TRF among the models.

2.4 Results

2.4.1 Avenue_1: Importance of Data for Pre-Training

Phase_1

To provide a comprehensive comparison, the results of predicted masks derived from training U-Net based on simulated RF, envelope, and B-mode images and testing on both simulated and phantom data are illustrated in this section. Furthermore, we outline the DSC and F_2-scores for both simulated and phantom data.

Fig. 2.5 represents an example of simulated RF data, B-mode image, the ground truth mask, and the predicted masks. In this particular example, the simulated data consists of 6 lesions including 5 hyperechoics and one anechoic. It is important to highlight that four of the lesions are located on the borders and therefore are only partly contained in the image. The predicted masks provide clearer boundaries of all aforementioned lesions compared to the ground truth mask. Mean and standard deviation of DSC and F_2-scores for predicted masks from the network trained on RF, envelope, and B-mode image are summarized in Table 2.4. The mean of evaluation scores for the predicted masks from the network trained

Table 2.4: DSC and F_2 scores for the simulated data.

Predicted Mask	DSC	F_2
RF data	0.83 ± 0.18	0.82 ± 0.2
Envelope data	0.85 ± 0.16	0.87 ± 0.19
B-mode image	0.85 ± 0.16	0.85 ± 0.2

on RF, envelope, and B-mode image are 83%, 85%, and 85% for DSC, and 82%, 87% 85% for

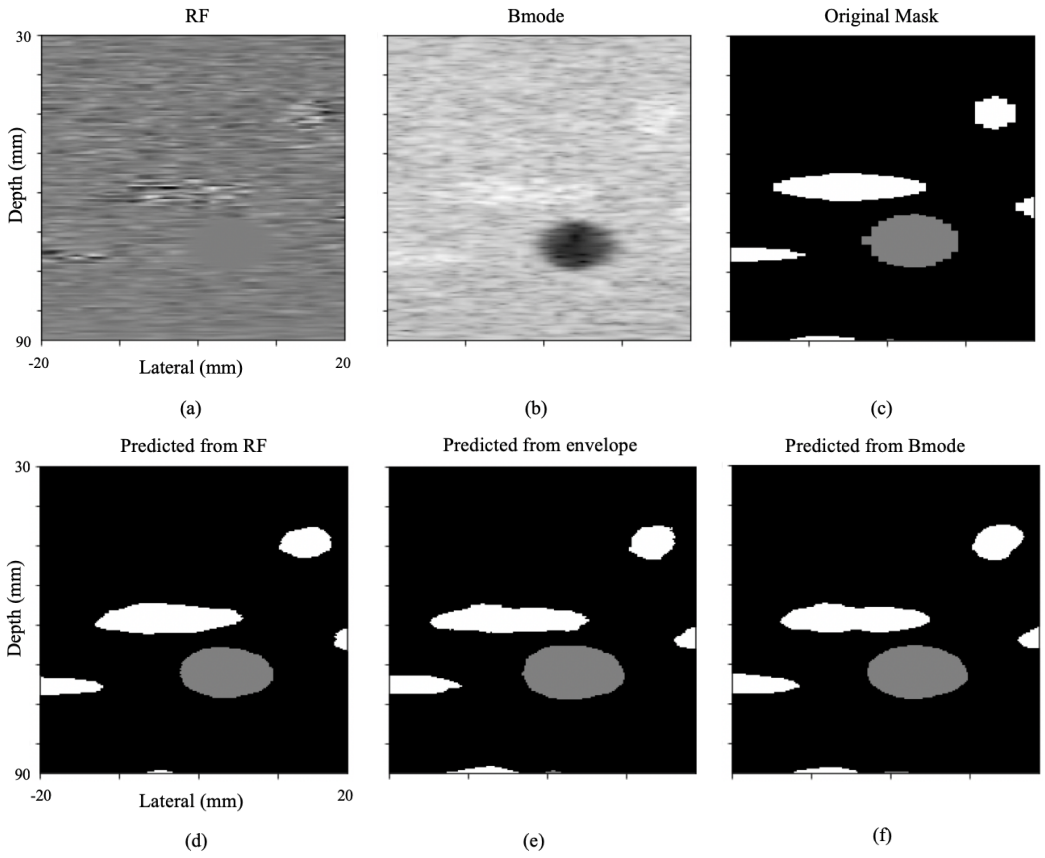


Figure 2.5: Field-II simulation images. An example of (a) RF data, (b) B-mode image, and (c) the ground truth mask. Predicted masks of the U-Net pre-trained on (d) RF, (e) envelope, and (f) B-mode images.

F_2-score, respectively. The high values in both DSC and F_2 scores indicate that U-Net has a promising structure in the segmentation of ultrasound images and is capable in learning the intrinsic features of simulated data. Furthermore, it shows that the network can learn mappings from the domain of RF, envelope, or B-mode image to pixel-level segmentation mask.

An example of the RF data and B-mode image of our tissue-mimicking phantom is shown in Fig. 2.6. Figures 2.6 (d), (e), and (f) show the results of training our network on RF, envelope, and B-mode images of simulated data and testing on the phantom data, respectively. In this particular example, the phantom data consists of three lesions. In all predicted masks, the anechoic lesion (dark cyst), which is more clearly visible, is segmented successfully. The mask derived from RF data clearly outperforms the envelope mask, which itself outperforms the B-mode mask. Table 2.5 presents the quantitative evaluation for phantom data. The mean of evaluation scores for the predicted masks from the network

trained on RF, envelope, and B-mode image are 31%, 31%, and 26% for DSC, and 27%, 27% 20% for F_2-score, respectively. It is important to highlight that the network has not seen any real images and is fully trained on simulation data. Two conclusions can be made from this observation. First, the Field-II simulation model creates US images quite similar to real US images which can be used for training deep learning techniques. Second, the network is not suffering from over-fitting and further has learned an efficient representation of US images.

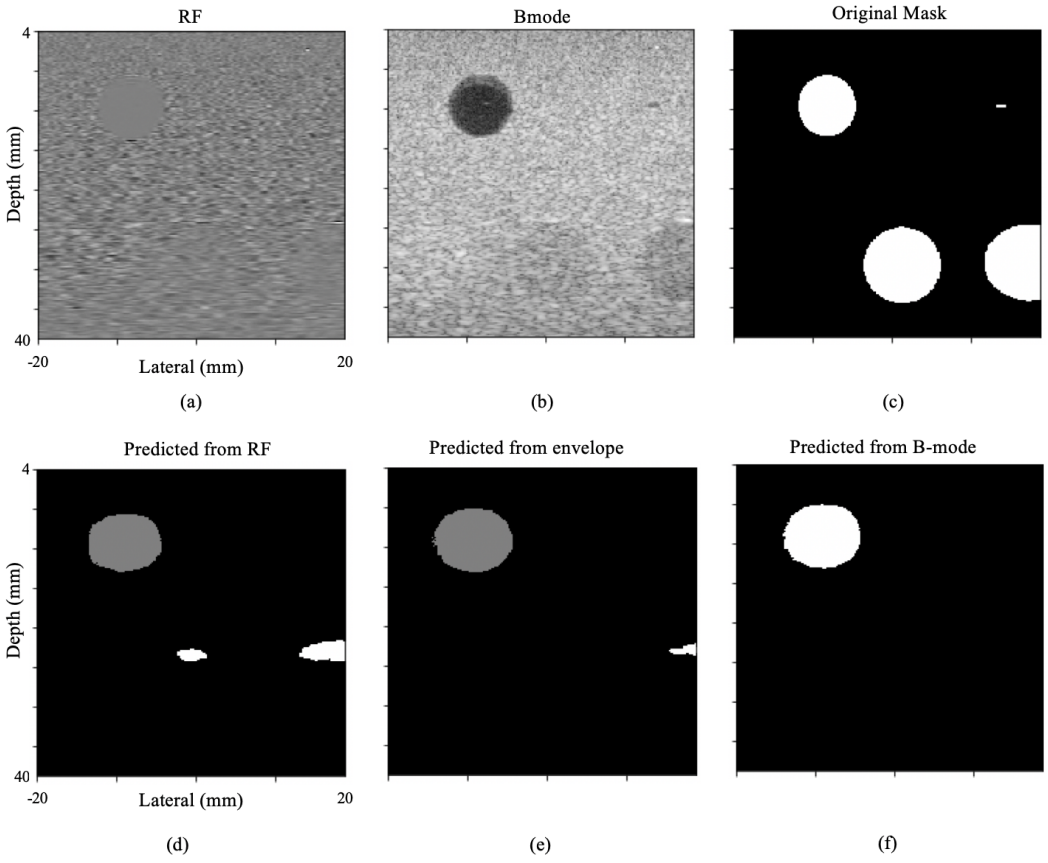


Figure 2.6: An example of real tissue-mimicking phantom (a) RF data, (b) B-mode image, (c) the ground truth mask (the envelope data is not shown due to space limitations). The predicted masks with (d) RF, (e) envelope, and (f) B-mode images.

Table 2.5: DSC and F_2 scores for the real phantom data.

Predicted Mask	**DSC**	**F_2**
RF data	0.31 ± 0.25	0.27 ± 0.23
Envelope data	0.31 ± 0.23	0.27 ± 0.2
B-mode image	0.26 ± 0.1	0.2 ± 0.1

Phase_2

In this section, the results of all three paths are compared in-depth. Table 2.6 presents the ***DSC*** scores of the predicted masks derived from *Pt_invivo*, *Pt_sim*, *Ft_sim_invivo*, *Pt_nat*, and *Ft_nat_invivo* networks for both training and testing *in vivo* sets. The ***DSC*** score for the test set increases when we fine-tune the pre-trained network no matter what type of images were used during pre-training. Therefore, pre-training the network performs better than training from scratch with limited training data. It is worth mentioning that we used 6000 number of natural images and 420 number of simulated images during pre-training. However, when we decreased the number of natural images in the third path from 6000 to 420 in order to be equal to the number of simulated images, the ***DSC*** score was reduced from 0.56 to 0.38 as shown in Table 2.6 (*Pt_nat420*, and *Ft_nat420_invivo* are referred as repetition of third path using 420 natural images). As a result, pre-training the network using simulated data is preferable as the auxiliary data than using natural images when the same number of images from both datasets is available. Figure 2.7 demonstrates examples of the predicted masks with their ***DSC*** scores.

We had 6000 natural images in which 29 hours were needed to pre-train the *Pt_nat* network on a device equipped with a Titan Xp NVIDIA GPU with 12 GB of memory on Ubuntu 16.04 LTS. However, for pre-training on simulation only 2 hours are required. Training/fine-tuning on *in vivo* needs 5 minutes. As more annotations become available, although the U-Net is better trained, more time is needed for the pre-training step.

Table 2.6: Mean and standard deviation of ***DSC*** scores for predicted masks of *in vivo* train and test sets over 5-fold cross-validation.

Network Name		**Train *in vivo***	**Test *in vivo***
Path 1	*Pt_invivo*	0.73 ± 0.03	0.37 ± 0.04
Path 2	*Pt_sim*	0.29 ± 0.22	0.27 ± 0.19
	Ft_sim_invivo	0.79 ± 0.05	**0.45 ± 0.03**
Path 3	*Pt_nat*	0.13 ± 0.27	0.14 ± 0.26
	Ft_nat_invivo	0.85 ± 0.04	**0.57 ± 0.02**
Ft_nat420_invivo		0.78 ± 0.15	0.40 ± 0.03

We highlight that pre-training the network performs better compared to training from scratch for *in vivo* data especially when limited annotations are available. Although pre-training on natural images leads to more accurate predictions for *in vivo* images, it requires more hours of training.

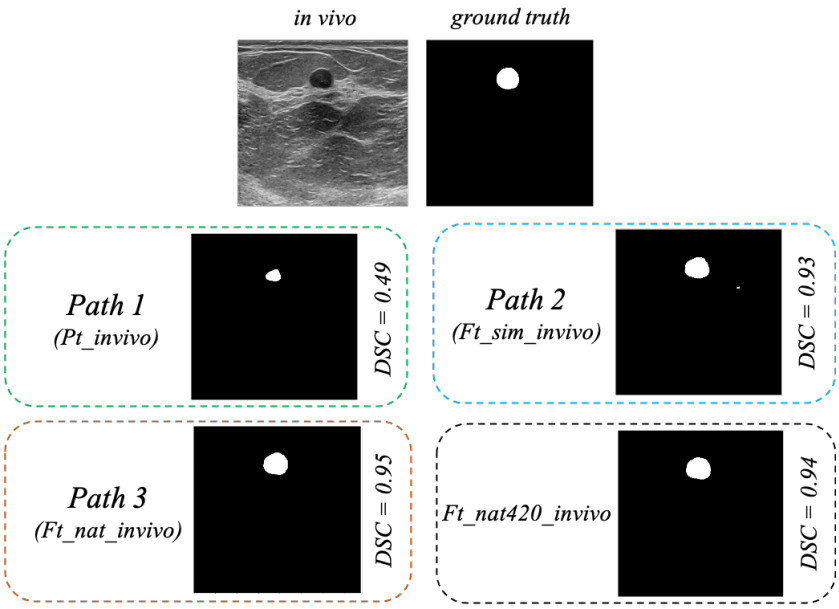


Figure 2.7: Examples of segmentation results and their *DSC* scores derived from *Avenue* 1, *Avenue* 2, *Avenue* 3, and *Ft_nat*420*_invivo*.

2.4.2 Avenue_2: Importance of Network Design

Phase_1

In *Dilation_nets* for breast lesion dataset, two models *D_nets_333_3_333* and *D_nets _111_7_111* perform very similarly and increase the DSC by 62% compared to the *D_nets _111_1_111*. For the muscle dataset, model *D_nets_111_7_111* outperforms other networks. These findings indicate the importance of having dilation in the bottleneck of the U-Net.

In *Pooling_nets* Comparing the results of two networks *P_nets _5MPools* and *P_nets _2MPools(84)* in both breast and muscle datasets, presents the high impact of pooling layers' kernel size in performance.

Phase_2

By combining the impact of the dilated convolutions with pooling kernel size, shallower networks (i.e. *DP_nets_111_3_111 _3MPools(244)* and *DP_nets_1111_3_1 111_4MPools (2224)*) are comparable to the deepest network (i.e. *DP_nets_11111_ 3_11111_5MPools*.

Effective and Theoretical Receptive Fields

ERF and TRF of six networks are shown in Fig. 2.8 for both breast and muscle US datasets. In Fig. 2.8, the white square shows the TRF, and ERF is a fraction of TRF. It is clearly

shown that by changing the dilation rate and the pooling kernel size, the size of ERF and TRF change. In *D_nets_111_ 7_111* and *DP_nets_11111_ 3_11111_5MPools* the TRF and ERF are divided into small rectangles indicating that pixels in between do not affect the receptive fields. *DP_nets_11_ 3_11_2MPools(48)*, as a shallow network, in both breast and muscle datasets provides the Dice score of 0.45 ± 0.28 and 0.75 ± 0.15, respectively, which are comparable to *P_nets_5MPools* as a deep network.

Figure 2.9 provides the predicted segmentation. By comparing the predicted segmentations of *D_nets_111_ 1_111* network with *D_nets_11_ 3_11_2MPools(48)* we see that the deeper network does not always lead to better results and dilated convolution and pooling kernel size are the factors that need to be taken into consideration for practical design of the network. Our results show that by adjusting the size of pooling layers and employing dilated convolution, we can control the size of the ERF, the key parameter in designing a network that is computationally effective.

Table 2.7: TRF size, the mean and standard deviation of Dice scores (DSC) for breast and muscle datasets.

	Model	***TRF size***	***DSC Breast***	***DSC Muscle***
		Phase_1		
Dilation_nets	*D_nets_111_1_111*	101×101	0.35 ± 0.28	0.60 ± 0.14
	D_nets_333_3_333	271×271	0.57 ± 0.29	0.76 ± 0.13
	D_nets_333_1_111	143×143	0.42 ± 0.29	0.64 ± 0.13
	D_nets_111_1_333	165×165	0.44 ± 0.29	0.70 ± 0.13
	D_nets_111_3_111	165×165	0.46 ± 0.29	0.69 ± 0.14
	D_nets_111_7_111	473×473	$\mathbf{0.56 \pm 0.28}$	$\mathbf{0.82 \pm 0.12}$
Pooling_nets	*P_nets_2MPools*	49×49	0.18 ± 0.13	0.67 ± 0.003
	P_nets_3MPools	101×101	0.25 ± 0.24	0.75 ± 0.011
	P_nets_4MPools	205×205	0.33 ± 0.3	0.79 ± 0.007
	P_nets_5MPools	445×445	0.44 ± 0.32	0.82 ± 0.010
	P_nets_2MPools(44)	101×101	0.24 ± 0.26	0.78 ± 0.006
	P_nets_2MPools(84)	149×149	$\mathbf{0.50 \pm 0.29}$	$\mathbf{0.82 \pm 0.003}$
		Phase_2		
	DP_nets_11_3_11_2MPools(48)	259×259	0.45 ± 0.28	0.75 ± 0.15
	DP_nets_111_3_111_3MPools(244)	524×524	0.56 ± 0.3	0.77 ± 0.15
	DP_nets_1111_3_1111_4MPools(2224)	540×540	0.55 ± 0.3	0.77 ± 0.13
	DP_nets_11111_3_11111_5MPools	701×701	$\mathbf{0.61 \pm 0.29}$	$\mathbf{0.77 \pm 0.15}$

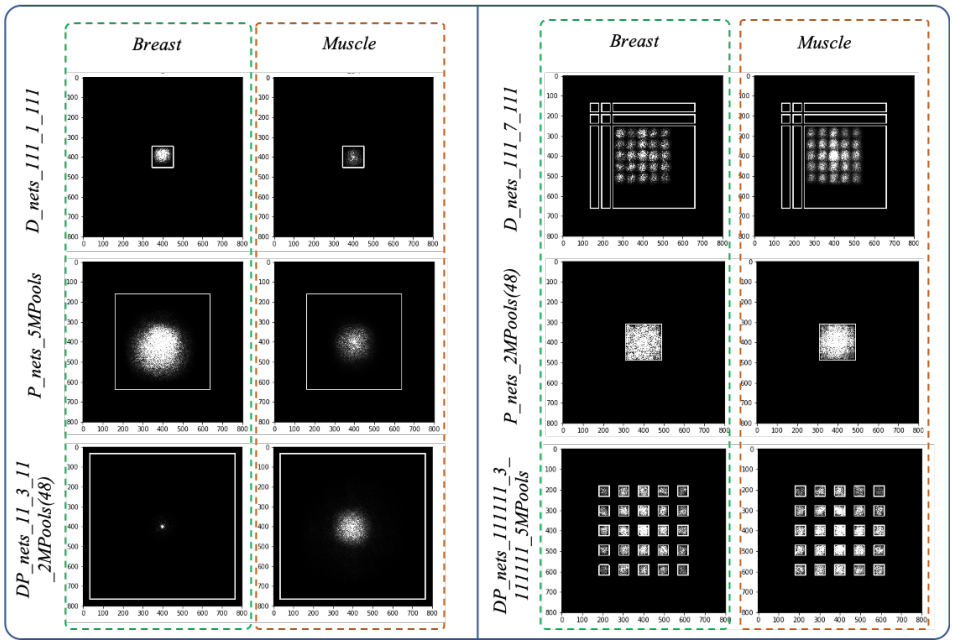


Figure 2.8: ERF and TRF of some of our proposed networks.

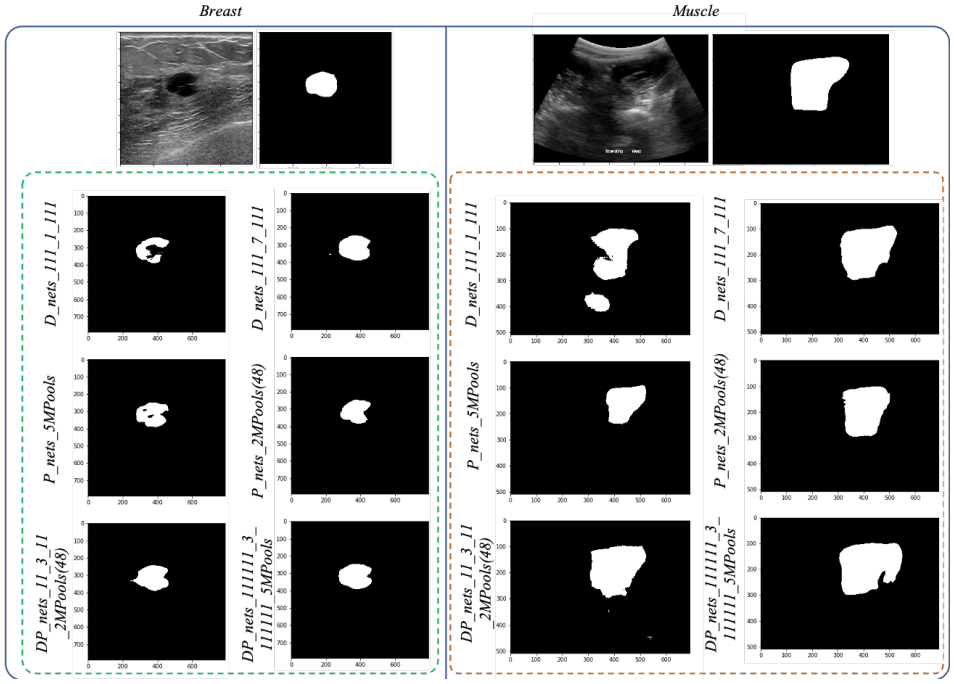


Figure 2.9: Predicted segmentation masks from some of our proposed networks.

2.5 Discussion

Among all state-of-the-art deep learning segmentation algorithms, few of them are successful for US images due to low resolution, speckle noise, and lack of enough annotations of US images. To this end, in this chapter, we investigated the two important factors that need to be taken into consideration when designing a segmentation network for US images.

The first factor is the choice of data for pre-training. In the case of US images, as mentioned earlier, access to enough number of annotations is not always available. Having this in mind, pre-training the segmentation network is a must for US images. In *Avenue_1*, we have proposed the use of simulation US images for pre-training and showed that the network can perfectly work for real phantom data without seeing any phantom data during pre-training. We have further demonstrated that for limited available *in vivo* US images (only 19 images with their annotations), pre-training on simulation phantom data is preferable to natural images when the same number of both of them for pre-training is available. Typically, it is expected that the standard deviation increases when the test DS) is low, indicating greater variability in the results. However, in this case, the standard deviation does not align with that expectation as illustrated in Table 2.6 which can be investigated in future works.

The second factor is the size of the receptive field which needs to be adjusted when designing the network. In *Avenue_2*, we investigated the importance of the dilated convolution and size of pooling layers in U-Net architecture. We showed that by adjusting these factors, we can control the size of the ERF in order to design a network computationally effective. Based on our experimental results, the receptive field is the key factor in designing U-Net-based architectures. If a network is shallow, the dilation rate and pooling kernel size should be adjusted in order to cover the sufficient receptive field. With this being said, networks should no longer be based on depth and convolutions without dilation.

Chapter 3

Automatic 3D Ultrasound Segmentation of Uterus Using Deep Learning

This chapter is based on our published paper [19].

3.1 Background

Cervical cancer as one of the most frequent cancer types in women, affects more than half a million females each year and results in 300,000 deaths worldwide [46]. It is, however, largely preventable, and the treatment is dependent on the severity of the condition and availability of local resources at the time of diagnosis [46]. Recent studies have shown that incorporating the results of advanced imaging technology and surgical staging leads to more enhanced prognosis and treatment planning [99]. Imaging modalities such as magnetic resonance imaging (MRI), computed tomography (CT), positron emission tomography (PET), and ultrasound (US) imaging have been utilized for treatment plans. However, MRI, CT, and PET imaging facilities, are costly, not uniformly available, require a long scanning time, and are not real-time. Thus, US imaging has emerged as the most suited modality for cervical cancer screening due to its cost-effectiveness, radiation-free, non-invasiveness, ease of use at the bedside, and real-time nature.

Radiotherapy is a type of treatment that delivers an effective dose of radiation to the target tissues, however, its effect and efficiency in the treatment of cervical cancer are controversial [247]. Therefore, online segmentation of the uterus can aid effective image-based

guidance for precise delivery of dose to the target tissue (the uterocervix) during cervix cancer radiotherapy. Furthermore, segmenting the uterus can aid in determining the extent of a tumor and the presence of metastatic disease. However, finding the position of the uterine boundary in US images is a challenging task due to large daily positional and shape changes in the uterus (shown in Fig. 3.1), large variation in bladder filling, and the limitations of 3D US images such as low resolution in the elevational direction. One group of studies on uterus segmentation mainly focused on developing semi-automatic algorithms that require manual initialization to be done by an expert clinician. Mason et al. [148] developed a semi-automatic algorithm such that a central sagittal plane is manually contoured. Then, the selected plane and contour are used as a starting point for fitting elliptical contours in semi-axial planes [148]. Another group focused on the use of conventional image processing techniques for automatic detection and segmentation of uterine fibroid [53, 174].

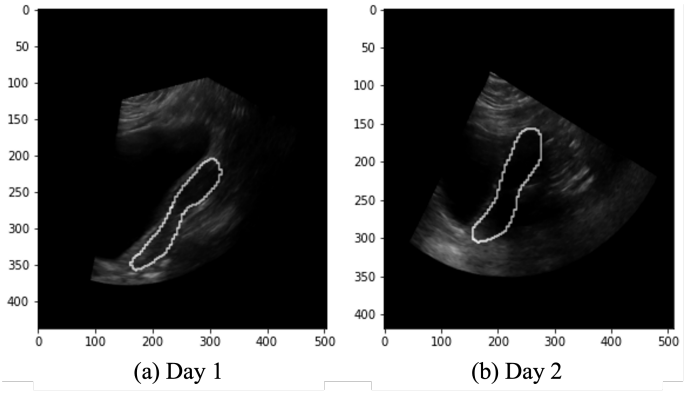


Figure 3.1: An illustration of uterine location variation (sagittal view) in one patient across two scans taken on different days.

Recent advances in image processing approaches, such as artificial intelligence (AI) and deep learning (DL) algorithms, have paved the way for solving a variety of problems. AI and deep learning approaches in medicine have a lot of potential, particularly in US diagnostic imaging, where large datasets must be managed. In US image analysis, many researchers have shown promising results in detection of breast lesions [4, 14, 18, 221], muscle [142], thyroid nodule [172], prostate[204], liver [236], and brain [211]. However, due to limited studies on the automatic segmentation of uterus US images, the main focus of the current study is to investigate more on the automatic segmentation of 3D uterus US images and to eliminate the need for manual initialization in the previous semi-automatic algorithms using the recent deep learning-based techniques. Deep learning techniques' success is heavily dependent on the amount of available data with annotations, and creating annotations for

US pictures is a time and money-intensive operation. To be more explicit, 3D networks have a higher number of parameters, which causes memory issues and a greater demand for annotated 3D data. Therefore, due to the limited available 3D uterus data, we explore 2D networks that use 2D planes of 3D volumes. Results in this chapter were published in [14].

3.2 Materials and Methods

3.2.1 Dataset

The dataset used in this chapter consists of 3D US images of 11 patients. On average, each patient went through 4 rounds of 3D US scanning, leading to a total of 38 3D US scans, with each 3D scan comprising >100 2D images. Two patients were chosen as the test set and the remainder as the train set, resulting in a total of 35 and 3 scans for the train and test sets, respectively. Table 3.1 presents the details of the number of scans for each patient. An example of US images with their overlaid annotations across all planes (i.e. axial, coronal, sagittal) is presented in Fig. 3.2. We scaled all the scans to an identical shape 576×576×576 as the 3D volumes varied in size.

Table 3.1: Number of 3D US scans per patient. Patients 1 and 10 were selected as the test set, and the rest were grouped as the train set.

Patient ID	No. 3D scans	Train/Test
1	2	Test
2	5	Train
3	3	Train
4	4	Train
5	5	Train
6	4	Train
7	5	Train
8	3	Train
9	5	Train
10	1	Test
11	1	Train

3.2.2 Protocol

Most of the recently developed deep learning algorithms suffer from generalization, and the performance of such algorithms for a new dataset needs to be investigated. Furthermore,

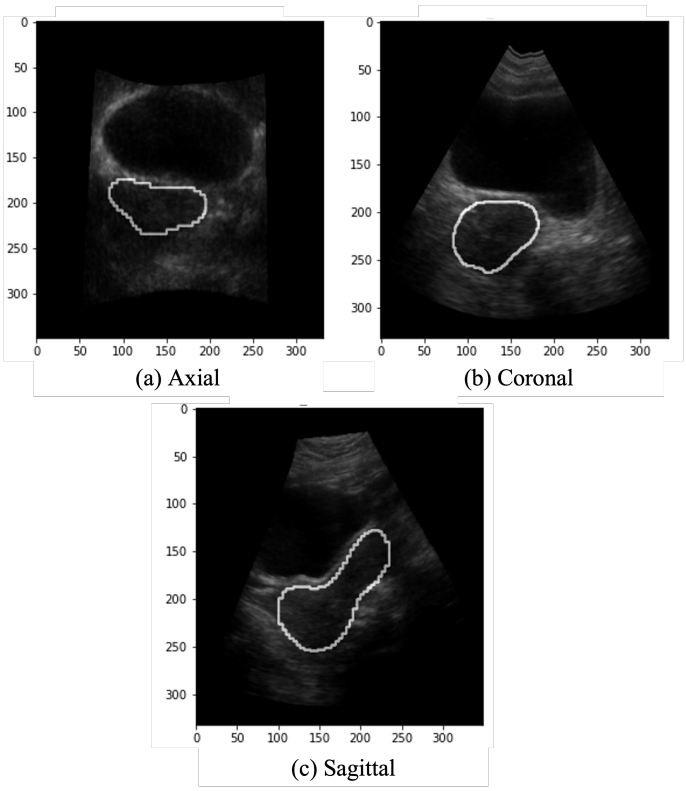


Figure 3.2: An example of a 3D US image with the uterus annotation across (a) axial, (b) coronal, and (c) sagittal planes.

training 3D networks with only 38 3D volumes is not possible. Therefore, we developed 2D networks for segmentation and stacked the outputs into a 3D volume as the final prediction. Each 3D volume is partitioned into 2D slices known as the coronal, sagittal, and axial planes. We proceeded with our analysis through two main scenarios. In the first scenario, we trained 3 different 2D networks on each 2D plane (i.e., sagittal, coronal, axial) individually. In the second scenario, our proposed 2D network was trained using 2D images across all the planes of each 3D volume.

3.2.3 Experiments

The proposed network was based on well-known segmentation architecture, U-Net [187], where its feature extractor is set to MobileNet-v2 [97]. Segmentation masks generated using the proposed algorithm were compared to expert manual contours. We had three and one network to train in the first and second scenarios, respectively. For simplicity, we refer to *net_X*, *net_Y*, *net_Z*, and *net_all* as networks trained on 2D images of axial, coronal, sagittal, and all planes. All the aforementioned networks were trained for 200 epochs, using

Adam [116] optimizer with learning rate $1e-4$ and weight decay 0.025. 2D images were reshaped to the size of 576×576 where a center crop augmentation with the cropping window size of 512×512 was applied as the augmentation. Additionally, images were flipped vertically and/or horizontally on a random basis. 5-fold cross-validation was conducted to prevent variation in network performance. The loss function was set to the combination of binary cross-entropy (BCE) and dice similarity (DSC) functions (Eq. 3.1).

$$loss = BCE + 0.5 * DSC \tag{3.1}$$

where $DSC = \frac{2|G \cap P| + ep}{|G| + |P| + ep}$, G and P are ground truth and predicted segmentation masks, respectively, and $ep = 1$. And, $BCE = ylog(P(y)) + (1-y)log(1-P(y)$, where y and P denote predictions and probability function, respectively.

3.3 Results

Figure 3.3 shows the train-validation loss across 2 networks. We only include the train-validation loss of *net_X* due to the similarity of train_validation loss in other networks (i.e., *net_Y* and *net_Z*) of our 1st scenario. We observed that when we combine all the planes of 3D volume (axial, coronal, and sagittal), Fig. 3.3 (b). Figure 3.4 (c), and (d) show an example of a sagittal slice of one patient, where the uterus is fully visible, predicted from *net_X* and *net_all* based on our first and second scenarios, respectively.

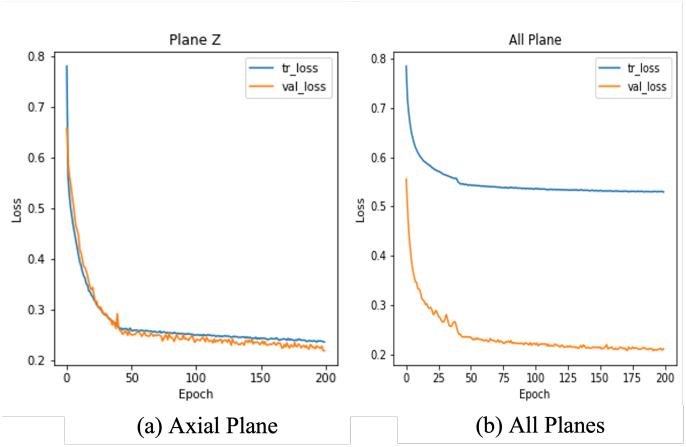


Figure 3.3: Train_validation loss for the 1st fold of 5-fold cross validation. ((a)-(c): 1st scenario, (d): 2nd scenario).

We observed that for the middle slices where the uterus is fully visible, the DSC is high for both test patients. However, for the slices close to the edges of the uterus, the DSC is

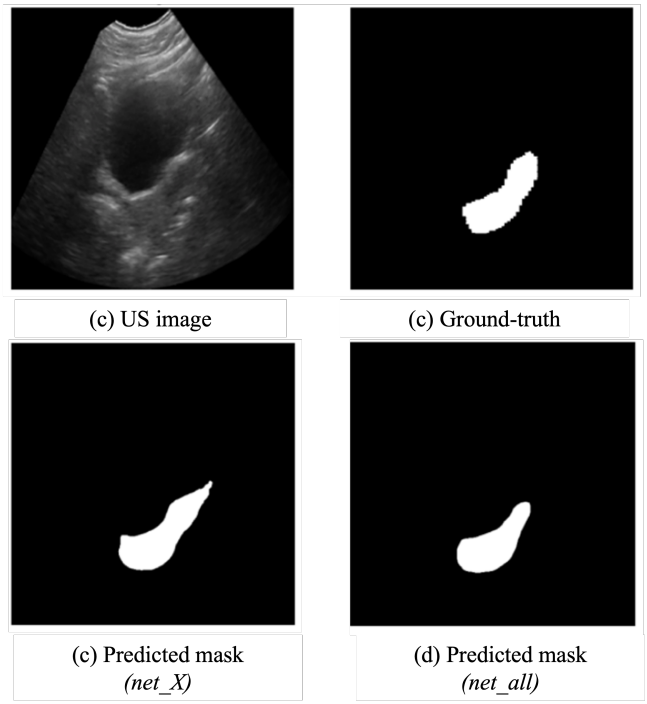


Figure 3.4: An example of ground truth versus predicted segmentation masks from *net_X* (DSC=0.88) and *net_all* (DSC=0.8) for a middle slice.

low, which means the network performs well, mainly on middle slices. The distribution of the DSC across slices in the axial plane for one scan in all 5 folds is illustrated in Fig. 3.5. The distribution of DSC in each fold is shown in (a)-(e), and the average of DSC is shown in (f). The red line in this figure shows the DSC of 0.7. Therefore, our proposed algorithms can overcome the need for manual selection of the middle slices for the semi-automatic presented in Mason et al. [148]. Some slices, however, are in the middle and have a low DSC (marked with red circles in Fig. 3.5). In the future, we will look at these cases more.

The quantitative results are reported in Table 3.2 and 3.3. The average DSC for most scans is low due to the difficulty in segmenting slices on the edges that we addressed earlier. However, we observed that the DSC of the middle slices is higher than we expected, and both scenarios behave pretty similarly.

3.4 Discussion

As mentioned earlier, uterus segmentation in US images is very challenging due to its location and inconspicuous boundaries. In the previous semi-automatic algorithm presented by Mason et al. [148], the start point of the algorithm is finding the slice where the uterus is completely

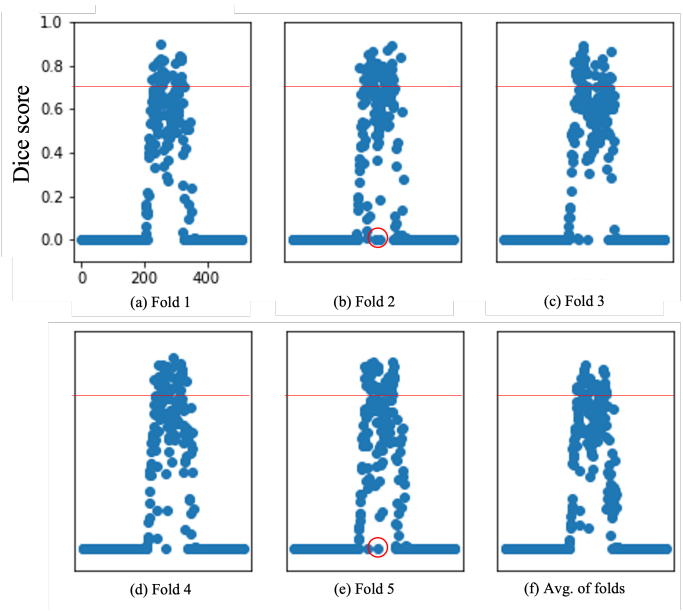


Figure 3.5: Distribution of the DSC across folds for patient ID 1.

visible. Therefore, our proposed schematic overcomes the initial manual selection of the previous semi-automatic algorithm and provides a comparable DSC with the semi-automatic algorithm. As we utilized MobileNet-v2, which is well-known in terms of being light in memory usage, the proposed network configuration is also sufficiently light, making it suitable for use in the clinic and requires results in a few seconds. In this chapter, we discovered that the proposed networks function inadequately on slices close to the uterus's boundaries.

Table 3.2: Quantitative results - Average DSC - Scenario 1.

Patient ID	**Scan No.**	***net_X***	***net_Y***	***net_Z***
All slices				
1	1	0.48 ± 0.24	0.61 ± 0.19	0.37 ± 0.15
	2	0.55 ± 0.21	0.45 ± 0.2	0.44 ± 0.17
10	1	0.58 ± 0.21	0.69 ± 0.24	0.58 ± 0.24
4 mid-slices				
1	1	0.68 ± 0.1	0.72 ± 0.05	0.56 ± 0.08
	2	0.64 ± 0.05	0.29 ± 0.04	0.46 ± 0.08
10	1	0.67 ± 0.11	0.85 ± 0.03	0.67 ± 0.11

Table 3.3: Quantitative results - Average DSC - Scenario 2.

Patient ID	**Scan No.**	**Axial**	**Coronal**	**Sagittal**
All slices				
1	1	0.55 ± 0.27	0.61 ± 0.21	0.37 ± 0.16
	2	0.53 ± 0.20	0.42 ± 0.21	0.44 ± 0.21
10	1	0.56 ± 0.22	0.64 ± 0.32	0.60 ± 0.25
4 mid-slices				
1	1	0.77 ± 0.05	0.59 ± 0.11	0.62 ± 0.08
	2	0.64 ± 0.06	0.32 ± 0.06	0.45 ± 0.06
10	1	0.63 ± 0.11	0.84 ± 0.02	0.71 ± 0.03

Chapter 4

Knowledge Distillation for Efficient Breast Ultrasound Image Segmentation: Insights and Performance Enhancement

This chapter is based on our published paper [20].

4.1 Background

Ultrasound (US) imaging is one of the most widely used medical imaging modalities, with benefits that include cost-effectiveness, non-invasiveness, portability, and real-timeliness. However, because of its low quality, the interpretation of US images necessitates professional competence, which must evolve with the variety of imaging techniques available. The delineation of an organ or region of interest (ROI) in a US image where the pixels inside the intended ROI share certain characteristics is known as US image segmentation. Image segmentation, often known as an important step in many computer-aided detection (CAD) pipelines, aids further quantitative analysis of clinical parameters related to volume and shape [134]. US image segmentation has been utilized in a variety of applications, including the creation of image atlases, determining the size or shape of the target ROI, target therapies, and performing image-guided procedures. These segmentation masks are manually delineated by an expert clinician and are considered the gold standard in medical

applications. Despite the importance of manual delineation, it is a time-consuming and labor-intensive task that is frequently subject to inter- and intra-observer variability due to differences in clinicians' experience, attention, and visual fatigue, as well as insufficient training of clinicians [222, 262]. Therefore, computerized semi- and fully-automatic segmentation techniques based on convolutional neural networks (CNN) have attracted great interest in expediting the delineation procedure while improving the reproducibility of the delineations [70, 93, 202, 239].

CNN-based algorithms have made revolutionary developments in CAD systems. However, despite the current growth of CNN-based algorithms for segmentation purposes, these techniques are rather complex, and their ability to generate satisfactory results on specific medical imaging problems is often limited [134]. Complexity in network design and configuration does not necessarily lead to better performance [109]. Furthermore, networks with large amounts of parameters that are both memory and computationally-demanding are often a hindrance to modern CNN-based segmentation approaches. To be more specific, although an increase in size is usually correlated with an improvement in representation power, it comes at a price: longer training time and more memory usage. With the present expansion of point-of-care US (POCUS) imaging equipment, it is critical to build networks that are computational and memory efficient. Compared to other imaging modalities, POCUS has the primary advantage of allowing investigations at the bedside, which is especially appealing for acutely ill patients who cannot normally be transported for such testing [157, 266]. One common use case of CAD-based systems equipped with CNN-based techniques is in POCUS imaging for breast cancer detection [10, 55, 65]. To achieve computational and memory efficiency, researchers have developed novel strategies for compressing large models so that the same or similar generalization performance can be achieved by training smaller networks [165].

In model compression techniques such as parameter pruning, quantization, and knowledge distillation (KD) to name a few, the goal is to minimize the associated computational and memory costs. In these techniques, the large model is encoded to a more efficient format with minimal performance impact [41, 231]. Parameter pruning involves training a large model and then removing unnecessary weights and parameters to get a considerably smaller yet effective model. This method also aids in the addition of regularisation to the model, resulting in improved generalization [84, 87, 123]. Pruning usually eliminates "unimportant" weights from a deployed model. This means that pruning is rarely useful for model efficiency in training and inference time; however, it can help with model storage. Similarly,

quantization-based compression techniques help in model storage by reducing the number of bits required to save the weights [72, 227, 235]. Pruning and quantization-based techniques provide a suitable compression rate without sacrificing accuracy. They are, nevertheless, better suited to applications that demand consistent model performance. As a result, KD-based techniques, also known as student-teacher networks, are better suited to applications involving small-size datasets or requiring large efficiency improvements [41]. In these approaches, the model is directly accelerated without special hardware or implementations. To be more specific, the student network (i.e., small network) is trained under the supervision of the teacher network (i.e., large network) [231]. The main idea of the KD-based approach is to transfer information from a complex teacher network into a small student network by simulating the distribution of the teacher network's representation. Previous experimental results have demonstrated that the student network can match or even beat the teacher's performance while being computationally efficient [8, 94, 186, 231].

While previous methods tend to capture rich information from various levels of teacher representation, they lack emphasis on identifying the most effective representation level. Moreover, many existing techniques propose complex strategies that pose implementation challenges. To this end, we address a gap in current KD techniques by focusing on the selection of optimal teacher representations from different levels, which has been overlooked in existing approaches. To be more specific, we study the impact of transferring knowledge from the teacher's output layer as well as from the intermediate layers of the teacher. Moreover, many existing techniques introduce complexities in selecting the appropriate teacher level for knowledge transfer [74]. In contrast, we conduct an extensive analysis of KD pathways, loss functions, and the impact of augmentation, providing valuable insights into the mechanisms underlying knowledge transfer from teacher to student networks. The proposed method simplifies the KD process by pinpointing the most beneficial teacher representation level, thus offering a more straightforward and practical solution for model compression and performance enhancement. The main contributions are summarized below:

- Highlighting the potential of leveraging teacher networks to facilitate significant performance gains in student models, indicating effective knowledge transfer.
- Developing a student network that achieves comparable performance to the teacher network while having significantly (100 times) fewer trainable parameters.
- Exploring the fundamental role of augmentation techniques and loss functions in facilitating knowledge transfer across different distillation pathways.

The rest of the chapter is structured as folows: Sec. 4.2 provides an in-depth literature review, while Sec. 4.3 outlines our proposed methodology. Our achievements and results are presented in Sec. 4.5, and concluding remarks are presented in Sec. 4.6.

4.2 Related Work

In this section, we present an extensive review encompassing previous methodologies for network compression based on KD, alongside an analysis of US segmentation techniques, specifically those employing KD methodologies. Please note that the structure of this section is designed to review KD studies utilizing both natural and medical image datasets. Additionally, since we are using a publicly available dataset introduced by Yap et al. [250] (i.e. *Dataset_A*), we include a review of recent studies that have employed this dataset, regardless of whether they used KD as their main methodology. Our aim is to compare our results with those of other studies that used the same dataset.

4.2.1 Studies on KD

KD-based techniques have been used in both classification and segmentation tasks [92, 94, 136, 224, 230, 244, 256]. The main idea of these approaches is to distill knowledge from the output probabilities with rich information of the teacher network to the student network. Xu et al. [244] focused on matching the distribution of logits while Zagoruyko and Komodakis [256] transferred knowledge from intermediate features. Tung and Mori [224] proposed the distillation of similarity-preserved knowledge such that the student network can preserve the pairwise similarities of paired inputs that provide similar activation maps from the teacher network. He et al. [92] developed a KD method for semantic segmentation that minimizes the inconsistency between student and teacher knowledge. Another KD-based strategy on semantic segmentation proposed by Liu et al. [136] performed structure distillation in pairwise and holistic distillation schemes.

4.2.2 Studies on KD in Medical Images

Recently, researchers have adopted KD-based techniques for various applications in medical imaging [39, 59, 67, 95, 128, 146, 179], and specifically in US imaging [32, 125, 173, 228]. Owen et al. [173] explored the efficacy of a student-teacher framework in training lightweight deep learning models, using unlabeled data to achieve fast, automated detection of abnor-

mality in optical coherence tomography B-scans. Vaze et al. [228] introduced a methodology for modifying and compressing the original U-Net model [188] while incorporating KD to ensure that the performance of the compressed model closely matches that of the original U-Net on 5635 US images. Cao et al. [32] proposed a noise filter network (NF-Net) that mitigates the negative impact of noisy labels through the incorporation of two softmax layers for classification and a teacher-student module for distilling the knowledge of clean labels in the classification of breast tumors. Fan et al. [61] introduces optimization trajectory distillation, a novel approach using a dual-stream distillation algorithm for unsupervised domain adaptation.

Table 4.1 reviews the key features of the aforementioned studies. Since the generalizability of the works discussed in Sec. 4.2.1 remains untested in the medical image domain, these studies are excluded from Table 4.1. In Table 4.1 most papers either utilize the output layer or the intermediate layers for distillation, and none investigate both simultaneously. Transferring knowledge solely from the logits can lead to a performance gap between teacher and student models. Each paper employs unique distillation losses, yet none explores the impact of these losses on the distillation process. By taking the L1-norm of all layers, knowledge transfer is ensured throughout the entire network, promoting more comprehensive learning.

Table 4.1: Key points of previous works using KD in medical images.

Article	Dataset	Task	Knowledge Distillation Method
Owen et al. [173]	Optical Coherence Tomography	Classification	From model logits using Binary Cross-Entropy
Vaze et al. [228]	Nerve US	Segmentation	From all the layers using L1-norm
Cao et al. [32]	Breast US	Classification	From model logits using squared error
Fan et al. [61]	Multiple[1]	Multi-task[2]	From gradients of one domain to another

[1] Multiple datasets used. For more details, please refer to Fan et al. [61], [2] Multiple tasks, including segmentation, classification, etc.

4.2.3 Studies on *Dataset_A*

In this work, as we utilize a publicly available 2D US dataset introduced by Yap et al. [250], we conduct a review of publications that have employed the same dataset to ensure a fair comparison of our segmentation results with existing works. It is worth noting that we employ the *Dataset_A* as explained in Yap et al. [250], and we maintain consistency in our terminology throughout this manuscript by referring to this dataset accordingly (please refer to Sec. 4.3.1 for more details on the dataset). Using *Dataset_A*, Yap et al. [251]

proposed an end-to-end approach for US lesion detection and recognition by utilizing a pre-trained segmentation network designed based on fully convolutional networks (FCN) [140] and achieved Dice similarity coefficient (DSC) score of 55%. Abraham and Khan [1] proposed generalized focal loss based on the Tversky index for the attention UNet and achieved a DSC of 80%. They achieved a DSC of 66% for the UNet model with focal Tversky loss. Zhuang et al. [265] proposed a Residual-Dilated-Attention-Gate-UNet (RDAU-Net) model obtaining a DSC of 85% and reported a DSC of 82% for UNet model. Costa et al. [47] proposed FCN-based segmentation models and reported a DSC of 82%. Liang et al. [131] developed a multi-stage elastic augmentation technique and achieved a DSC of 84% using a Mask-RCNN-based segmentation network [90].

Amiri et al. [4] developed two-stage segmentation UNet to first detect the tumor region and then segment the detected region. They reported a DSC of 86%. Lee et al. [124] proposed an attention module and obtained a DSC of 76%. Shareef et al. [199] proposed Small Tumor-Aware Network (STAN) that involved CNN layers with various kernel sizes in order to extract multi-scale information from US images. They achieved a DSC of 78%. In their next study [198], they improved their work by proposing an Enhanced STAN network and achieved a DSC of 82%. Singh et al. [207] proposed a contextual information-aware network based on conditional generative adversarial networks (cGAN) [155] that integrates atrous-convolution [38], channel attention [64] along with channel weighting [100]. They obtained a DSC of 86%. Hussain et al. [104] proposed a combination of deep learning-(i.e. UNet network) and a traditional learning-based algorithm (i.e. level-set framework) as their proposed methodology. They reported only the DSC of 98% and 72% for benign and malignant tumors. Qu et al. [180] introduced an attention-supervised full-resolution residual network (ASFRRN) inspired from full-resolution residual networks (FRRN) [177] and achieved DSC of 84%. Ning et al. [166] achieved a DSC of 85% from their proposed coarse-to-fine fusion network alongside a weighted-balanced loss function. In one of our previous works [15], we explored the different pre-training strategies for training a UNet when only 20 images were used for training and obtained a maximum DSC of 57%.

Gao *et al.* [66] investigated class imbalance in segmentation by proposing their multi-scale fused network with additive channel–spatial attention and achieved a DSC of 85%. Su et al. [216] proposed a multi-scale UNet that involves layers with different receptive fields and led to a DSC of 82%. Xu et al. [245] introduced a multi-scale self-attention network by integrating local features and global contextual information that led to a DSC of 83%. Huang et al. [102] proposed different approaches for transfer learning. In one of their experiments,

Table 4.2: Summary of previous works and their reported DSC scores (%) on *Dataset_A*.

Article	Method	DSC (%)
Yap et al. [251]	Pre-trained model	55
Abraham and Khan [1]	Tversky focal loss	80
Zhuang et al. [265]	RDAU-Net	85
Costa et al. [47]	FCN-based model	82
Liang et al. [131]	Multi-stage AUG[1]	84
Amiri et al. [4]	Two-stage UNet	86
Lee et al. [124]	Attention module	76
Shareef et al. [199]	STAN model	78
Shareef et al. [198]	ESTAN model	82
Singh et al. [207]	cGAN-based model	86
Hussain et al. [104]	DL+LS[2] framework	98 (B)[3] 72 (M)[4]
Qu et al. [180]	ASFRRN model	84
Ning et al. [166]	Coarse-to-fine fusion	85
Behboodi et al. [15]	Pre-trained model	57
Gao et al. [66]	MS[5] fused model	85
Su et al. [216]	MS[5] UNet	82
Xu et al. [245]	MS[5] self-attention model	83
Huang et al. [102]	Transfer learning	83
Yeung et al. [253]	Unified focal loss	82
Xu et al. [243]	Adaptive RF[6] model	88
Lou et al. [143]	IRPB+CFB modules	90
Yang et al. [248]	CTG-Net	79
Lee et al. [125]	TTFT KD-based	89

[1]Augmentation, [2]Deep learning+level-set, [3]Benign,
[4]Malignant, [5]Multi-scale, [6]Receptive field

first, they pre-trained various networks on Achilles tendon US images, and then fine-tuned on breast US images (i.e. *Dataset_A*) and reported the best DSC of 83%. A unified-focal loss was introduced by Yeung et al. [253] achieving a DSC of 82%. Xu et al. [243] reported a DSC of 88% for their proposed adaptive receptive field network. Lou et al. [143] achieved a DSC of 90% by introducing inverted residual pyramid block (IRPB) and context-aware fusion block (CFB) modules to UNet architecture. Yang et al. [248] introduced CTG-Net that integrates lesion segmentation and tumor classification tasks in breast US image analysis, achieving improved performance compared to existing multi-task learning approaches. Table 4.2 summarizes previous works that utilized *Dataset_A*.

4.3 Proposed Method

This section describes the proposed KD-based method that highlights the potential of leveraging teacher networks to facilitate significant performance gains in student models. We develop a student network that achieves comparable performance to the teacher network while having a remarkable 100 times fewer trainable parameters. Additionally, we explore the fundamental role of augmentation techniques and loss functions in facilitating knowledge transfer across different distillation pathways, providing new insights into the optimization of model distillation processes.

4.3.1 Dataset

As mentioned earlier, we used a publicly available 2D US dataset introduced by Yap et al. [250] referred to as *Dataset_A*. It consists of 163 breast US images with their manual delineations, each presenting either cancerous masses or benign lesions with a mean image size of 760×570 pixels. In our experiments, we created three random splits of 130 images for the train-validation set and 33 images for testing.

4.3.2 Teacher and Student Models

In our segmentation framework, we employed a U-Net-based architecture [188] for both student and teacher networks. After conducting extensive experimentation with various backbone architectures including ResNet34, ResNet101, ResNext50, and ResNext101, we ultimately selected ResNeXt101 [241] as the backbone for our teacher model. Among the options tested, ResNeXt101, with 96 million parameters, outperformed others, demonstrating superior performance in terms of accuracy and robustness. For our student model, we modified MobileNetV3-small-100 [98] in a way that only had 0.82 million parameters. MobileNetV3-small-100 stood out as the sole model with a significantly reduced parameter count while still possessing pre-trained weights on the ImageNet dataset. This characteristic was pivotal for our choice, as it allowed us to strike a balance between model complexity and computational efficiency, making it the most suitable candidate to serve as the student model in our knowledge distillation process. By leveraging distinct encoders tailored to computational requirements, our approach aims to optimize distilling knowledge from teacher to student, achieving a favorable balance between model complexity and performance in our proposed segmentation framework. For both teacher and student models, we initialized the backbone with pre-trained weights obtained from ImageNet [51].

4.3.3 Knowledge Distillation (KD) Paths

In our proposed knowledge distillation (KD) paths, we explore two distinct strategies by distilling knowledge either from the final predictions or from the hidden representations, i.e., the output of the teacher model's encoder. This approach allows us to examine the most relevant source of knowledge transfer. By incorporating these alternative pathways, we ensure that the student model can effectively learn from either the teacher's output or the intricate feature representations captured by its encoder. Illustrated in Fig. 4.1, our approach delineates three primary pathways for distilling knowledge from the teacher to the student model: *KD-Logits (L)*, *KD-Hidden (H)*, and *KD-HiddenRegressor (HR)*

KD-Logits (L)

In this particular distillation pathway, our objective is to transfer knowledge in the form of final predicted logits from the teacher to the student model. Logits represent the raw predictions produced by the teacher model before applying any activation function, offering a view of the model's confidence scores across different classes or categories. By distilling these logits, the student model gains access to valuable information regarding the teacher's level of uncertainty, enabling a more nuanced optimization process than can be achieved with only a binary training signal. The design of *KD-Logits (L)* is shown in Fig. 4.1-(a).

KD-Hidden (H)

In this designated pathway, as illustrated in Fig. 4.1-(b), we aim to distill knowledge from the output of the teacher's encoder, specifically focusing on the hidden features. These hidden features encapsulate rich representations of the input data captured at various levels of abstraction within the teacher's architecture. However, a challenge arises due to discrepancies in the dimensions of the hidden features between the teacher and student models. To address this, we employ a strategy to adjust the size of the hidden features to ensure compatibility between the two models. Specifically, since the number of channels in the teacher's hidden features may differ from that of the student's, we harmonize their dimensions by taking the average over the channels (denoted as K in Fig. 4.1-(b)) from both sets of hidden features. This normalization process facilitates a seamless transfer of knowledge, aligning the representations from both models and enabling effective learning by the student. Moreover, by leveraging this method, we ensure that the student model can benefit from the comprehensive insights encoded within the teacher's hidden features. The hidden feature size of the

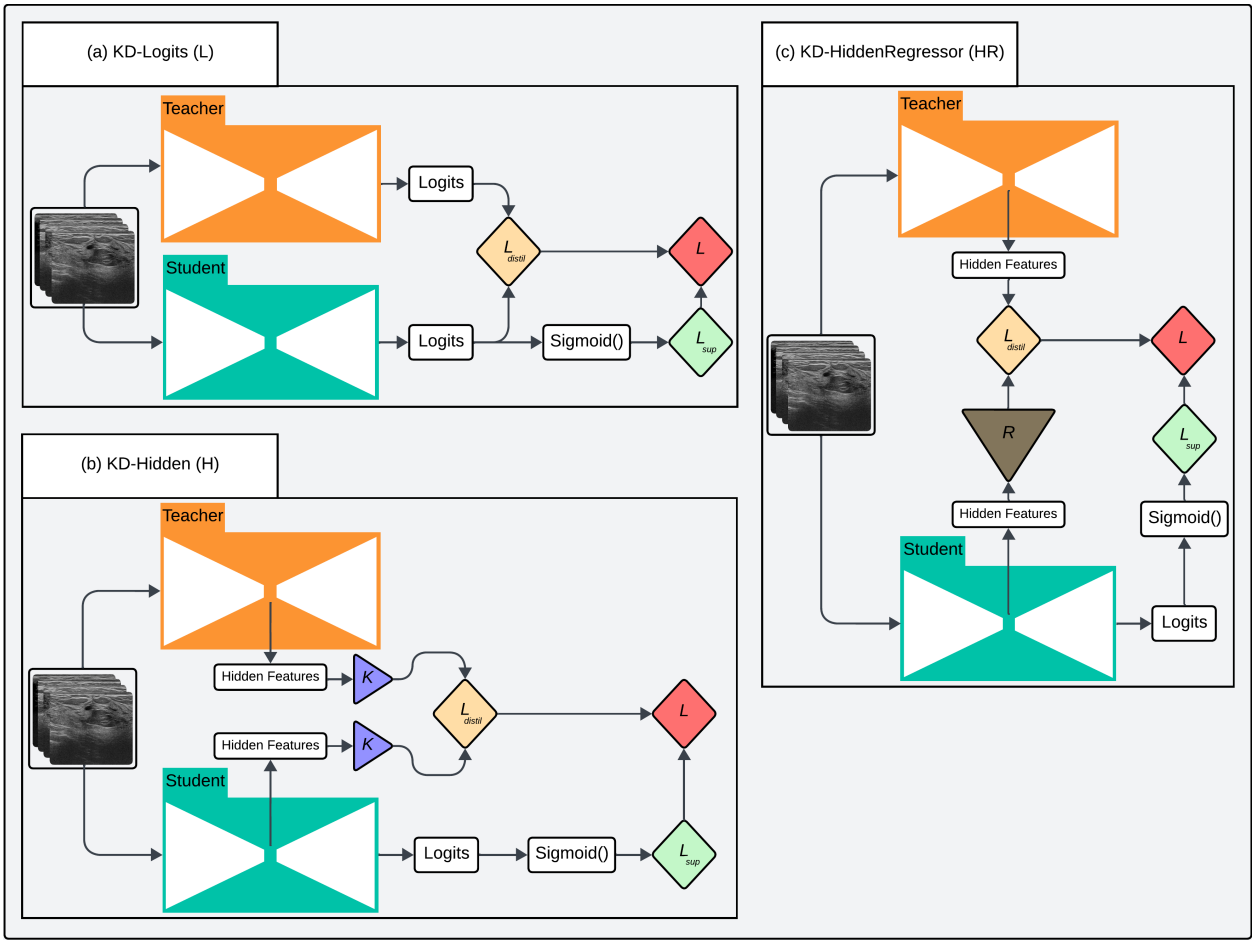


Figure 4.1: The proposed KD paths. (a) Transfer of knowledge from the output layer (logits) of the teacher network. (b) Transfer of knowledge from the hidden representations of the teacher network. The hidden representations of the teacher and student networks differ only in the number of channels. Hence, to align the shapes, an average over the channels (K) is computed. (c) Similar to (b), knowledge is transferred from hidden representations, but to match the shape, a CNN-based regressor (R) is employed.

teacher and student are denoted as $B \times C_t \times H \times W$ and $B \times C_t \times H \times W$, respectively, where B is the batch size, C_* is the number of channels in teacher (C_t) and student (C_s) models, H is the height, and W is the width.

KD-HiddenRegressor (HR)

As demonstrated in Fig. 4.1-(c), in this pathway, akin to the previous approach, our objective remains to transfer knowledge from the hidden features of the teacher model to the student. However, in this instance, to mitigate the challenge of mismatching the hidden feature maps, we introduce a novel regressor model (denoted as R in Fig. 4.1-(c)) compris-

ing two convolutional layers followed by a Rectified Linear Unit (ReLU) activation function. This regressor model is strategically designed to adjust the number of channels in the student's feature map to match that of the teacher. By incorporating this regressor model into the distillation process, we effectively bridge the gap between the differing feature map sizes. The convolutional layers within the regressor model learn to map the student's feature representations onto a higher-dimensional space, aligning them with the richer feature representations of the teacher model. The subsequent ReLU activation function introduces non-linearity, facilitating the extraction of complex patterns and enhancing the fidelity of the knowledge transfer process.

4.3.4 Loss Functions

The training loss (L), defined in Eq. 4.1, is the combination of two key components: the distillation loss ($L_{distill}$) and the supervised loss (L_{sup}). The distillation loss aims to optimize the transfer of knowledge from the teacher to the student model, leveraging the insights encoded within the teacher's representations to refine the student's internal representation. On the other hand, the supervised loss allows the student model to learn directly from the ground-truth labels, aligning its predictions with the true distribution of the training data. By combining these two losses, the training process balances the richer signal coming from the internal representation of the much larger teacher network against the supervision signal from the actual task.

$$L = 0.5 \times L_{distill} + 0.5 \times L_{sup}, \tag{4.1}$$

Please note that in the following loss functions N, i, y_i, $\hat{p}_i$, $\hat{y}_i$, $P(i)$, and $Q(i)$ respectively refer to total number of samples, one sample, ground-truth label, predicted values, predicted label, teacher's representation, and student's representation. The predicted label is the rounded predicted value.

Knowledge Distillation (KD) Loss ($L_{distill}$)

To evaluate the effectiveness of loss functions in distillation, we have employed two commonly used metrics: the Mean Squared Error (MSE) and the Kullback-Leibler Divergence (KLD).

The MSE loss measures the average squared difference between the predicted and target values. It is defined as:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(P(i) - Q(i))^2. \tag{4.2}$$

The KLD loss quantifies the divergence between two probability distributions, in our case, teacher and student representation distributions, defined as:

$$KLD = \sum_{i=1}^{N} P(i) \log\left(\frac{P(i)}{Q(i)}\right). \tag{4.3}$$

Supervised Loss (L_{sup})

For the supervised loss, we have used cross-entropy (CE). CE measures the dissimilarity between the predicted pixel-wise probability distribution and the ground truth labels. The formula for cross-entropy is given by:

$$CE = -\frac{1}{N} \sum_{i=1}^{N} y_i \log(\hat{p}_i). \tag{4.4}$$

4.3.5 Augmentation

In our experimental setup, we separated our approach into two distinct strategies regarding data augmentation. In one set of experiments, we employed weak augmentation solely for the teacher model while implementing strong augmentation exclusively for the student. In another series of experiments, we applied strong augmentation to both teacher and student models. This difference in augmentation strategies was purposefully designed to explore and assess the impact of differential augmentation levels on the performance and robustness of the resulting models. By varying the augmentation schemes, we aimed to gain insights into how each model responds to different levels of data perturbation and how this influences their learning dynamics and generalization abilities.

By isolating weak augmentation for the teacher and strong augmentation for the student, we sought to emphasize the role of the teacher as a stable source of distilled knowledge, while allowing the student to leverage augmented data for enhanced generalization. Conversely, employing strong augmentation for both models aimed to evaluate the effectiveness of augmenting data at both stages of the distillation process, potentially leading to further improvements in model performance through increased exposure to diverse training instances.

Table 4.3: Summary of experimental factors in the tested knowledge distillation approaches

KD Representation	Model Notations	KD Loss	Teacher Augmentation
KD (Logits)	*L_MSE*	Mean Squared Error	Strong
	L_MSE_WAug	Mean Squared Error	Weak
	L_KLD	Kullback-Leibler Divergence	Strong
	L_KLD_WAug	Kullback-Leibler Divergence	Weak
KD (Hidden)	*H_MSE*	Mean Squared Error	Strong
	H_MSE_WAug	Mean Squared Error	Weak
	H_KLD	Kullback-Leibler Divergence	Strong
	H_KLD_WAug	Kullback-Leibler Divergence	Weak
KD (Hidden-Regressor)	*HReg_MSE*	Mean Squared Error	Strong
	HReg_MSE_WAug	Mean Squared Error	Weak
	HReg_KLD	Kullback-Leibler Divergence	Strong
	HReg_KLD_WAug	Kullback-Leibler Divergence	Weak

Weak Augmentation

Our weak augmentation strategy employed a conservative approach tailored to the teacher model's training. Only random cropping is applied before normalization to the ImageNet mean and standard deviation. Since we utilized pre-trained networks as our base architecture, adhering to these standard normalization procedures helped maintain consistency with the pre-existing feature representations.

Strong Augmentation

Our strong augmentation strategy employed a more aggressive set of techniques. Random cropping, shift, scale, rotation, Gaussian noise injection, and pixel dropout, were all applied before normalization to the ImageNet space.

4.3.6 Experimental Overview

As we have highlighted, our research delves into several crucial factors within the realm of knowledge distillation, including different distillation paths, loss functions, and augmentation strategies. We have summarized these experiments in Table 4.3, employing distinct notations for each experiment to enhance clarity and comprehension.

4.4 Experiments

In our experiments, we utilized an Nvidia Titan XP GPU running on Ubuntu 20.04.6 LTS as our computational platform. We employed PyTorch as our primary framework for model implementation. For building our models, we relied on Segmentation Models [105], a PyTorch-based library. Additionally, we utilized Albumentations [28], another PyTorch-based library, for implementing augmentation techniques.

In all our experiments, we standardized the image size to 224×224. Given the limited size of our training dataset, we used 3-fold cross-validation to showcase the generalizability of the models. Initially, we trained both the teacher and student models separately to establish our baseline performance. Throughout this manuscript, "Teacher" (with a capital T) refers to the predictions of the selected teacher model, while "Student" (with a capital S) refers to the predictions of the student model trained from the dataset alone. All KD-based supervised student models are referred to by their "Model Notation" as defined in Table 4.3.

For optimization, we employed the Adam optimizer with a learning rate of 10^{-4}. The batch size was set to 64, and training was conducted for 500 epochs. To prevent overfitting, we implemented early stopping, and terminated training if the validation loss failed to improve for 50 consecutive epochs. Model checkpoints were saved based on the best validation loss achieved during training. For performance evaluation, we utilized the DSC that quantifies the overlap between the predicted and ground truth segmentation masks and captures both the precision and recall aspects of model performance. DSC is calculated as follows:

$$DSC = \frac{2 \times \sum_{i=1}^{N} (y_i \times \hat{y}_i)}{\sum_{i=1}^{N} (y + \hat{y})}, \tag{4.5}$$

where N, i, y_i, and $\hat{y}_i$, represent the total number of samples, one sample, ground-truth label, and predicted label, respectively.

4.5 Results

In this section, we delve into the obtained outcomes and analyze the implications of each suggested KD pathway. Additionally, we assess the effect of MSE and KLD loss functions on knowledge transfer from teacher to student. Furthermore, we examine the impact of augmentation on the teacher model. Finally, we compare the best-performing KD paths with SOTA methods that utilized the same dataset. It is worth noting that none of these SOTA methods have publicly disclosed their training and testing splits, nor have they shared

their codes. As a result, we can only report the results as presented in their respective papers.

The Dice similarity scores, averaged over a 3-fold cross-validation, are summarized in Table 4. The proposed *L_KLD_WAug* model achieved the highest DSC of 80.00 ± 18.86. Please note that the statistical analysis in Table 4 illustrates the significance of the differences between the DSC of the proposed models compared to *Teacher* and *Student*. For instance, the DSC for the *L_KLD_WAug* model does not significantly differ from the *Teacher* model, with a p-value of 0.2075. This validates that the *L_KLD_WAug* model performs similarly to the *Teacher* model. Conversely, it is significantly outperforming the *Student* model, with a p-value of 0.0012.

4.5.1 Ablation Study

In this section, we conduct an ablation study and analyze the results from various perspectives. As presented in Table 4.4, the proposed KD paths consistently exhibit performance closely aligned with that of the teacher model. This is evidenced by their p-values relative to the teacher's predictions, which generally do not demonstrate significant differences. However, these KD paths typically reveal significantly better performance when compared to the student model, as indicated by their respective p-values. A visualization example of our ablation study is presented in Fig. 4.2.

Effect of KD Paths

By investigating the performance evaluation of the KD paths, it becomes evident that each pathway showcases noteworthy achievements in enhancing student performance. In the KD (Logits) path, where knowledge is transferred between the logits of the teacher and student, the highest DSC of 80.00 was attained by the *L_KLD_WAug* model. Moving to KD (Hidden), which involves exchanging knowledge between hidden features, the top DSC of 79.00 was achieved by the *H_KLD* model. Similarly, in KD (Hidden-Regressor), where knowledge passes from hidden features through a regressor model, the highest DSC of 79.00 was reached by the *HReg_KLD* model. These findings collectively suggest that all proposed KD paths exhibit comparable performance, enhancing student performance by approximately 9%. Such consistent enhancements underscore the robustness and versatility of the proposed KD paths, demonstrating their effectiveness in knowledge exchange between teacher and student.

Effect of KD Loss Function

Further analysis of the results presented in Table 4.4 shows that both the MSE and KLD loss functions are effective for knowledge transfer. Notably, across various KD pathways, including KD (Logits), KD (Hidden), and KD (Hidden-Regressor), the DSC reveals a consistent pattern wherein both loss functions demonstrate similar effectiveness. In the KD (Logits) pathway, for instance, the DSC achieved by *L_MSE* and *L_KLD*, namely 77.75 and 78.00, respectively, highlight the marginal outperformance of *L_KLD*. Similarly, in other KD pathways such as KD (Hidden) and KD (Hidden-Regressor), the comparative analysis reveals a similar pattern between MSE and KLD. This slight outperformance of KLD in average DSC scores suggests that KLD can be a preferable choice, indicating that the selection between MSE and KLD loss functions could notably influence the effectiveness of knowledge transfer in knowledge distillation.

Effect of Augmentation

Exploring the impact of weak augmentation on the teacher model reveals more insights into the knowledge distillation process. The utilization of weak augmentation for the teacher did not yield a significant impact on performance. Models with and without weak augmentation for the teacher demonstrated comparable performance. Despite the negligible effect of weak augmentation on the teacher model, all models incorporating teacher guidance showcased improvements compared to students without such supervision. This observation demonstrates the fundamental role of the teacher network in guiding and enhancing the learning process of the student network. While weak augmentation may not directly influence the performance of the teacher model, its presence facilitates the extraction and transfer of valuable knowledge, thereby contributing to the overall improvement in student performance.

4.5.2 Results with Respect to SOTA Methods

In this section, we compare our best model with SOTA models, as outlined in Table 4.5, which have utilized the same dataset employed in our study. It is important to emphasize that none of these SOTA models have provided access to either their codebase or their training and testing splits. Consequently, our comparison is based solely on the results reported in their respective papers. Please note that Table 4.5 exclusively showcases our best model alongside the top 3 SOTA models that have reported the number of trainable parameters in their corresponding paper.

Table 4.4: Experimental Dice similarity scores (DSC): average over 3-fold cross-validation.

Model Notations	**DSC (%)** (mean±std)	***p*-value w.r.t[1]**	
		Teacher	***Student***
Teacher	81.50±18.40	-	0.0048**
Student	73.16±23.78	0.0048**	-
L_MSE	77.75±22.55	0.1414	0.0227*
L_MSE_WAug	77.52±17.61	0.1565	0.0948
L_KLD	78.00±22.05	0.0950	0.0037**
L_KLD_WAug	**80.00±18.86**	0.2075	0.0012**
H_MSE	77.50±21.49	0.0645	0.0215*
H_MSE_WAug	77.85±20.06	0.0341*	0.0125*
H_KLD	79.00±20.45	0.1320	0.0067**
H_KLD_WAug	78.50±19.83	0.1417	0.0316*
HReg_MSE	78.06±19.25	0.0325*	0.0201*
HReg_MSE_WAug	77.63±19.97	0.0402*	0.0269*
HReg_KLD	79.00±20.37	0.2132	0.0021**
HReg_KLD_WAug	79.31±19.33	0.1247	0.0068**

[1]w.r.t.: with respect to. * and ** denote a statistically significant difference with p-value < 0.05 and p-value < 0.01, respectively.

Our proposed best model demonstrates comparable performance to SOTA models, despite having significantly fewer trainable parameters. This observation highlights the efficiency of our model architecture in achieving competitive results while keeping the number of parameters minimal. By leveraging innovative design in distilling knowledge, our model strikes a balance between computational complexity and performance, making it well-suited for resource-constrained environments or applications where model size is a critical consideration.

Table 4.5: Results w.r.t SOTA methods.

Article	**DSC (%)**	**No. of Params (M[1])**
Liang et al. [131]	84	20.5
Gao et al. [66]	85	2.34
Lou et al. [143]	90	26.63
Lee et al. [125]	89	7.7
Ours (*L_KLD_WAug*)	**80**	**0.82**

[1]Millions.

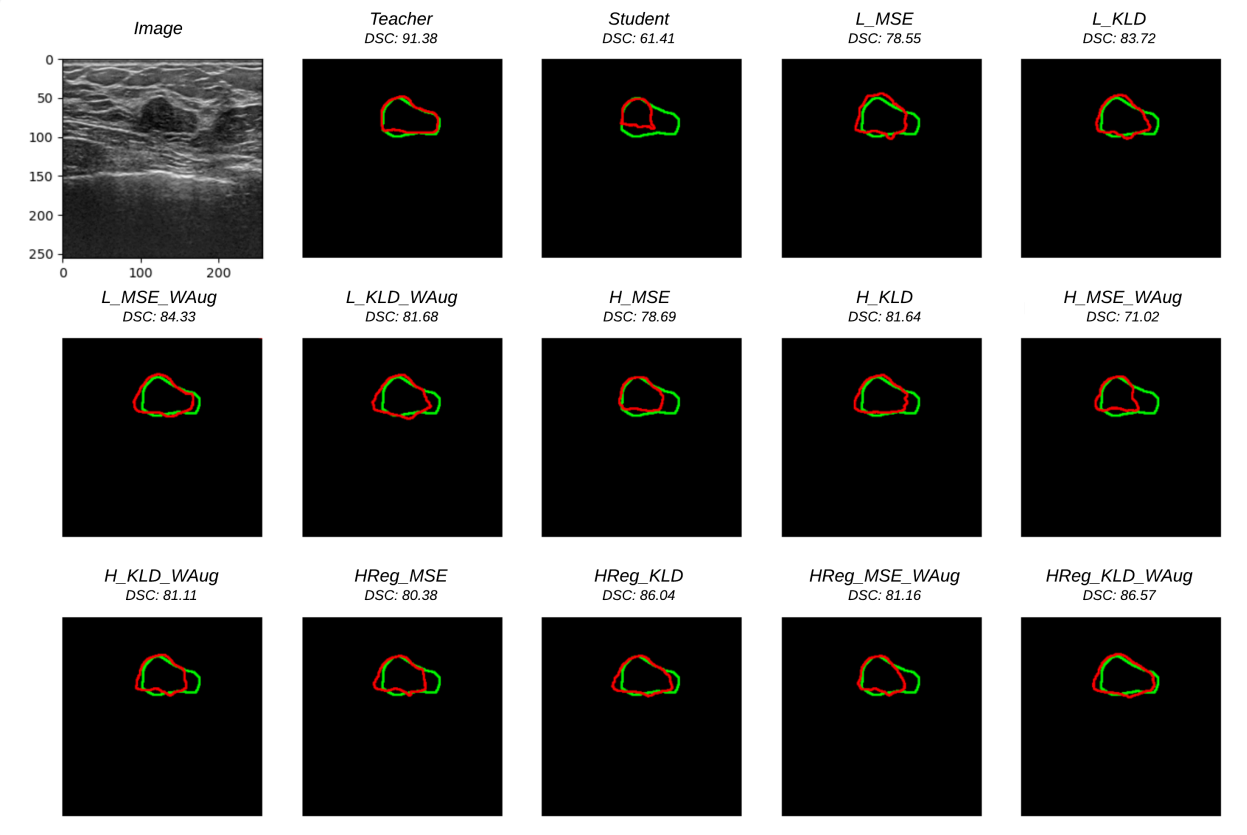


Figure 4.2: Visual comparison of our ablation study. The original test image, the prediction of the teacher model, and the prediction of the unsupervised student model are shown in (a), (b), and (c), respectively. The predicted segmentations of the proposed KD-based models are shown in (d)-(o). Green contours represent the ground truth mask, while the red contours illustrate the corresponding predictions.

4.6 Discussion and Conclusion

In this study, we investigated various aspects of knowledge distillation techniques and their implications for enhancing student performance. Through an extensive analysis of KD pathways, loss functions, and the impact of augmentation, we gained valuable insights into the mechanisms underlying knowledge transfer from teacher to student networks. Our findings revealed that the proposed KD paths consistently demonstrated performance closely aligned with that of the teacher model, indicating effective knowledge transfer. Additionally, the comparative analysis between MSE and KLD loss functions showed comparable efficacy in facilitating knowledge transfer across different KD pathways. Furthermore, exploring the impact of different augmentations on the teacher model showed the fundamental role of teacher guidance in improving student performance, despite the negligible effect of augmentation on the teacher model itself.

Finally, our comparison with SOTA models showcased the efficiency of our proposed model architecture. Despite having significantly fewer trainable parameters, our best model demonstrated comparable performance to SOTA models, highlighting the effectiveness of our approach in achieving competitive results while minimizing model complexity. Therefore, by leveraging the rich knowledge encapsulated within the teacher network, students can effectively learn from the expertise encoded in the teacher's parameters, leading to significant performance gains. Such endeavors hold promise for advancing the state-of-the-art in model compression and facilitating the deployment of efficient deep learning solutions across various domains and applications.

Even though our study provides valuable insights, it would be advantageous to explore various student models with differing numbers of trainable parameters to assess the trend of their performance relative to parameter count. This investigation would offer a deeper understanding of the scalability and efficiency of the proposed KD-based framework. Furthermore, expanding our research to encompass additional publicly available US datasets with diverse applications would improve the generalizability and robustness of the proposed framework.

Chapter 5

Deep Classification of Breast Cancer in Ultrasound Images: More Classes, Better Results with Multi-task Learning

This chapter is based on our published paper [17]. It is noteworthy that our paper was selected as one of the finalists for the SPIE Medical Imaging 2021 Robert F. Wagner Award.

5.1 Background

Image classification refers to producing an output classification label given an input image. In the domain area of medical image analysis, one of the imperative problems is identifying whether a disease is present or not. Breast cancer is one of the most common leading causes of death among women worldwide. For screening breast cancers, mammography is usually used as the primary imaging modality. However, mammography fails to distinguish dense and cancerous tissues. Additionally, it utilizes harmful radiation which makes it impractical for patients with certain conditions such as pregnancy. Due to its high rate of false positives, unnecessary biopsies might be required [37]. Therefore, US has become one of the most important alternatives for breast cancer detection as well as screening [40]. However, US imaging is operator-dependent and requires well-trained and experienced radiologists

[37]. Therefore, automatic image analysis approaches play the most important role in US classification.

5.2 Related Works

In previous methods of breast US classification, US images were either used in full size or divided into subregions, and then texture-related features were manually extracted and input into a classifier (i.e. support vector machine (SVM), random forest (RF), etc) [34, 54, 71, 225, 249]. Yang et al. [249] proposed a texture-based analysis to extract gray-scale invariant features from US images by adopting a multi-resolution ranklet transform. By using support vector machines (SVM) as the classifier, they compared the extracted features with wavelet feature extraction methods and showed the higher performance of their proposed methodology in extracting features. Similarly, Gómez et al. [71] extracted texture features using the combination of co-occurrence statistics with several gray-level quantizations. Ding et al. [54] proposed two steps of SVM learning based on multi-instance learning wherein the first SVM trained on the local features extracted from the whole image, and the second one trained on features extracted from region of interests and sub-regions. Uniyal et al. [225] employed RF time series features in addition to B-mode texture features for the classification task of benign and malignant breast lesions. Chang et al. [34] first, identified the suspicious regions by applying watershed segmentation, then, extracted statistic and geometric features of the tumor.

The main focus of previous works was the extraction of hand-engineered features, more precisely, texture-related features from a US image. Recent state-of-the-art deep learning methods have paved the way for automatically extracting the most meaningful features by adopting convolution layers. The promising results of deep learning methods in the domain of US images for classification tasks have raised researchers' attention to this field [13, 29, 43, 83, 103, 178, 205]. Han et al. [83], employed a pre-trained GoogleNet with some modifications for classifying benign and malignant lesions of histogram equalized of 7408 US images. Becker et al. [13] employed a generic deep learning software for classifying benign versus malignant lesions and compared the results with human readers. The work by Chiao et al. [43] was performing a detection step prior to classification and segmentation of breast lesions using Mask R-CNN architecture. Huang et al. [103] performed a two-stage CNN in categorizing breast lesions based on Breast Imaging Reporting and Data System (BI-RADS). Moon et al. [156] proposed a computer-aided system (CAD) from an ensemble of several CNN

architectures for breast lesion identification. They showed that by combining the ROI of US B-mode image, segmented tumor, and binary mask as the input of the CNN network, the classification results improved significantly. Byra et al. [29] introduced a matching layer where gray-scale images were converted to RGB format as a color conversion step before feeding to a pre-trained network. In the work of Shin et al. [205], a framework based on weakly and semi-supervised scenarios was proposed to tackle the overfitting problem wherein limited data with strong annotations was available. The work by Qi et al. [178] presented two networks trained in a cascade manner where the input of each network in addition to US image, accepted the class activation map of the other.

5.3 Problem Statement

In most of the studies related to the classification of US breast lesions, the main focus is the binary classification of benign versus malignant lesions [29, 178, 205]. Consequently, the adopted deep learning network learns only one task. In the domain of multi-task learning (MTL) studies, it has been shown that networks perform better when they are assigned with multiple tasks compared to only one task [190]. Due to the appearance of invasive ductal carcinomas, the task of distinguishing them from fibroadenomas is the most challenging task compared to the binary classification of benign versus malignant. Therefore, we propose a deep learning-based scheme that performs better when it is assigned to a multi-class classification task. To be more specific, we propose a multi-class classification of fibroadenoma, cyst, and invasive ductal carcinomas in US images with limited data. We further propose a novel technique in taking the background of the US image into account as an additional class leading to a 4-class classification task for breast US images. We also show that the proposed scheme of multi-class classification including background as the additional class, holds for different deep learning networks. To cope with uncertainty in the network's estimations, we adopt test-time augmentation for classification evaluation[206]. Our contributions can be summarized as below:

- Multi-class classification of breast US with limited available images
- A novel technique in adding background as an additional class
- Our proposed scheme holds for different networks
- Test-time augmentation for evaluation

5.4 Methodology

5.4.1 Dataset

We use a publicly available US dataset [250]. It was collected in 2012 from the UDIAT Diagnostic Centre of the Parc Taul Corporation, Sabadell (Spain) with a 17L5 HD linear array transducer (8.5 MHz). The dataset consists of a total of 163 images with a mean image size of 760 × 570 pixels including one or more lesions. Out of 163 lesions, 53 images have cancerous masses, which include the subcategories of 40 invasive ductal carcinomas, 4 are ductal carcinomas *in situ*, 2 invasive lobular carcinomas, and 7 unspecified malignant lesions. The remaining 110 lesions are benign (65 cysts, 39 fibroadenomas, and 6 other types of benign). In the chapter, only 40 invasive ductal carcinomas (IDC), 65 cysts (Cyst), and 39 fibroadenomas (FA) are used as the other classes contain very few samples. Figure 5.1 presents an example of US images used from the dataset.

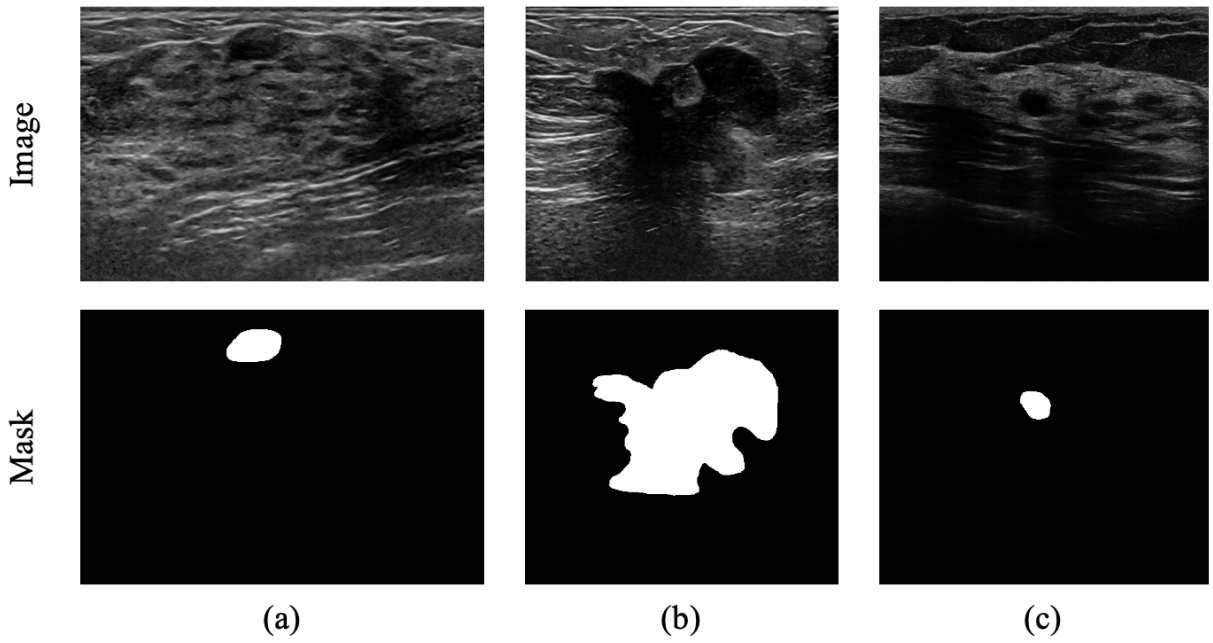


Figure 5.1: An example of breast US images from the dataset (a) FA, (b) IDC, and (c) Cyst.

5.4.2 Preprocessing

As mentioned earlier, we propose a novel technique for adding background as an additional class. To this end, before feeding the images to the deep learning network, they are cropped in order to help the network learn the characteristics of each lesion type more precisely.

Please note that we did not apply any denoising techniques during our preprocessing steps; we utilized the images in their original form. Based on available segmentation masks of images, a marginal cropping window surrounding the lesion (blue window in Fig. 5.1) is used which is 40% larger than the window surrounding the lesion borders (red window in Fig. 5.1). All the cropped images are then resized to the size of 400 × 400, and their intensities are normalized to the range of 0 to 1. In order to keep the balance between the number of lesion types in training and test sets, 80% of each lesion type is randomly selected for the training set and the rest is used in the test set. As for background class, a window is used to randomly crop the background (BG) of US image excluding any part of the lesion. In US imaging, the more the US waves travel deeper, the more they are attenuated. Therefore, in order to have a similar range of attenuation in the BG class, the BG images are cropped from the same depth of the lesion's location. In Fig. 5.1 the yellow square represents the area where BG images are cropped from. All the randomly selected BG images from the yellow area of Fig. 5.1 have the same size as the marginal cropping window (i.e. the blue window). In order to have a balanced number of BG classes, we randomly selected the BG images derived from all US images. As a result, in our training and validation sets the total number of FA, CYCT, IDC, and BG respectively are 31, 52, 32, and 38. Consequently, in our test set the number of each class is 8, 13, 8, and 10, respectively.

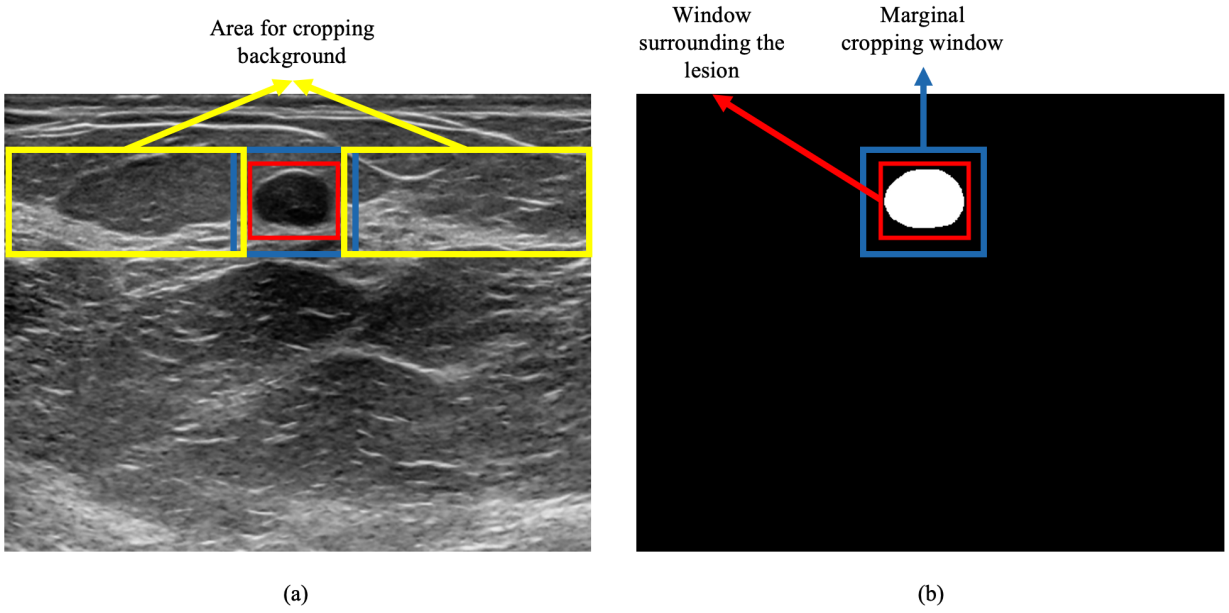


Figure 5.2: Cropping schematic used in this study. The red window is based on the lesion border in the mask. The blue window is 40% larger than the red window. The yellow rectangles show the desired area for selecting BG images.

5.4.3 Experiments

Due to the small number of training data and large inter-class variations in the database, training the network from scratch is impossible. Therefore, we use two pre-trained networks, ResNet-34 and MobileNet-v2, separately as backbone feature extractors with some modifications for our dataset. A schematic of our network is shown in Fig. 5.3. In order to show the impact of MTL on the performance of our network, we present our results based on two avenues: *2-class Avenue* and *4-class Avenue*. Our *2-class Avenue* is a 2-class classification problem (IDC versus all other classes), versus *4-class Avenue* is a 4-class classification (FA versus Cyst versus IDC versus BG). In both avenues, a random on-the-fly data augmentation of horizontal flip, width and height shifts, and zooming, is applied to the batches during training. The number of batches is set to 25 and training lasts for 100 epochs while saving the best model based on the validation accuracy. Adam optimizer is used [162] and its learning rate is tuned using cyclical learning rate [140]. It worth noting that to mitigate the effect of imbalanced data during training, we use a weighted cross-entropy loss function in both avenues wherein the weights are initialized based on the distribution of images in each class. For improving the predictions, we employ test-time augmentation [206] where the same augmentation strategy in the training set, is applied to the test set. Therefore, we enlarge the number of images in the test set by augmenting each image 4 times leading to 195 (i.e. 39×5) images.

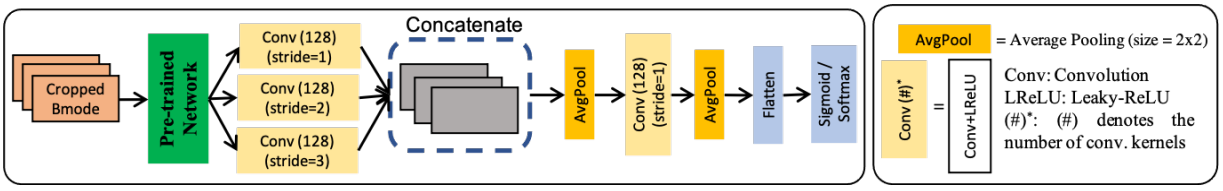


Figure 5.3: A schematic of our proposed network.

5.5 Results

The predicted labels are evaluated using the receiver operating characteristic curve (ROC) and the area under the curve (AUC) for both ResNet-34 and MobileNet-v2 in two proposed avenues. For ResNet-34, in *2-class Avenue* we achieve AUC of 0.66 for IDC. The AUC of IDC is improved by 31% in our *4-class Avenue* for the same network. Furthermore, the AUC scores for FA, Cyst, IDC, and BG in *4-class Avenue* are 0.87, 1.0, 0.87, and 1.0, respectively, for ResNet-34. Similarly, for MobileNet-v2, the AUC of IDC is improved from 0.82 in *2-class*

Avenue to 0.90 in *4-class Avenue*. The AUC scores of FA, Cyst, and BG for MobileNet-v2 are 0.87, 1.0, and 1.0, respectively.

Table 5.1 summarizes the classification reports for IDC in two proposed avenues. Please note that, as we used one-hot encoding for our analysis, the results shown in Table 5.1 focus solely on IDC predictions. This means that, in both the *2-class Avenue* and *4-class Avenue* approaches, the reported accuracies pertain specifically to IDC versus all other classes. For ResNet-34, the accuracy is improved by 18% in *4-class Avenue* showing that increasing the number of classes helps the network for better predictions. Precision and F_1-score are improved from 0.38 to 0.67 and 0.48 to 0.57, respectively. However, recall is decreased from 0.62 to 0.5 when using more classes (i.e. *4-class Avenue)*. Similarly, for MobileNet-v2, the accuracy is enhanced from 0.84 in *2-class Avenue* to 0.90 in *4-class Avenue*. We observe improvements in recall and F_1-score for MobileNet-v2, however, precision dropped from 1.0 in *2-class Avenue* to 0.80 in *4-class Avenue*.

Table 5.1: Classification report for IDC.

Avenue	***Accuracy***	***Precision***	***Recall***	***F_1-score***
		ResNet-34		
2-class	0.71	0.38	0.62	0.48
4-class	0.84	0.67	0.50	0.57
		MobileNet-v2		
2-class	0.84	1.0	0.25	0.4
4-class	0.90	0.8	0.50	0.62

5.6 Discussion

Breast cancer as the most common cause of death worldwide, can have a reduction in its mortality if it is diagnosed early and reliably. Automated breast cancer detection can improve the screening paradigm and assist radiologists in better examinations. Most of the previous studies on automatic US image classifications focused on the binary classification of benign versus malignant lesions. The main challenge in breast lesion classification is the detection of FA versus IDC due to their similar appearance. In Chapter 5, we showed the importance of MTL in better detection of IDC in breast US images. We investigated that increasing the number of classes led to better performance of the deep learning networks. We further proposed a novel strategy in adding the background of US images as an additional class. We

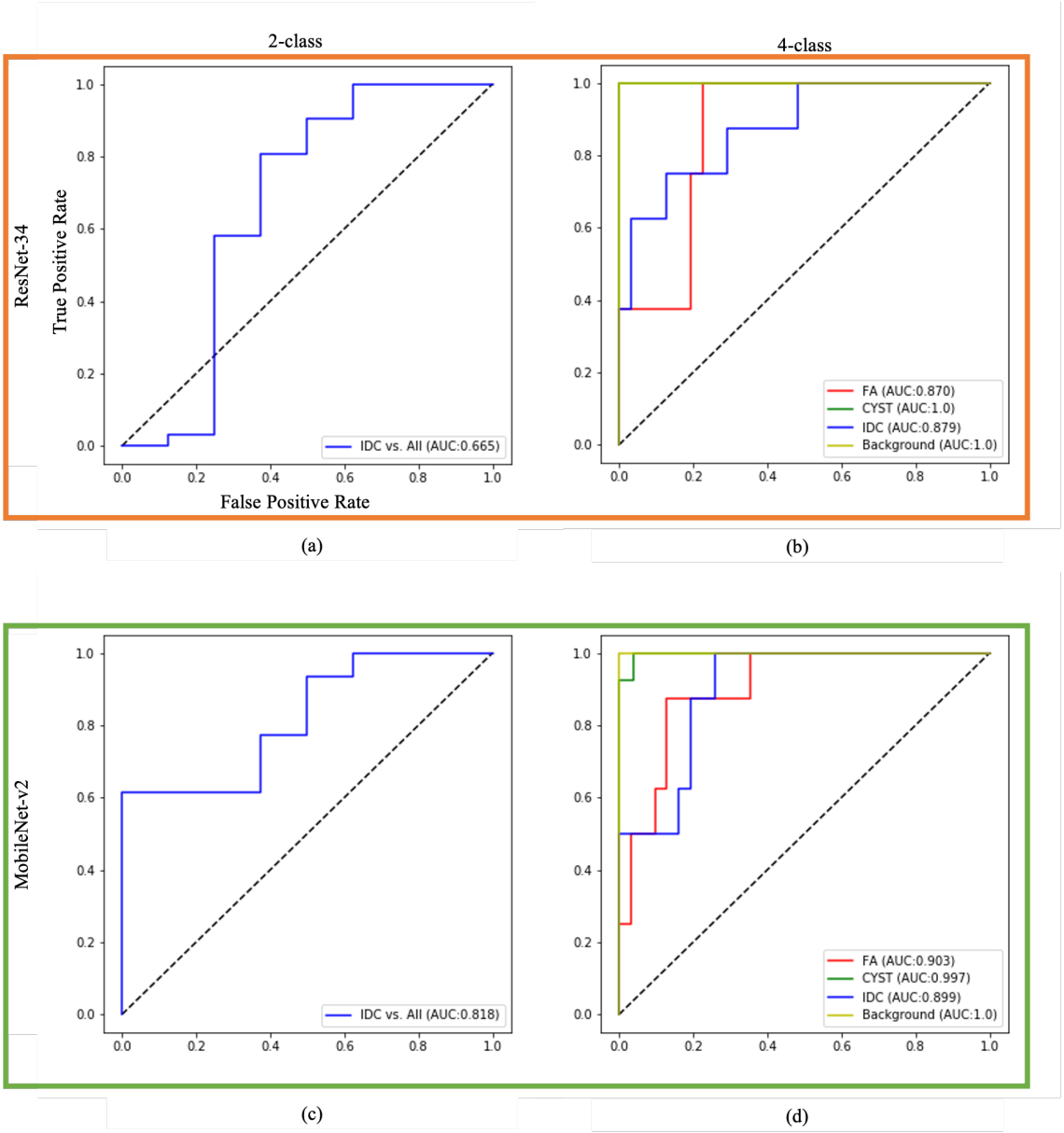


Figure 5.4: ROC curves and AUC results. The first and second rows show the ROC curves for ResNet-34 and MobileNet-v2, respectively. The first column ((a) and (c)) presents the results for *2-class Avenue* whereas the second column ((b) and (d)) presents for *4-class Avenue*.

showed that our proposed scheme holds for different deep learning networks by adopting 2 pre-trained networks, ResNet-34 and MobileNet-v2. By adding more classes, we illustrated that the AUC score was improved by a factor of 31% and 9% for ResNet-34 and MobileNet-vs, respectively. Also, to control the network's uncertainty in its predictions, we adopted test-time augmentation.

Chapter 6

DeepSarc-US: A Deep Learning Framework for Assessing Sarcopenia from Ultrasound Images

This chapter is based on our published paper [22].

6.1 Background

Frailty syndrome is a growing public health concern, especially among older adults with multiple medical conditions, in whom it is associated with high rates of disability, morbidity, mortality, and healthcare resource use [158]. Frailty is highly prevalent in cardiac patients, affecting 25–50% of those above age 70 and being associated with adverse outcomes in the settings of coronary disease, valvular disease, heart failure, and arrhythmia, to name a few [2]. A large body of evidence has shown the negative impact of frailty following invasive procedures like cardiac surgery [2, 96]. As such, cardiologists and cardiac surgeons have embraced the concept of frailty to better characterize their older patients, predict risk, and guide treatment decisions. This is crucial to ensuring benefits for patients and avoiding costly yet futile procedures. In addition, it is helpful to prepare patients before and after cardiac surgery through cardiac rehabilitation, exercise programs, nutritional supplementation, and comprehensive geriatric interventions. Consequently, proactive detection of frailty in cardiac patients may allow for the deployment of cost-effective, easy-to-implement preventive health measures that have been shown to improve clinical and patient-reported outcomes [50, 81].

There are multiple ways to operationalize frailty and measure it in the clinical setting [3, 63, 126]. One of the biggest challenges is the lack of clinician-friendly tools to measure frailty in acutely ill patients who are otherwise unable to perform the usual physical performance tests and questionnaires, especially those seen in the emergency department and cardiac intensive care unit. A proposed solution has been to measure muscle mass and quality as an objective biomarker of frailty, which does not require any patient effort to acquire and can be used in all patients, regardless of acuity [26].

The age-related loss of muscle mass and quality is known as sarcopenia, and it is one of the cornerstones of frailty syndrome [48, 267]. Sarcopenia has been defined as a systemic condition that reflects the functional and physical aspects of aging; thus, its assessment in a clinical environment can provide useful information about the patient's underlying frailty [57]. When evaluating sarcopenia, various physical tests and questionnaires are available, including assessments of hand grip strength, muscle mass, clinical frailty scale (CFS), frailty index (FI), and others [45, 118]. Measurement of muscle mass can be achieved using a variety of imaging modalities, including computed tomography (CT), magnetic resonance imaging (MRI), dual X-ray absorptiometry, or ultrasound (US) [27, 49, 113]. Additionally, measurement of intramuscular fat and inflammation (indicative of muscle quality) can be achieved using many of these modalities. US, compared to other modalities, has the major advantage of being portable and feasible at the patient's bedside, which is particularly attractive for acutely ill patients that cannot otherwise be electively transported for such tests. In addition, US is a non-invasive, portable, cost-effective, and safe imaging modality that is widely used in medicine [234]. The evidence for using US to measure the quality and quantity of a variety of muscle groups is compelling [49, 214, 215]. Unfortunately, US has noteworthy drawbacks such as a low inherent signal-to-noise ratio, low contrast to differentiate muscle from adjacent soft tissues, operator-dependent acquisition of images, and the need for significant time and training to quantitatively measure muscle thickness and intramuscular fat. Progress is needed to overcome these drawbacks before US can become a mainstream tool for the assessment of sarcopenia by clinicians.

Recent developments in image processing techniques such as deep learning (DL), particularly convolutional neural networks (CNNs) and transformers, have been instrumental in automating and standardizing the analysis of medical images and deriving informative features not otherwise apparent to the human eye. Blanc-Durand et al. developed a DL method that allowed for the automated and reliable quantification of skeletal muscle mass for sarcopenia assessment from CT images, which may be integrated into a clinical workflow [27].

Bian et al. used DL methods to optimize and improve cardiac US images in hospitalized patients with chronic heart failure (CHF). Their findings further revealed a correlation between CHF and sarcopenia [25]. Pintelas et al. used an autoencoder DL method to condense valuable information from multi-frame US muscle images, followed by a classification strategy for the diagnosis of sarcopenia [176]. In a similar study, Sobral et al. compared several DL methods to diagnose sarcopenia and differentiate it from normal muscle tissue [209]. Yet another study used DL methods to segment muscle tissue from US images [147].

ConvNext is an advanced CNN model showcasing notable performance in various computer vision applications [139]. It leverages a sophisticated architecture, incorporating deep convolutional layers that enable it to capture intricate patterns and features within images. The model benefits from its ability to automatically learn hierarchical representations, making it well-suited for complex image recognition tasks. ConvNext has demonstrated competitive accuracy and efficiency, making it a valuable asset in the realm of DL-based image analysis, specifically in US image analysis [56, 86, 115].

Current vision-transformer (ViT) models, inspired by natural language processing (NLP) studies, have shown promising performances compared to CNNs [58]. ViT-based models do not have the inductive locality bias of CNNs, which makes CNNs less effective at modeling long-range dependencies. Instead, ViT-based models take advantage of their data-driven self-attention mechanism, which helps them to better understand the contextual information derived from not only the region of interest but also its surrounding [58, 91, 127]. Despite all the aforementioned benefits compared to CNNs, these models are data-hungry, and their performance can be limited by the size of the training dataset. To tackle this limitation of ViT-based models, He et al. proposed the masked auto-encoder (MAE), a self-supervised learning approach that includes image inpainting [91]. The MAE paradigm is a self-supervised technique for ViT-based models that enables the network to learn useful information by predicting masked targets. MAE has shown potential for faster training and better generalization of the ViT-based models. In ViT models, feature maps are created based on a single low-resolution image by adopting a fixed-scale windowing step. Liu et al. [137, 138] proposed Swin Transformer V2 (SwinT) that builds hierarchical feature maps by adopting a non-overlapped shifted windowing step into the encoder of vanilla ViT, and it is capable of training with images of up to 1536×1536 pixels (vanilla ViT refers to ViT model proposed by [58]). SwinT has shown capabilities for learning functional dependencies between features.

ViT-based models have demonstrated superior performance over simple CNN-based models in various computer vision tasks as well as US image analysis [69, 129, 135, 159, 181, 203,

218, 238, 257]. However, ConvNext, employing a convolutional architecture, has surpassed ViT-based models in certain contexts of natural image analysis [139]. Notably, neither ViT nor ConvNext have been specifically applied to the measurement of muscle quality as of the current knowledge update. Furthermore, despite the promising performance of ConvNext and ViT-based models compared to simple CNN-based models, the training step is challenging, particularly when there are a lack of sufficient data. In the clinical scope, due to security and privacy policies, the publicly available US data are very limited. Labeled data are especially scarce, since manually annotating medical data is expensive. Therefore, utilizing complex models in clinical settings is quite challenging. To this end, a strategy is proposed for training the complex models on a small set of US images. We aimed to explore the performance of three recent CNN-based and three ViT-based DL models to estimate quadriceps muscle thickness (QMT) based on a limited number of US images using two main strategies that provide a fair comparison of ViT- and CNN-based models. To better comprehend the decisions made by the models, examples of the visualization maps produced by the ViT- and CNN-based models are further examined. These visualization maps can also offer supplementary information that can be used as a guide for practitioners at the time of data acquisition. The contributions of this work are summarized as follows:

- Three CNN- and three ViT-based models are proposed to estimate QMT using ultrasound images acquired in a clinical setting.
- A strategy is proposed for optimizing the training of DL models to estimate QMT more accurately, especially when limited data are available.
- The activation maps are explored to provide clinicians with real-time feedback. This feedback can potentially be used to help clinicians collect better US images to help DL models estimate QMT more accurately.
- To the best of our knowledge, it is shown for the first time that DL can be used to automatically estimate QMT from US images taken from phased array probes.

This chapter is organized as follows. In Section 6.2, first, the data collection and experimental setup are outlined. Then the results of the proposed QMT estimation strategies and the derived visualization maps are presented in Section 6.3. Finally, the outcomes of our experiments are concluded in Section 6.4.

6.2 Methods

6.2.1 Dataset

The dataset is based on a prospective cohort of 486 adult patients undergoing a clinically indicated cardiac US examination at the Jewish General Hospital (JGH), Montreal, Canada, between 1 October 2018, and 30 June 2019. These patients provided verbal informed consent to participate in this study. At the end of the cardiac US examination, with the patient remaining in a supine position, the sonographer acquired a static image of the left quadriceps muscles (rectus femoris and vastus intermedius) at the level of the anterior thigh, midway (approximated visually) between the anterior superior iliac spine and patella. A GE E95 machine (GE HealthCare, Chicago, IL, USA) with a sector 4Vc-D phased-array probe was used for data acquisition, where the acquisition setting was set to standard adult transthoracic settings with the center frequency of 1.4 MHz. The phased-array probe used to image the heart was also used to image the quadriceps muscles, even though the latter are usually imaged using linear probes that are better suited for superficial structures. This was done for convenience of clinical implementation, as cardiac US systems do not typically include linear probes.

The quadriceps US images were extracted in DICOM format and manually annotated by a trained observer (J. O.) to define the region of interest corresponding to the skeletal muscle tissue between the superior margin of the femur and the inferior margin of the subcutaneous fat, which is defined as QMT. Specifically, for each image, the upper border of the femur was delineated by marking nine equally spaced points (with the fifth point centered on the femur bone), and the upper border of the muscle was delineated by marking five equally spaced points (with the third point centered on the quadriceps muscle, aligning roughly with the central point of the femur). QMT was then determined by measuring the vertical distance between the central points of the femur and quadriceps. To validate the QMTs measured by J.O., referred to as the ground-truth QMTs in this study, and to prevent training the models with potentially inaccurate QMTs, the muscle thickness in each US image was additionally measured by two independent medical researchers using the same subset of images. Table 6.1 presents the inter-rater variabilities for redundant QMT measures. Inter-rater reliabilities were evaluated using the intraclass correlation coefficient (ICC) with a 95% confidence interval (CI) [119, 132]. All ICC values exceeded 0.80, indicating very good to excellent reliability.

The size of the images was 708×1080 pixels, which was resized to 224×224. It is worth

Table 6.1: Inter-rater Variability of QMT measurements.

QMT measurement (values are in *cm*.)				**ICC(2,1)**	**95% CI**	**ICC(2,3)**	**95% CI**
Rater 1	Rater 2	Rater 3	Rater 4				
3.29 ± 1.07	2.94 ± 0.86	2.99 ± 0.87	3.24 ± 1.12	0.83	[0.77-0.87]	0.95	[0.93-0.96]

noting that the ground truth QMT was computed before resizing the images to 224 × 224. Given the variability in image acquisitions across subjects, pixel spacing was duly considered in this calculation. Five-fold cross-validation was used in all the experiments. Therefore, 67% of the images (330 images) were used for training, 17% (83 images) were used for validation, and the remaining 15% (73 images) were used as the test set. Figure 6.1 presents three examples of US images with annotations, and Figure 6.1a–d presents the distribution of QMT in centimeters (cm). Please note that pixel spacing varies across patients, so one pixel does not correspond to the same length in centimeters for different patients. Therefore, in Figure 6.1a,b, the QMT is 1.57 cm and 2.17 cm, respectively, even though the number of QMT pixels in (a) is higher than in (b).

6.2.2 Experimental Setup

The overarching goal was to automate the evaluation of US-based QMT as a clinically useful and accessible indicator of sarcopenia. Figure 6.2 summarizes the proposed framework.

Regression and Classification for QMT Measurements

Both regression and classification approaches were employed to achieve this goal, and within each approach, a total of six DL models were compared using either their ImageNet weights [191] or our experimentally derived weights. The first approach consisted of training a regression model to predict the QMT value (in cm) as a continuous output. The second approach consisted of training a classification model to predict the QMT class (10 classes binned in 0.5 cm increments starting at 1.0 cm) as an ordinal output and also to generate activation maps for visualization. Using this experimental setup, the following hypotheses were tested:

1. Models with transformer and CNN architecture would achieve good results in predicting QMT.

2. Regression models with pre-trained weights experimentally derived from classification

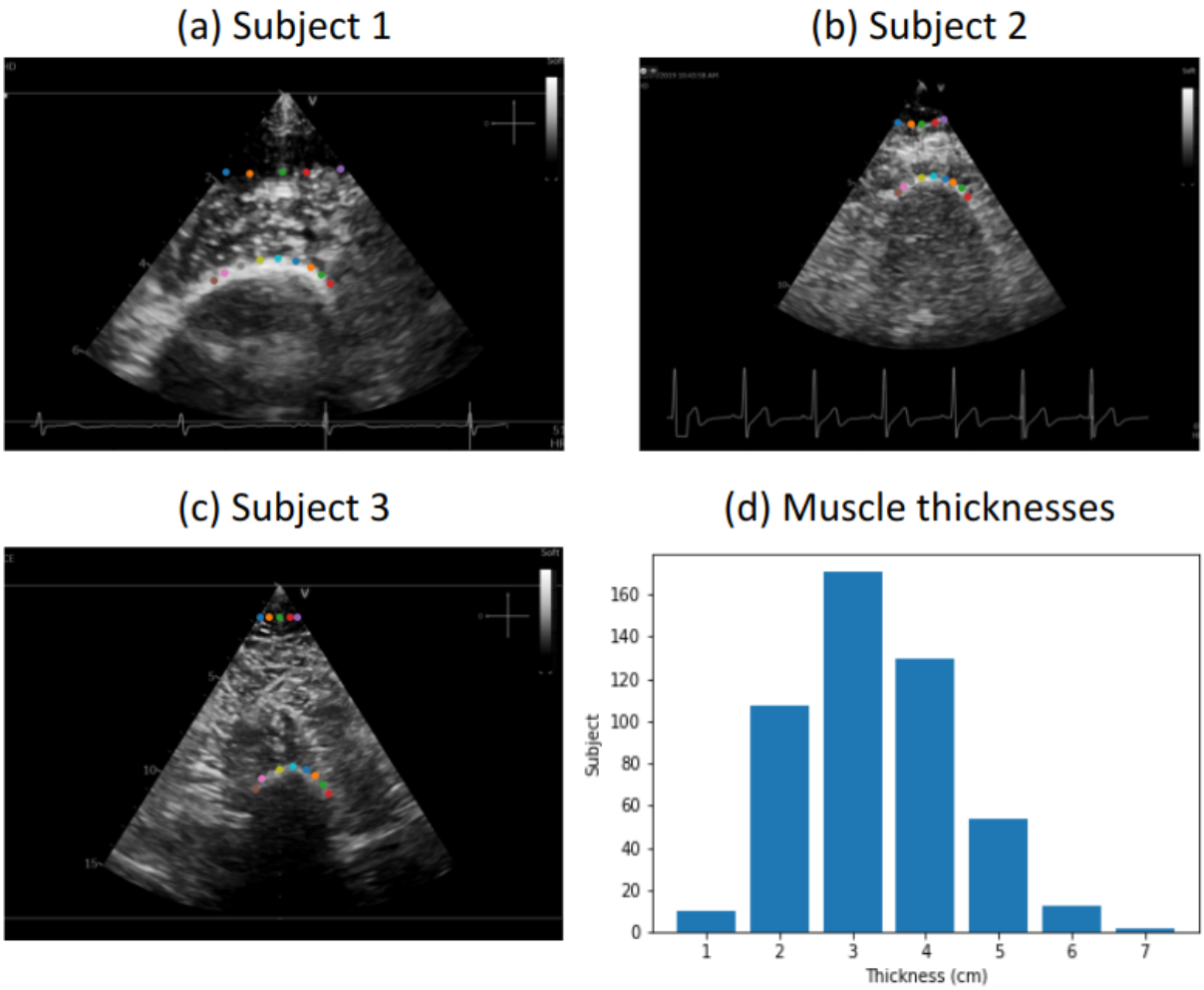


Figure 6.1: Examples of the dataset with QMT of (a) 1.57 cm, (b) 2.17 cm, (c) 6.75 cm (Note: The pixel spacing varies between images (a)–(c), so one pixel does not correspond to the same length in cm across these images). The colored dots represent the annotations of the quadriceps muscle and femur bone surfaces (better seen in colored prints). (d) The distribution of QMT across all 486 subjects.

training runs would outperform the same models with pre-trained weights from ImageNet.

3. Classification models with pre-trained weights experimentally derived from regression training runs would outperform the same models with pre-trained weights from ImageNet.

4. Activation maps that correctly highlight the anatomical structures of interest would be more likely to correspond to accurate predictions of QMT.

The six DL models investigated for regression and classification tasks were ResNet101

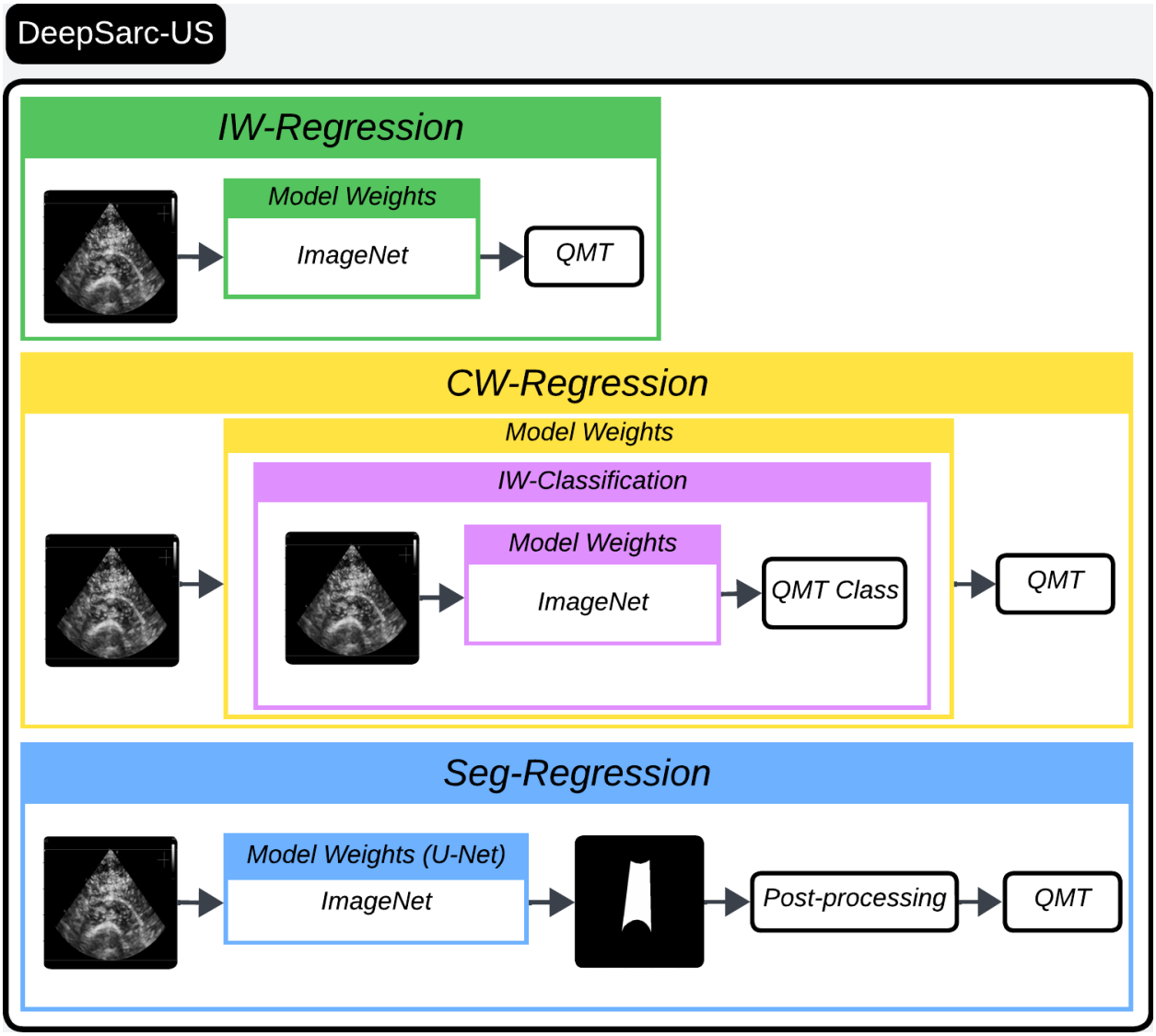


Figure 6.2: Summary of the proposed QMT measurement framework: *IW-Regression*, *CW-Regression*, and *Seg-Regression*. *IW-Regression* model initialized with ImageNet weights, *CW-Regression* model initialized with *IW-Classifictaion* weights, and *Seg-Regression* model initialized with ImageNet weights.

[89], DensNet121 [101], ConvNext [139], ViT [58], MAE [91], and SwinT [138]. For transformer-based models (i.e., ViT, MAE, and SwinT), the base (i.e., ViT-B, MAE-B, and SwinT-B) and large (i.e., ViT-L, MAE-L, and SwinT-L) architecture designs of the models were used in our experiments. For ConvNext, the base architecture design was utilized.

The initialization weights investigated for the regression models were experimentally derived from classification training runs (CW-Regression) vs. a priori derived from ImageNet weights (IW regression). The rationale is that pre-training a model on an easier task (i.e.,

classification) can help it learn basic features, and then training it on a more sophisticated task (i.e., regression) can fine-tune it to learn complex features [106, 217, 232].

The initialization weights investigated for classification models were experimentally derived from regression training runs (RW classification) vs. a priori derived from ImageNet weights (IW classification). Since the ImageNet weights were designed for 1000 classes, the last layer was modified accordingly for 10 classes (1–1.5 cm, 1.5–2 cm, 2–2.5 cm, 2.5–3 cm, 3–3.5 cm, 3.5–4 cm, 4–4.5 cm, 4.5–5 cm, 5–5.5 cm, above 5.5 cm).

For all training, the Adam [116] optimizer with a learning rate of 0.0001 was used. Five-fold cross-validation was used, and the results were presented as the average of each fold for the patients comprising the test set. Furthermore, multiple augmentations, including horizontal flipping, cropping, Gaussian noise addition, and blurring, were randomly applied to the training set.

Training Regression Models for QMT Estimation: The regression models in the proposed strategies were trained for hundreds of epochs until the loss did not change for 50 consecutive epochs, using the mean squared error (MSE) loss function defined as:

$$MSE_{loss} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2, \tag{6.1}$$

where y and $\hat{y}$ represent the ground truth and predicted QMT, respectively, in a batch of N (32) US images. As previously noted in Section 6.2.1, the ground truth values of QMT were derived from manual annotations of skeletal muscle.

Training Classification Models for QMT Estimation and Activation Map Visualization: The classification models were trained using the focal loss function [133]. Focal loss, as defined in Equation (6.2), tackles the class imbalance problem by reducing the loss contribution from common samples (which are usually correctly classified) and increasing the importance of rare cases (which are often misclassified).

$$Focal_{loss} = -\alpha(1 - p_t)^{\gamma} log(p_t), \tag{6.2}$$

where p_t represents the estimated probability for the corresponding class. α and γ are defined as the weight and modulating factors, respectively. The recommendations in [133] were followed for the initialization of these factors. Therefore, α and γ were set to 0.25 and 2, respectively. Predefined class intervals set by our clinician were used. The intervals, defined

as [1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 7.5], led to a total of 10 classes (0 to 9). For example, the class for a QMT of 1.3 cm was set to 0. For visualizing activation maps, Grad-CAM [197] is utilized for CNN-based models. For transformer-based models, only the activation maps of the ViT and MAE models were investigated by adopting the transformer interpretability method [35].

Segmentation for QMT Measurement

An alternative method to attain the objective of automating the assessment of QMT measurements in US images involved generating segmentation masks using annotations of quadriceps muscle and femur surfaces. Subsequently, the length between these surfaces was measured. Figure 6.3 illustrates examples of US images, where the first column on the left displays annotation points on the US images, and the second column on the left showcases generated segmentation masks derived from these annotation points for three patients (Figure 6.3a–c). The generated segmentations can be employed to calculate muscle thickness in US images. The muscle thickness can be calculated by identifying the lowest pixel of the muscle surface and the uppermost pixel of the femur surface, as indicated by the dashed lines in Figure 6.3. It is noteworthy that while segmentations can be used to measure muscle thickness, their application may not extend to the assessment of other biomarkers of muscle quality that are commonly utilized in sarcopenia detection.

For the training phase, binary masks were initially generated based on annotations of the surface areas of the muscle and femur. Following a standard approach for training the proposed regression and classification models, 5-fold cross-validation technique was employed. The model architecture utilized for segmentation was a modified U-Net [188] with ResNet50 as its backbone. The chosen configuration included Dice loss as the loss function and a learning rate of 0.001. Dice loss can be defined as follows:

$$\text{Dice Loss} = 1 - \frac{2 \times \text{Intersection}}{\text{Union} + \epsilon}, \tag{3}$$

where ϵ was set to 0.00001.

Subsequently, majority voting was applied to consolidate the binary masks for the 73 subjects in the test set. To be more specific, each test image had five masks generated by the five models trained on the five train-validation sets (i.e., 5-fold cross-validation). The final mask for each test image was determined via majority voting among these five generated masks. After this, muscle thickness was computed in a post-processing step involving the

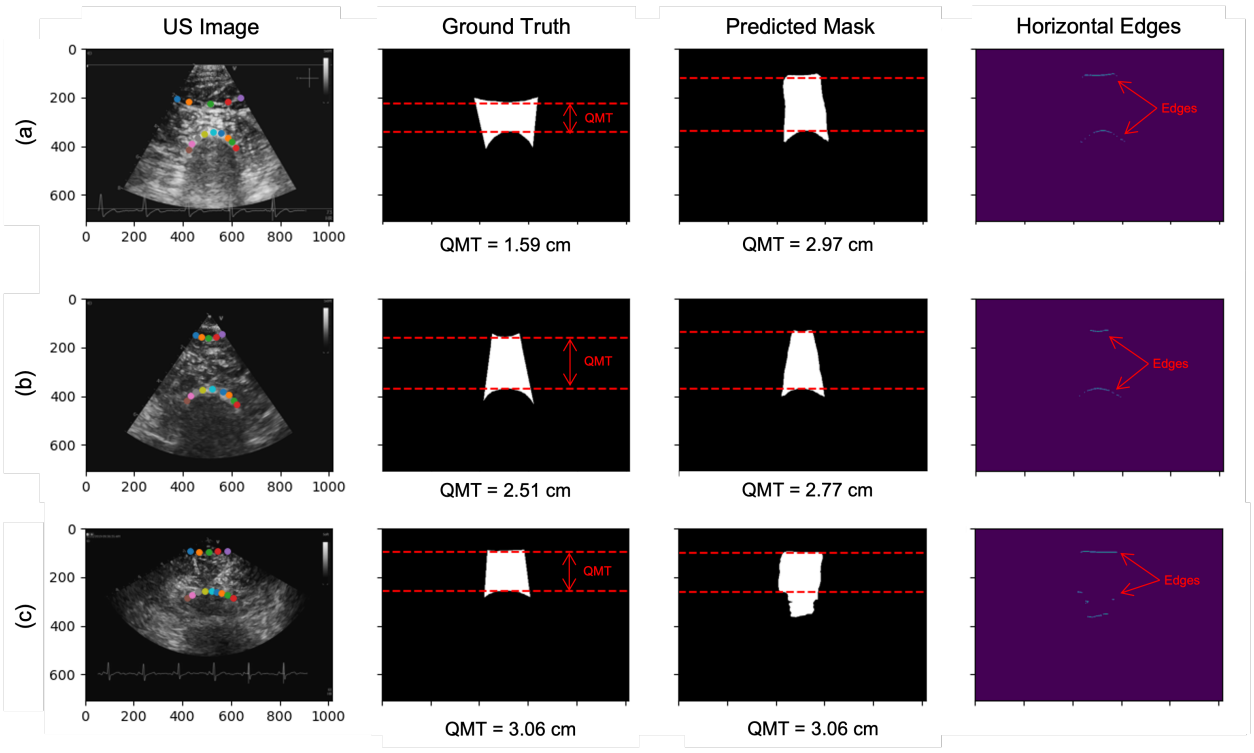


Figure 6.3: Examples of segmentation masks generated from manual annotations (*ground truth*) for three patients (**a**–**c**) (better seen in colored prints). The predicted masks showcase the segmentation outcomes achieved by the *Seg-Regression* model (*predicted mask*). The Dice scores of predicted masks for patients (**a**), (**b**), and (**c**) were 0.63, 0.89, and 0.76, respectively. In this process, the QMT was derived through a post-processing step that involved determining the distance between the horizontal edges of the muscle surface and the femur surface, utilizing Canny edge detection (*horizontal edges*).

measurement of the distance between the surfaces of the muscle and the femur. In the post-processing phase, the horizontal edges of the binary mask were identified using the Canny edge detection method. This facilitated the determination of surface curves for both the muscle and femur. The QMT was subsequently determined by identifying the bottom-most pixel on the muscle surface and the highest pixel on the femur surface. An example of a QMT measurement from predicted segmentation masks is shown in Figure 6.3. For simplicity, the QMT measured based on predicted segmentation masks has been denoted as *Seg-Regression* throughout the remainder of this manuscript.

The overall overview of the proposed framework is summarized in Figure 6.2. The *IW-Regression* model is initialized with ImageNet weights, the *CW-Regression* model is initialized with *IW-Classifictaion* weights, and *Seg-Regression* model is initialized with ImageNet weights.

6.3 Results

As previously mentioned, the results presented below are based on an average of 5-fold cross-validation for the 73 subjects in the test set. Here, we present a summary of the terminologies utilized throughout this manuscript for reference.

- *IW-Regression*: This denotes the regression model utilizing ImageNet pre-trained weights (*I* referring to ImageNet weights). For instance, *IW-Regression* ResNet101 signifies the fine-tuned ResNet101 with ImageNet weights specifically tailored for the regression task of QMT.

- *IW-Classification*: This represents the classification model leveraging ImageNet pre-trained weights. Similarly, *IW-Classification* ResNet101 signifies the ResNet101 with ImageNet weights fine-tuned for the classification of QMT.

- *CW-Regression*: This designates the regression model fine-tuned using *IW-Classification* model weights. For example, *CW-Regression* ResNet101 is the ResNet101 initially initialized with ImageNet then fine-tuned for the classification of QMT (as denoted by *IW-Classification*), and subsequently fine-tuned once more for the regression task of QMT.

- *RW-Classification*: This signifies the classification model fine-tuned using *IW-Regression* model weights. For example, *RW-Classification* ResNet101 is the ResNet101 initially initialized with ImageNet and fine-tuned for the regression of QMT (as denoted by *IW-Regression*) and then further fine-tuned for the classification task of QMT.

- *Seg-Regression*: This denotes the measurement of QMT through a post-processing step applied to the predicted segmentation masks.

6.3.1 Regression of QMT

The median absolute error over the test set is summarized in Table 6.2. In Table 6.2, the asterisk (p-value < 0.05) and double asterisks (p-value < 0.01) show a significant difference between the error distributions of the *IW-Regression* model and its *CW-Regression* counterpart. For the significance test, the Wilcoxon signed-rank test (histograms of errors did not show Gaussian distributions) was used. It is also shown that the average median absolute error across all *CW-Regression* models is significantly less than that of the *IW-Regression* models.

Table 6.2: Median absolute error of QMT estimations.

Median absolute error (values are in *cm*)			
Model	***IW-Regression***	***CW-Regression***	***Delta***
ResNet101	0.40	0.30	+0.10 ↑
DensNet121	0.50	0.38	+0.12 ↑*
ConvNext-B	**0.27**	**0.23**	+0.04 ↑
ViT-B	0.31	0.27	+0.03 ↑*
ViT-L	0.32	0.29	+0.03 ↑
MAE-B	0.34	0.31	+0.03 ↑
MAE-L	0.28	0.28	0.00
SwinT-B	**0.27**	0.24	+0.03 ↑*
SwinT-L	0.30	**0.25**	+0.05 ↑*
All	0.28	0.25	+0.03 ↑ **

* and ** denote a statistically significant difference with p-value < 0.05 and p-value < 0.01, respectively, between *IW-Regression* , models with ImageNet weights, and *CW-Regression*, models with corresponding classification weights. ↑: *CW-Regression* outperformed *IW-Regression*, ↓: *IW-Regression* outperformed *CW-Regression*.

As shown in Table 6.2, most of the *CW-Regression* models showed improvements compared to their *IW-Regression* counterparts, as demonstrated in the Delta column. Therefore, it can be concluded that the prior step of classification training for a regression task can help to boost the QMT estimations in US images. Additionally, the results demonstrate that SwinT-L, the transformer-based model, with median absolute errors of 0.30 and 0.25 for its *IW-Regression* and *CW-Regression* models, respectively, significantly benefit more from our proposed training strategy than the other transformers-based models. Among the CNN-based models, ConvNext showcased superior performance by achieving a median absolute error of 0.23 in *CW-Regression*. Furthermore, it demonstrated a notable enhancement of 0.04 cm in QMT estimation when comparing *CW-Regression* with *IW-Regression*.

Given the limited number of US images available, pretraining a model for a classification task first, which is perhaps seen to be an easier challenge, can enable the model to be better prepared for the regression task in both transformer- and CNN-based models. However, the substantial impact of the proposed strategy is dependent on the model architecture, where significant improvements in DensNet121, ViT-B, SwinT-B, and SwinT-L were observed. There were no significant improvements in ResNet101, ConvNext-B, ViT-L, and MAE-B. The complete statistical analysis is presented in Figure 6.4.

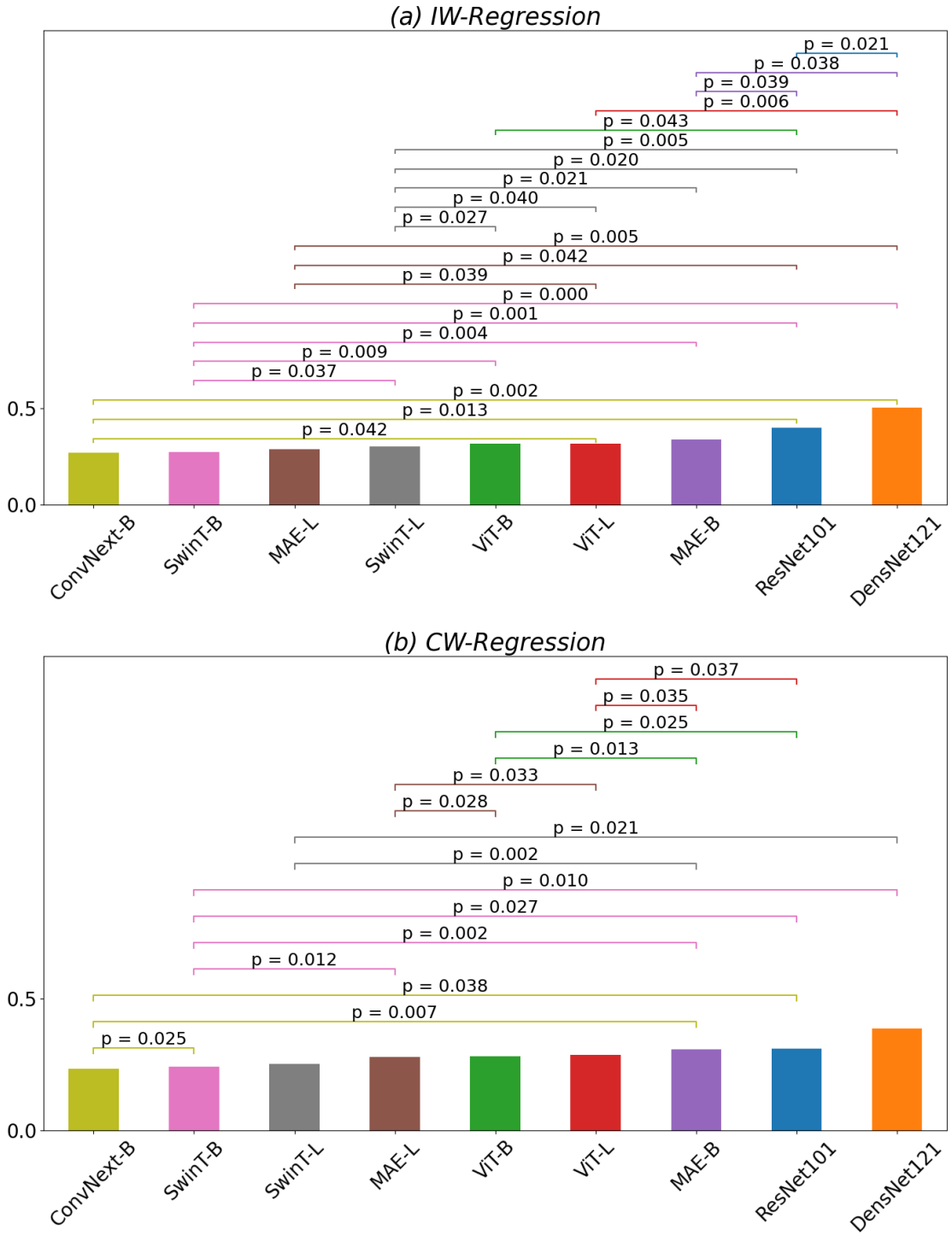


Figure 6.4: Statistical analysis for the (**a**) *IW-Regression* and (**b**) *CW-Regression* models.

6.3.2 Classification of QMT

This section summarizes the results of the classification tasks. As previously explained, the *RW-Classification* models were initialized with the weights of our regression training runs; however, the *IW-Classification* models were initiated with publicly available weights and were pretrained on the ImageNet dataset. The classification accuracy results are shown in Table 6.3. Among the *IW-Classification* models, SwinT-L showed the best accuracy of 43.84%, and among the *RW-Classification* models, ViT-L showed the best accuracy of 43.84%. It was observed that the *IW-Classification* models showed better performances compared to the *RW-Classification* models. SwinT-B and SwinT-L were the only models that showed improvements in their *RW-Classification* compared to *IW-Classification* models. Additionally, the transformer-based models outperformed CNN-based models in both the *IW-Classification* and *RW-Classification* models, demonstrating their superior performance in the classification task .

Table 6.3: Accuracy of QMT classification in classification models.

Accuracy (%)			
Model	***IW-Classification***	***RW-Classification***	***Delta***
ResNet101	36.99	30.14	−6.85 ↓*
DensNet121	39.73	38.36	−1.37 ↓
ConvNext-B	31.51	32.88	+1.37
ViT-B	42.47	38.36	−4.11 ↓**
ViT-L	41.10	**43.84**	+2.74 ↑
MAE-B	38.36	36.99	−1.37 ↓
MAE-L	41.10	41.10	0.00
SwinT-B	41.10	36.99	−4.11 ↓*
SwinT-L	**43.84**	42.47	−1.37 ↓
All	43.84	41.10	−2.74 ↓

* and ** denote a statistically significant difference with p-value < 0.05 and p-value < 0.01, respectively, between *IW-Classification*, models with ImageNet weights, and *RW-Classification*, models with corresponding regression weights. ↑: *RW-Classification* outperformed *IW-Classification*, ↓: *IW-Classification* outperformed *RW-Classification*.

The activation maps from *Classification* models for ResNet101, DensNet121, ViT-B, and MAE-B are shown in Figure 6.5 and 6.6. Based on the achieved activation maps, it was discovered that in both the *IW-Classification* and *RW-Classification* models, the activation

maps in the ViT-B and MAE-B models mainly focused on detecting the edge of the femur bone or the skeletal muscle tissue itself in most of the correctly classified US images, as shown in Figure 6.5c,d. There were no distinguishing patterns in the activation maps between the *IW-Classification* and *RW-Classification* models. Furthermore, no discernible differences in the activation maps of ViT-B against ViT-L or MAE-B versus MAE-L were found; thus, only the results for ViT-B and MAE-B are provided. The activation maps for two sample test subjects are shown in the first row for correctly classified cases and in the second row for misclassified cases. It is important to note that the activation maps were not relevant in some misclassified cases of CNN-based models, especially when the classification errors were greater than two classes, as shown in Figure 6.5e,f.

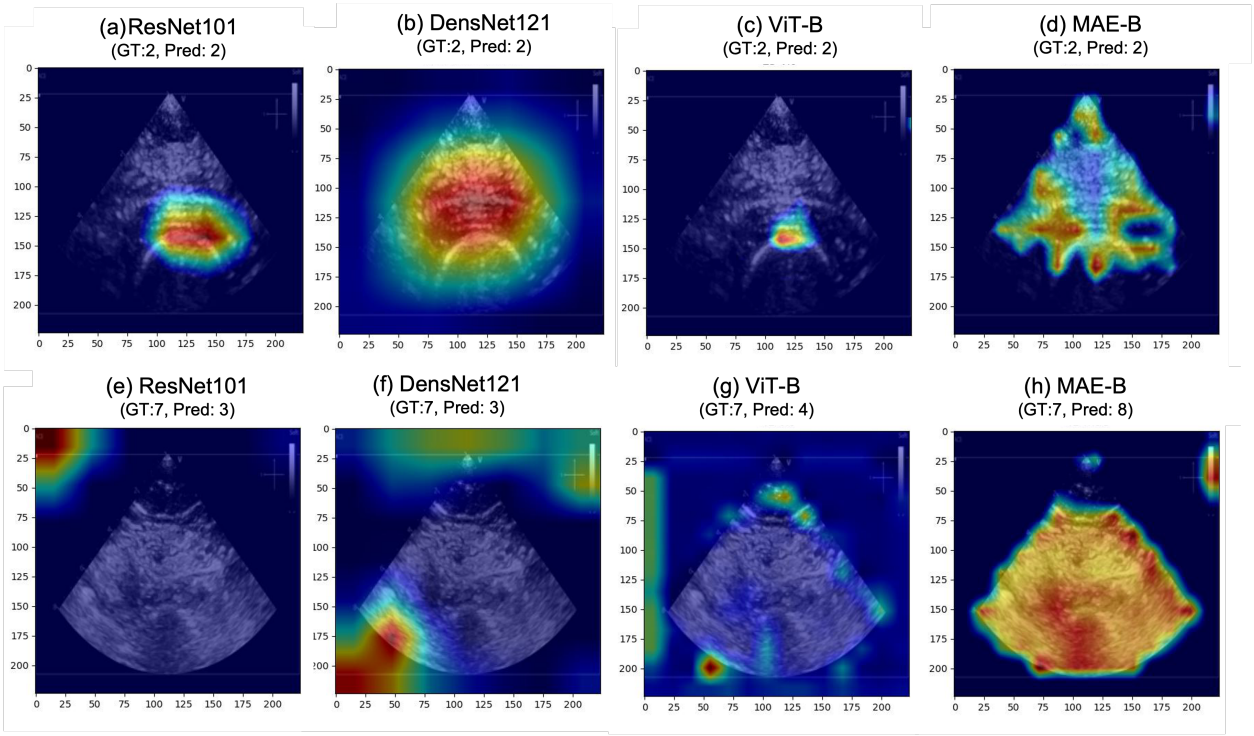


Figure 6.5: Activation maps in classification models: ResNet101 (**a**, **e**), DensNet (**b**, **f**), ViT-B (**c**, **g**), and MAE-B (**d**, **h**). The first and second rows represent activation maps for two different subjects. The first row was correctly classified, and the second row was misclassified. (GT: ground truth class label, Pred: predicted class label).

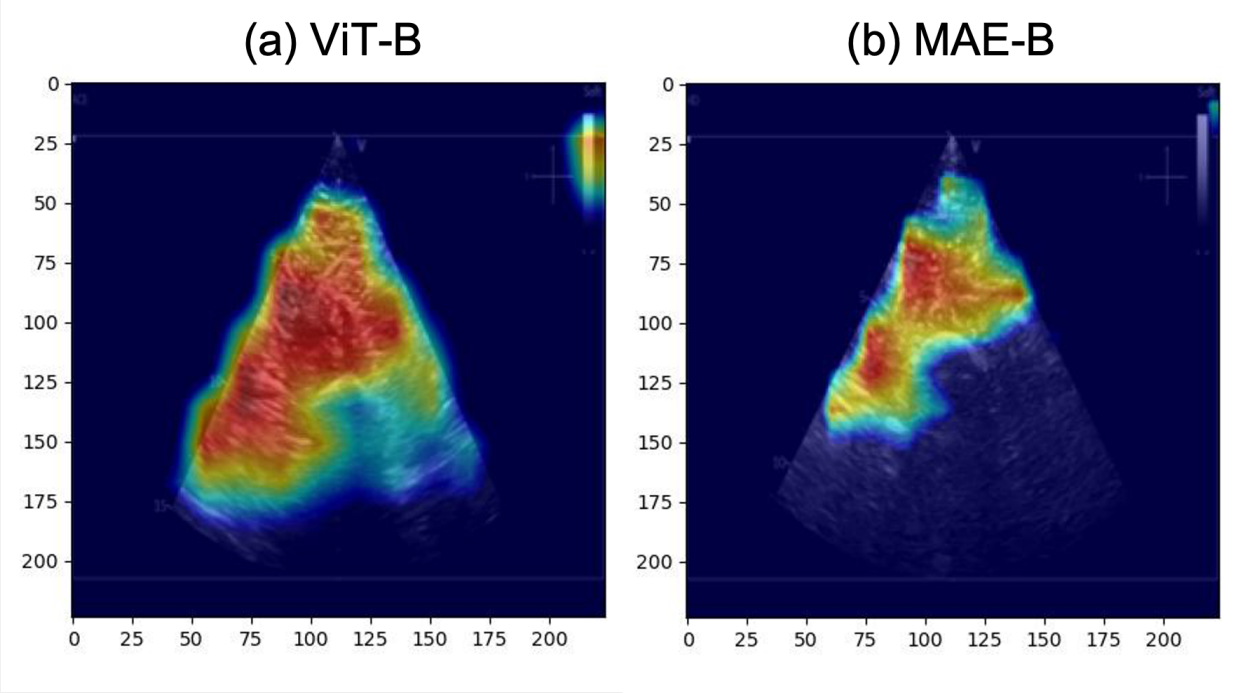


Figure 6.6: Sample activation maps of (**a**) ViT-B and (**b**) MAE-B, detecting the body of the muscle.

6.3.3 Segmentation of QMT

This section presents the results of QMT measurements derived from segmentation masks. An evaluation of the predicted masks was performed using the Dice score, defined as follows:

$$\text{Dice score} = \frac{2 \times \text{Intersection}}{\text{Union} + 0.0001}. \tag{6.3}$$

The aggregate Dice score across 73 subjects yielded 0.90 ± 0.06, indicating perfect predictions. Figure 6.3 showcases examples of predicted masks for three patients. The horizontal edges were identified using the Canny edge detection method applied to the predicted masks. Following this, the QMT was computed by measuring the distance between the lowest point of the muscle surface and the uppermost part of the femur surface, as represented by the dashed lines in Figure 6.3. The median absolute error of QMT measures based on *Seg-Regression* was found to be 0.13 cm. In the context of estimating QMT, the model consistently demonstrated an outstanding performance, showcasing its superior ability in this particular task. Table 6.4 presents a comparison of the median absolute error derived from *Seg-Regression* with that of *IW-Regression* and *CW-Regression*. Table 6.4 displays the averages of all the models for *IW-Regression* and *CW-Regression*. As the error distributions

did not adhere to normality, the medians of the absolute errors are reported in this table. Subsequently, a Wilcoxon signed-rank test was conducted to assess statistical significance.

Table 6.4: Median of absolute errors in QMT estimations

Median of absolute errors (cm)		
Seg-Regression	***IW-Regression***	***CW-Regression***
0.13	0.28**	0.25**

** denotes a statistically significant difference with p-value < 0.01 between *Seg-Regression* and two other methods.

6.4 Discussion and Conclusions

In this study, a method was developed for the measurement of QMT using US images acquired using a phased array probe in a clinical setting. QMT has been put forth as an objective biomarker for frailty and a diagnostic criterion for sarcopenia. A number of DL models, including ResNet101, DensNet121, ConvNext-B, ViT-B, ViT-L, MAE-B, MAE-L, SwinT-B, and SwinT-L, were compared to predict this measurement.

First, there was no significant and consistent difference when comparing all the CNN-based models with the transformer-based models in predicting QMT. In the *IW-Regression* models, as summarized in Table 6.2, ConvNext-B, a CNN-based model, achieved a median absolute error of 0.27, while of the transformer-based models, SwinT-B also achieved the same median absolute error of 0.27. Furthermore, upon conducting a significance test, as illustrated in Figure 6.4a, the observed difference between the two models was not statistically significant. In the *CW-Regression* models, as depicted in the results presented in Table 6.2, ConvNext-B once again demonstrated a superior performance among the CNN-based models, achieving the lowest median absolute error of 0.23. Similarly, within the transformer-based models, SwinT-B and SwinT-L closely followed, with median absolute errors of 0.24 and 0.25, respectively, showing competitive results comparable to ConvNext-B. Therefore, it can be concluded that both CNN- and transformer-based models exhibit a satisfactory performance when adequately trained.

Second, performance improved for the *CW-Regression* models compared to the *IW-Regression* models for the task of QMT prediction. In the domain of CNN-based models, all the models experienced improvements through the utilization of the *CW-Regression* strategy. To this end, ResNet101, DensNet121, and ConvNext-B demonstrated improvements of

0.1 cm, 0.12 cm, and 0.04 cm, respectively. Similarly, within the transformer-based models, the implementation of the proposed *CW-Regression* strategy generally resulted in enhanced performance. Excluding MAE-L, which showed no improvements, ViT-B, ViT-L, MAE-B, SwinT-B, and SwinT-L achieved improvements of 0.03 cm, 0.03 cm, 0.03 cm, 0.03 cm, and 0.05 cm, respectively. These findings highlight the fact that for those models that are harder to train using small datasets, simplifying the task (i.e., pretraining on the classification task first) is highly beneficial for their performance. This is in line with a large body of evidence in curriculum learning [24] that has shown that training deep models on easier tasks first leads to better results. This is often due to the non-convex nature of the loss landscape, where training on an easy task helps the network to not be trapped in poor local minima. Given that transformer-based models have shown a large potential in the natural language processing field, our proposed *CW-Regression* transformer models will be advantageous for US images where there are limited images available.

Third, our experiments on the classification of QMT, i.e., the *IW-Classification*, and *RW-Classification* models, showed that prior pretraining on a regression task is not beneficial for the classification models, which emphasizes the fact that regression is a harder task compared to classification. When compared to the *IW-Classification* models, the *RW-Classification* models generally displayed decreased performance across all eight SOTA models. Furthermore, there was no significant difference between the activation maps of the *IW-Classification* and *RW-Classification* models. Although the current work concentrates on non-ordinal classification, investigating ordinal classification in further studies may provide insightful information. Assessing inter-observer variability, improving annotation techniques, and creating strong algorithms that can process ordinal data effectively could all help improve the accuracy of measuring QMT in US images. Such initiatives would lay the groundwork for enhanced clinical applications and a more thorough understanding of muscle health.

Fourth, the activation maps derived from the classification of QMT can further provide complementary information that can be beneficial for sonographers and clinicians when collecting and interpreting the images. It was found that either the femur bone or its surroundings were the main focus in activation maps for correctly classified cases. When the surface of the femur bone was not clearly visible in the US image, the activation maps of CNN-based models did not display any useful visualizations, and the model usually misclassified that case. As a result, since these visualizations can be obtained online during the data collection, they can aid in ensuring that the US image is collected in such a way that the surface of the femur bone is clearly visible in the image in order to improve the QMT

predictions. While the activation maps provide qualitative insights, the lack of ground-truth annotations for the entire body of muscle limits the ability to conduct quantitative experiments on the generated activation maps. This constraint highlights the challenge of comprehensively assessing muscle features without complete annotations and prompts consideration for future investigations to incorporate more extensive annotation datasets. It is worth noting that variations in data acquisition settings can impact the model's activation maps. In this context, the utilization of transfer learning techniques becomes crucial for improving the model's adaptability to diverse acquisition settings. In future work, it could also be highly beneficial to visualize activation maps during the training phases. Saving checkpoints of training weights and tracking and analyzing the progression trends of activation maps throughout the training process can provide valuable insights for further optimization.

The present study encountered a notable limitation regarding the absence of complete annotations for the entire muscle body in our dataset. While we successfully trained a classification model to estimate muscle thickness based on annotations of the surface of the femur bone and the surface of the muscle, the lack of ground truth annotations for the entire muscle limits our ability to conduct quantitative experiments on the generated grad-cam images. This constraint highlights the challenge of comprehensively assessing muscle features without complete annotations and prompts consideration for future investigations to incorporate more extensive annotation datasets.

Moreover, the adoption of segmentation masks to automate QMT measurements in US images presents a compelling avenue for streamlining the evaluation of muscle thickness. Specifically, the QMT measures derived from the segmentation-based approach (*Seg-Regression*) exhibited noteworthy distinctions when compared to those obtained through *IW-Regression* and *CW-Regression*. Despite *Seg-Regression* showcasing superior performance over its counterparts, it is imperative to acknowledge certain limitations. There were instances where the prediction of the segmentation masks encountered failures. As illustrated in Figure 6.3a, for instance, the generated mask is inaccurately produced, leading to an erroneous measurement of QMT (error of 1.38 cm). Similarly, in Figure 6.3c, while the QMT is accurately measured, the predicted mask itself is inaccurate. Therefore, the accuracy of the post-processing step relies heavily on the segmentation model's performance. Furthermore, the application of *Seg-Regression* is primarily tailored for the precise measurement of muscle thickness and may not be seamlessly extended to assess other essential biomarkers such as hand grip strength, FI, CFS, etc., required for sarcopenia assessments. This constraint underscores the need for a nuanced understanding of the method's scope, emphasizing its

suitability for specific QMT-related assessments while recognizing its limitations in a broader context.

A novel aspect of this study is the demonstrated ability to generate accurate predictions of QMT despite the use of suboptimal phased-array US source images. Such measurements would otherwise require training, practice, and added time for clinicians to perform manually—all of which are major barriers to clinical translation. Phased-array images were used because of their convenience for clinical acquisition without the need to switch probes in the context of a cardiac US exam. Further research is underway using linear probes in the context of a point-of-care US exam using a handheld system. The images provided by these probes are ideally suited to imaging superficial skeletal muscles and should, in theory, generate even more accurate predictions of QMT.

It is important to note that predicting sarcopenia involves more than just measuring muscle thickness. Accurate assessment requires integrating various patient measurements to provide a comprehensive evaluation. The proposed technique is designed to automate the entire sarcopenia assessment process by relying solely on patient data, minimizing the need for practitioner input. This approach is particularly useful in scenarios where immediate access to equipment for measuring muscle thickness is not available, such as bedside data collection. By incorporating our method, we can facilitate the development of fully automated, online sarcopenia assessment systems, enhancing the efficiency and accuracy of patient evaluations.

One limitation of this study is that only one leg was imaged. A more accurate approach would involve imaging both legs and averaging the measurements, although this would significantly increase the scan time. Future research could explore the costs and benefits of imaging both legs. Another constraint of the current study is the phased-array transducer that was used for data acquisition. Limb muscles, as mentioned earlier, are typically evaluated using linear probes that are better adapted for superficial structures. However, in the current study, phased-array images were used because of their convenience for clinical acquisition without the need to switch probes in the context of a cardiac US exam.

The contribution of this study is a necessary step preceding the clinical application of QMT at scale. The objective was to provide detailed benchmarking and a comparative analysis of the performance of various models, helping us to understand the strengths and weaknesses of different architectures in the context of muscle thickness measurements. By providing reliable and precise muscle thickness measurements, this study contributes essential data that, when combined with other diagnostic criteria, will help categorize patients more

specifically within the spectrum of sarcopenia. Ultimately, this will enhance the ability of healthcare providers to diagnose and manage sarcopenia more effectively. Having derived DeepSarc-US, the stage is now set to apply this method to a large-scale clinical cohort to validate its diagnostic performance against a gold standard determination of sarcopenia. It is important to note that muscle size in isolation may not be sufficient to diagnose sarcopenia, and ancillary criteria assessing muscle quality and strength may very well be needed. Further research involving radiomic features of muscle quality is currently an active area of ongoing investigations that has the potential to complement the proposed measures of muscle size.

Chapter 7

Open Access Segmentations of Intra-operative Brain Tumor Ultrasound Images

This chapter is based on our published paper [21].

7.1 Background

Gliomas are the most common malignant primary brain tumors originating from glial cells and are classified into grades 1-4 by the World Health Organization (WHO) [144, 170]. Grades 1-2 are low-grade, while grades 3-4 are high-grade tumors [195]. Surgical resection is a standard treatment for gliomas, and pre-operative magnetic resonance (MR) imaging is used for tumor characterization. However, brain tissue deforms during surgery due to factors like edema and gravity (i.e. brain shift) [68], rendering pre-operative MR images inaccurate. Acquiring data at different stages during surgery helps the surgeon better monitor the progress of the tumor resection and, consequently, operate more precisely. Intra-operative imaging, particularly intra-operative MR (iMR) and intra-operative ultrasound (iUS) aids surgeons by providing updated guidance[11, 161, 185]. While iMR offers superior image quality, it is costly, adds a long time to the operation, and requires dedicated operating rooms[252, 258]. In contrast, iUS is a cost-effective, flexible, and versatile modality that presents real-time scanning without altering the surgical workflow[12, 189, 213]. Due to the easy procedure of acquiring iUS rather than iMR, several recent studies have demonstrated

the use and interest in iUS in neurosurgery [189, 193, 213, 252].

While iUS presents several advantages in the context of brain tumor resection, ultrasound (US) images can be difficult to interpret. Non-standard imaging planes and unfamiliar contrast are major factors limiting the efficient and widespread use of US in neurosurgery. To mitigate such limitations and fully leverage the advantages of iUS, automated image segmentation of structures such as tumors within iUS images can provide valuable assistance to neurosurgeons during procedures. Recent automated image segmentation algorithms, such as deep learning (DL) algorithms have made advancements in brain tumor segmentation from both US [30, 33, 107, 163] and MR [31, 130, 152, 212] images. However, access to high-quality datasets expertly annotated is essential for the development and validation of DL algorithms [219]. In the context of medical imaging, MR images have seen more readily available datasets compared to other modalities, making them the primary focus for DL algorithm development. The BRATS challenge, among others, stands out as a prominent dataset with valuable annotations that have significantly contributed to the evolution and refinement of DL algorithms [149]. Acquiring such data, especially for iUS images, is expensive and rare.

Currently, there are only three publicly available datasets that provide iUS brain images, the BITE dataset [150], the RESECT database [240], and ReMIND [114]. The BITE dataset contains pre- and post-operative MR scans as well as multiple iUS images of 14 patients. The RESECT database contains pre-operative MR scans and iUS images from 23 patients with low-grade gliomas. The ReMIND dataset contains 369 pre-operative MR scans, 320 3D iUS scans, 301 iMR scans, and 356 pre-operative MR segmentations of 114 patients. None of the abovementioned datasets contain segmentation of anatomical structures in the iUS images, thereby hindering the development and validation of iUS processing methods. For the RESECT database, a few research groups have previously conducted segmentations of iUS images [31, 33, 160], However, a portion of these annotations remained inaccessible to the public, and in some instances, only a small subset of cases was segmented, with limited validation procedures in place.

In this work, we present the most comprehensive and validated expert segmentations of cerebral structures in iUS images from the RESECT database. The focus is on delineating the tumor in pre-resection iUS 3D volumes and identifying the resection cavity during and after the surgery. To enhance the surgeon's ability to achieve more precise tumor resection, *sulci* and the *falx cerebri*, whenever they were within the field of view, are also delineated. These structures commonly serve as crucial anatomical references for surgeons, given their

clear visibility in iUS images. The following sections detail the segmentation and validation protocols for all the mentioned structures in the iUS images and brain tumor segmentation in pre-operative MR images.

7.2 Acquisition and Validation Methods

In this section, a comprehensive overview of the dataset is presented, along with detailed annotations for various anatomical structures, including tumors, resection cavities, *falx cerebri*, and *sulci*. The annotation process involved the utilization of two primary tools: ITK-SNAP [255] and 3D Slicer [62], chosen based on individual preference and familiarity. Furthermore, specific built-in features of these software platforms were leveraged as necessary, such as smoothing and interpolation functionalities, to enhance the accuracy and completeness of the annotations. Further elaboration on these tools and their respective functionalities is provided in the subsequent paragraphs. Given that both ITK-SNAP and 3D Slicer are widely used tools in the field, the decision to select one over the other was solely based on our inter-group preferences. It is important to highlight that manual segmentation entails human judgment and expertise, enabling nuanced interpretation and adjustments tailored to the anatomical complexities of each case. Consequently, the choice of segmentation software or algorithm does not introduce bias, as it depends on the proficiency and diligence of the annotators.

7.2.1 RESECT Database

The RESECT database [240] comprises pre-operative contrast-enhanced T1-weighted and T2 FLAIR MR scans alongside three 3D volumes of iUS scans from 23 patients with low-grade gliomas (grade 2) who underwent surgeries between 2011 and 2016 at St. Olavs University Hospital, Trondheim, Norway. The iUS scans were acquired at three different stages of the procedure: before resection, during resection, and after resection for control. These US images were captured by an expert surgeon, and the database includes manual neuroanatomy landmarks, facilitating MR-to-US volume registration and inter-US volume alignment. The details of the image acquisition procedure can be summarized as follows:

- Pre-operative MR scans: T1-weighted and T2 FLAIR sequences were acquired on 3T Magnetom Skyra MR scanners, both with 1 mm isotropic voxel size, except three patients who underwent the MR imaging on a 1.5T Magnetom Avanto MR scanner with 1 mm slice thickness.

- Intra-operative US scans: 3D US images collected using a 12FLA-L linear probe of Sonowand Invite neuronavigation system with a frequency range of 6-12 MHz integrated with the NDI Polaris optical tracking system.

7.2.2 iUS Tumor Segmentation Protocol

Tumoral tissue in US images is typically identified through abnormal echogenicity or texture variations compared to healthy tissue. Echogenicity refers to the level of reflectivity or brightness of tissue on a US image. In this context, variations in echogenicity indicate potential areas of malignancy, allowing medical professionals to identify and examine potential cancerous lesions. In the study, 19 out of 23 cases' iUS images (cases 1 to 23) were segmented, initially following Munkvold et al.'s method [160]. In these cases, initial US volume segmentations were already available. Four cases (cases 24 to 27) without prior iUS segmentations relied on MR segmentations to define the tumor region of interest in iUS images. To be more specific, for these four cases, the MR segmentation served as a starting point to define the region of interest (ROI) for the tumor in the iUS images. Further details of the MR segmentation protocol can be found in the work of Munkvold et al. [160].

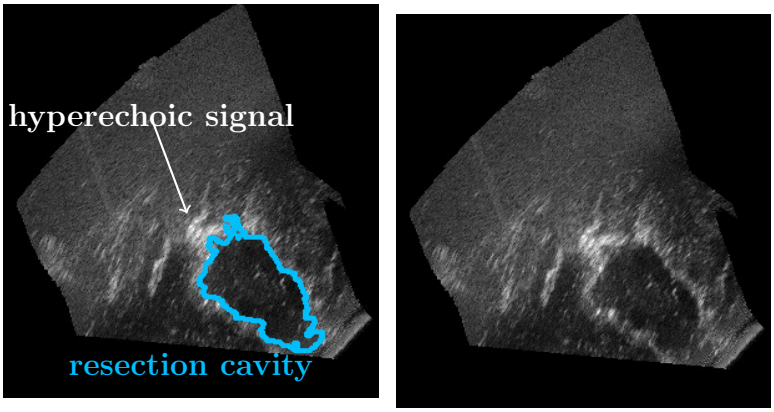
In the iUS segmentation protocol, 3D Slicer [62], a free and open-source medical image analysis software, was employed to perform ground truth segmentations on the acquired iUS images. For cases 24 to 27 that were initiated with MR segmentations due to brain shift during resection surgery, the boundaries of the tumor in the US images and MR tumor segmentations did not align [240]. This discrepancy necessitated the registration of MR tumor segmentations to iUS images using available landmarks from the RESECT database. This registration process effectively mitigated the misalignment of tumor borders between the iUS images and their corresponding MR tumor segmentations. Therefore, for cases 24 to 27, the MR segmentations were imported into the 3D Slicer scene and after the registration step, they were utilized as the initial delineation for iUS.

In all cases, to refine the initial 3D US tumor segmentations, the Label Map Smoothing module, an existing feature in 3D Slicer, was employed. Subsequently, the smoothed tumor segmentations were manually fine-tuned to ensure accurate coverage of the tumor region in the iUS image. To facilitate further correction, FLAIR MR images were registered to iUS images since MR images were unaffected by brain-shift effects. These complementary overlays served as guidance for generating more precise iUS tumor segmentations. An illustrative example of tumor segmentation is provided in Fig. 7.3 (a)-(d).

7.2.3 iUS Resection Cavity Segmentation Protocol

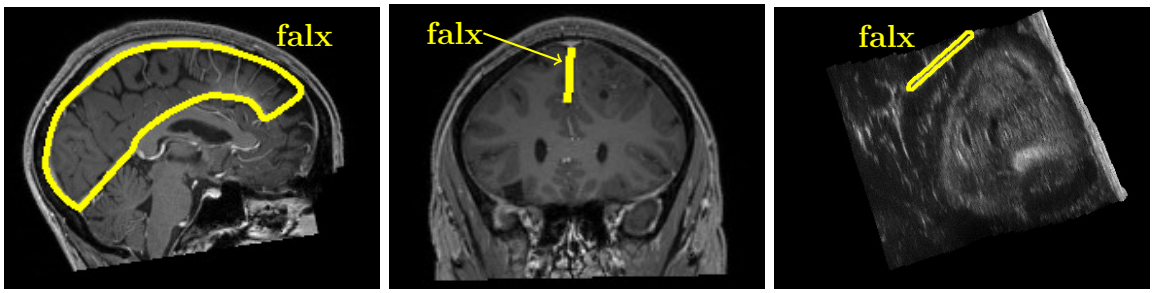
Resection cavity segmentation in iUS images encompasses the volume where tissue has been resected or retracted during image acquisition. Resection, a surgical procedure characterized by the complete removal of tissue or tumors, contributes to the formation of a distinct three-dimensional space within the brain known as the resection cavity. On the other hand, retraction, another surgical maneuver, entails cutting tissue and displacing it to the side without complete removal. Hyperechoic signals surrounding cavities in iUS images result from sound attenuation differences between brain tissue and saline water, as well as the presence of blood (see Fig. 7.1) [196]. To prevent false positives, only homogeneous, dark signals were considered as cavities, potentially leading to slight underestimation in cases involving blood-filled cavities. In challenging cases with small, entirely blood-filled cavities, segmentation was not possible due to indistinguishable borders. For example, in three exceptional instances (Case 11 during and after resection, and Case 15 during resection), the cavities appeared notably small and completely inundated with blood, without any noticeable dark signals.

It is worth noting that, due to the inherent variability in surgical procedures, determining the precise timing of image capture was challenging since it could occur at different stages of the resection process. The crucial factor was ensuring that US images were taken before the surgeon completed the resection entirely, even if residual tumors remained. The segmentation of resection cavities was conducted using ITK-SNAP [255]. Initially, regularly spaced slices, approximately one in every five slices, were manually delineated. Subsequently, ITK's morphological interpolation, facilitated by ITK-SNAP's Convert3D command-line tool, was utilized to fill in the remaining slices. Convert3D is one of the companion tools of ITK-SNAP that provides additional features. It is a command-line tool that enables the combination of multiple image processing tasks into efficient mini-programs, making it an integral tool in studies involving hundreds of 3D images. In cases where necessary, additional slices were manually segmented to optimize the interpolation outcome. The segmentation process for most RESECT cases was originally carried out by two raters as part of a previous study [33]. Following an assessment of intra- and inter-rater variability, these segmentations were reviewed by a neurosurgeon and were then modified accordingly. Subsequently, for this study, the remaining RESECT cases were segmented, and all cases underwent refinement during the validation protocol detailed in section 7.2.7.



(a) (b)

Figure 7.1: Ultrasound image of a resection cavity (a) with and (b) without segmentation.



(a) (b) (c)

Figure 7.2: T1 MR (a), (b), and US (c) images of the falx cerebri.

7.2.4 iUS *Falx Cerebri* Segmentation Protocol

The *falx cerebri* is the membrane that separates the left and right hemispheres of the brain. This cerebral falx has a hyperechoic signal in iUS images. It presents a characteristic quasi-planar shape that appears as a straight line in coronal and axial slices. The falx is also visible in the MR images, especially T1-weighted (Fig. 7.2). This structure is thus a convenient landmark that can be particularly useful to anchor registration. The falx is not always within the iUS volume due to the limited field of view but can be visible depending on the tumor location. The falx segmentation is, therefore, present for some volumes only. Since the falx' bright signal is similar to *sulci* it is difficult to localize the inferior border of the membrane. We, therefore, used the registered T1-weighted MR images to adjust the falx segmentation in height. Regularly spaced slices were manually delineated using ITK-SNAP and then interpolated with Convert3D.

7.2.5 iUS *Sulci* Segmentation Protocol

For segmentation purposes, *sulci* were defined as folds filled with cerebrospinal fluid (CSF) between brain tissue sections. CSF surrounding the brain, such as between the tissue and the *dura mater* or *tentorium*, was not labeled, although it had a similar iUS signal. *Sulci* were initially segmented with manual delineation every five slices and morphological interpolation using Convert3D when moving through the volume in a single direction (e.g., axial slices). Unlike volumetric structures, *sulci* are thin, complex, folded surfaces. To capture their irregular shapes, each volume was annotated in three viewing directions (axial, sagittal, and coronal), and these segmentations were first interpolated separately and later combined into a union. While this process resulted in slight over-segmentation, it significantly improved *sulci* delineation, according to annotators and neurosurgeons. In each case, the volume preceding resection underwent segmentation aided by registered MR images to accurately delineate *sulci* structures. This initial segmentation served as a reference for segmenting the volume during and after resection. The manual segmentation process was facilitated by ITK-SNAP, employing a Wacom One pen tablet, which demonstrated superior speed and precision compared to a conventional computer mouse.

7.2.6 Pre-operative MR Tumor Segmentation

For completeness, we provide segmentations of the tumors in the pre-operative T2 FLAIR images. As the cases in the database are lower-grade gliomas, there is no contrast uptake in the T1 weighted images and the T2 FLAIR images are used to define the tumor boundaries. The tumors were semi-automatically segmented in 3DSlicer using the *GrowCut* algorithm [60]. The resulting segmentations were manually corrected when needed and smoothed with a 2×2 mm median filter.

7.2.7 Data Validation

All segmentations presented in this work were validated by two experienced neurosurgeons (S.D.R., O.S.). The segmentations were presented to the specialists through a case-by-case 3D Slicer scene, including the original iUS image, segmentation masks, and MR images. Figure 7.3 represents an example of such a scene. We asked specialists to grade all segmentation masks based on five criteria for three types of structures:

- Quality of tumor: smoothness of the boundaries (SMT), identification of tumoral tissues (IdT), exclusion of non-cancerous tissues (ExT)

- Quality of resection cavity: identification of resection cavity (IdR)
- Quality of *sulci* and falx: identification of sulci and falx (IdS)

The grading scheme for each criterion was on a scale of 1 to 5 defined as major improvement needed, minor improvement needed, acceptable, good quality, and excellent, respectively. For each criterion, a score of 3 from both surgeons was needed to pass the quality control of segmentation masks. Otherwise, the masks were revised according to the surgeons' comments. In determining the choice of a passing score of 3, it is important to clarify that this decision was rooted in the specific criteria established for surgeons evaluating the dataset. The selection of 3 as the pass score was deliberate, as it represented the midpoint on the validation scale of 1 to 5. This choice was informed by the summary of evaluation forms, aiming to identify a common rating among inter-rater assessments. Given the intricate and complex nature of brain structures, achieving consensus on a moderate score like 3 ensured a balanced assessment that accounted for the variability inherent in such evaluations. The final grades for each patient are presented in Table 7.1.

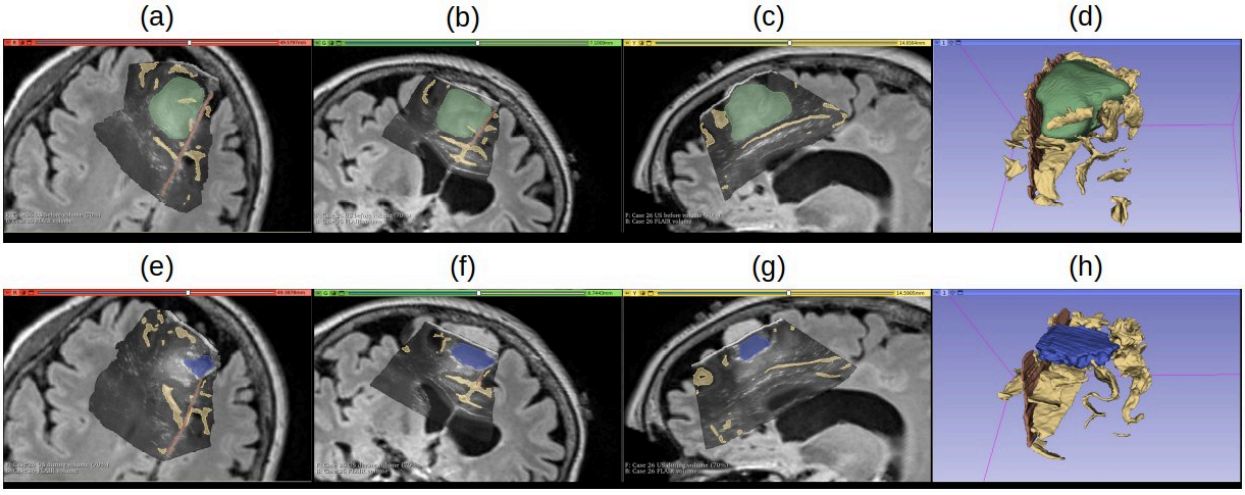


Figure 7.3: An example of segmentations overlaid with intra-operative ultrasound (iUS) and MR images (green: tumor; yellow: sulci; red: cerebral falx; blue: resection cavity). iUS volume before resection: (a)-(d); iUS volume during resection: (e)-(h).

7.3 Data Format and Usage Notes

The proposed segmentations are distributed in the NIFTI format. Upon the acceptance of this paper, they will be available via the OSF open-science `https://osf.io/jv8bk` for

public viewing and downloading and can be freely used by research laboratories as well as clinical institutes. However, gaining any financial benefits from the distribution of the proposed segmentation dataset is prohibited. The database is under the CC BY-NC-SA 4.0 License.

7.4 Discussion

The border and shape of brain tumors have long been established as important diagnostic markers in resection surgeries. Several image processing techniques have been adopted to segment tumors which rely on the creation of mathematical descriptions of the tumor border. Similarly, identifying resection cavity contours has been used to evaluate the completeness of tumor resection. Finally, segmenting surrounding cerebral structures can greatly benefit image analysis during the surgeries. However, validation of image processing techniques needs to be investigated in the case of new data. It is important to highlight that there are few brain datasets accessible to the public, and even those available, such as the BITE [150], RESECT [240], and ReMIND [114] datasets, do not include iUS segmentation of brain anatomies. The absence of US segmentation datasets for the brain is largely attributed to the complexity of the task. To this end, we have provided the manual segmentations of cerebral structures in iUS images of the 23-patient RESECT dataset, verified by two expert surgeons through detailed evaluation criteria per brain anatomy. Our proposed expert-annotated dataset comprises the segmentation masks of brain tumors, resection cavities, the *falx cerebri*, and *sulci*. Tumor segmentations of the pre-operative MR images are also provided as a reference. To the best of our knowledge, this is the first study that publicly provides a comprehensive expert-annotation segmentation of iUS images. The challenging procedure of delineating brain iUS images has impeded the publications of such segmentations. This validated dataset serves as a crucial asset for evaluating and benchmarking various segmentation methods, thereby driving advancements in brain imaging research.

The proposed dataset offers substantial utility, focusing primarily on two pivotal applications that have the potential to revolutionize brain tumor diagnosis and treatment. As the first application, it accelerates the development of advanced image analysis algorithms for brain tumor detection and segmentation, whether based on deep learning or energy minimization. With the growing number of segmentation algorithms, there is a need for comprehensive evaluation, and this dataset provides a standardized metric for rigorous testing, propelling advancements in brain cancer treatment. Therefore, the proposed dataset offers an

Table 7.1: Quality control grade chart for segmentation masks before, during, and after resection. Grades of the two neurosurgeons are given side-by-side in each cell. For Case 11 (during and after resection) and Case 15 (during resection), the resection cavity was not labeled (see section 7.2.3).

Validation Scores								
	Before Resection				During Resection		After Resection	
Patient ID	SMT	IdT	ExT	IdS	IdR	IdS	IdR	IdS
1	4 4	4 5	4 3	3 4	3 3	4 4	3 4	4 4
2	4 4	4 3	3 5	4 4	3 4	4 4	3 5	4 5
3	4 4	3 3	4 3	3 3	3 4	3 3	4 4	4 3
4	3 4	3 5	4 5	3 4	3 4	4 4	4 5	3 4
5	3 4	4 4	4 4	3 4	3 5	3 4	4 5	4 3
6	3 4	4 5	3 5	3 4	4 5	4 4	4 5	4 4
7	3 4	3 5	4 4	3 4	3 3	4 4	4 5	4 4
8	4 3	4 4	4 5	3 4	3 5	3 4	3 5	4 4
11	4 4	4 4	3 5	4 4	- -	4 3	- -	4 3
12	3 4	4 3	4 3	4 4	3 4	3 4	4 5	4 5
13	4 4	4 4	3 4	4 3	3 5	4 5	3 5	4 3
14	4 4	3 4	4 4	4 4	4 3	4 4	4 4	3 4
15	3 4	3 3	4 3	4 4	- -	4 4	4 4	4 4
16	4 4	4 4	4 4	4 4	3 4	4 4	3 4	3 4
17	3 4	4 5	4 5	4 4	4 5	4 4	3 4	4 4
18	4 4	3 4	3 4	4 4	3 4	3 4	3 5	3 5
19	3 4	3 5	3 5	4 3	3 5	3 5	3 5	4 4
21	3 4	3 4	4 4	3 4	4 5	4 4	4 5	4 4
23	3 4	3 5	3 5	3 4	3 3	4 4	3 5	4 4
24	4 4	4 4	3 5	3 5	3 4	4 4	3 4	4 4
25	4 4	3 4	3 4	3 4	4 3	3 4	3 4	3 3
26	4 5	4 5	3 5	3 5	3 5	3 5	3 5	3 3
27	4 5	3 3	3 5	4 4	3 3	3 4	3 5	3 4

SMT: smoothness of the boundaries, IdT: identification of tumoral tissues, ExT: exclusion of non-cancerous tissues, IdS: identification of sulci and falx. IdR: identification of resection cavity

opportunity for both technical and clinical communities to rigorously test their algorithms. Additionally, it offers a unique opportunity for algorithms to excel in multi-instance detection and segmentation, enhancing performance beyond binary segmentation tasks. Multiple methodologies are available for binary segmentation problems; however, recent studies suggest that integrating instances into deep learning algorithms not only enhances performance through parallel multi-instance segmentation but also achieves a comprehensive improvement overall [36]. Consequently, this dataset acts as a catalyst in refining computer-aided diagnosis (CAD) systems, enabling multi-instance and multi-organ analyses, thereby revolutionizing

brain tumor diagnosis and treatment.

The second application is transformative, focusing on developing and validating segmentation-based registration algorithms to address brain shift challenges during surgery. Brain shift, involving tissue deformation and displacement, poses precision hurdles in surgery. Integrating this dataset into registration algorithms provides them with expert-annotated tumor segmentations as a foundation. These resources empower algorithms to dynamically recalibrate pre-operative images in real-time, aligning them with evolving intraoperative conditions. The result is an advanced neuronavigation system, offering surgeons accurate, real-time patient anatomical visualization. This leads to enhanced resection control, minimizing structural damage risk and optimizing tumor removal. The synergy between segmentation and registration algorithms has the potential to redefine neurosurgery, equipping surgeons with a powerful tool to navigate brain shift complexities, ultimately ensuring safer surgeries, better patient outcomes, and improved resection control. This advancement holds promise for revolutionizing the field of neurosurgery.

The proposed RESECT-SEG dataset stands as a pivotal resource for advancing image processing techniques in neurosurgery. While it offers valuable segmentations of cerebral structures, it's essential to acknowledge its limitations. One such concern is the potential lack of representation of diverse clinical scenarios. Ensuring the dataset encapsulates a broad spectrum of anatomical variations, tumor types, and imaging modalities is crucial for its relevance and applicability in broader contexts. Evaluating the generalizability of proposed techniques beyond the confines of the RESECT-SEG dataset is imperative. Nevertheless, by adhering to rigorous evaluation protocols, conducting thorough validation processes, and maintaining transparent reporting standards, the RESECT-SEG dataset can significantly improve its credibility and impact in advancing neurosurgical image processing techniques.

7.5 Conclusion

In this study, the most comprehensive and validated expert delineations of cerebral structures within iUS images from the RESECT database were proposed. The primary focus lay in outlining tumor boundaries within pre-resection iUS 3D volumes and tracking the resection cavity both during and post-surgery. Additionally, delineated *sulci* and the *falx cerebri* were further provided to enhance surgical precision. This dataset presents an invaluable resource for both the training and evaluation of DL-based segmentation algorithms and registration methodologies, thereby presenting a rigorous challenge to their capabilities. This collective

effort is poised to catalyze advancements in brain tumor treatment and surgical interventions, ultimately benefiting patients and furthering the realm of medical science.

Chapter 8

Conclusions and Future Work

This thesis explored new techniques developed for the segmentation and classification of US images that have been submitted for review or published in peer-reviewed journals and conferences. In each chapter, we have detailed our achievements and contributions, along with potential future work specific to each area of research. In this chapter, we aim to provide a general overview of the entire thesis and present a broader perspective on the potential directions for future work. This holistic view will highlight the overarching themes and opportunities for further advancements across the various aspects of our research.

8.1 Conclusions

Clinical US imaging faces significant dataset limitations, primarily because annotating these images is labor-intensive and relies on expert radiologists. Additionally, hospital restrictions on data sharing due to privacy policies further hinder the progress of DL-based algorithms in this area. In light of these challenges, this thesis has focused on creating DL-based methodologies tailored to different US applications, especially when limited data is available. Segmentation of US images presents significant challenges, but it offers a wider array of applications. It is particularly essential in guided surgeries and enables the tracking and monitoring of changes in organ geometry during treatments. In this thesis, by prioritizing segmentation, we aimed to address the specific challenges associated with US applications and to pave the way for more effective clinical solutions. Additionally, we engaged in the classification and regression of US images since, from a development perspective, classification is closely related to segmentation. In classification, we assign labels to entire images, while in segmentation, labels are applied at the pixel level. Classification is especially valuable

for detecting abnormalities during patient assessments and diagnostic processes. We also focused on regression analysis, which assigns a value to each image. Unlike classification, which provides discrete integer labels, regression yields continuous values. Together, these methodologies enhance our understanding and application of US imaging in clinical settings.

In Ch. 2, we explored critical considerations for designing deep learning (DL) segmentation networks by addressing two key factors: the choice of data for pre-training and the size of the receptive field when designing the network architecture. First, we proposed the use of US simulation images for pre-training, demonstrating that pre-training on simulated data effectively improves network performance on real phantom data and is preferable to using natural images when working with limited *in vivo* data [14, 15]. Second, we analyzed the impact of dilated convolution and pooling layer size in the U-Net architecture design to adjust the proper receptive field, emphasizing that these adjustments were crucial for designing computationally efficient networks[16]. Our findings suggest that receptive field size, rather than network depth alone, should guide the design of U-Net-based architectures. In 3D ultrasound (US) uterus data in Ch. 3, where only 8 scans of patients were available, we employed 2D models to overcome the limitation of available 3D scans[19]. We showed that MobileNet-v2 was suitable for clinical use. However, it was noted that the network performed inadequately on slices near the uterus's boundaries. To further overcome segmentation challenges, particularly when data availability is limited, in Ch. **??**, we employed knowledge distillation (KD) from teacher to student model using well-defined KD pathways. This approach allowed us to effectively transfer knowledge from a well-trained teacher model to a smaller student model with only 0.82 million trainable parameters, which is crucial when training larger models becomes difficult due to the scarcity of data. Our segmentation results can be validated using a more comprehensive database of US image segmentations [80] and 2D echocardiography US [122]. In Ch. 5, we investigated the fact that increasing the number of classes in breast classification led to better performance of the deep learning networks. We further proposed a novel strategy in adding the background of US images as an additional class [17]. In Ch. 6, we integrated advanced techniques in segmentation and classification to propose a DeepSarc-US framework for evaluating sarcopenia [22]. DeepSarc was developed to measure QMT in clinical settings. We compared various deep learning models and results indicated that there was no significant difference in performance between CNN and transformer models using the proposed framework. We demonstrated that simplifying the task and pretraining on easier tasks could enhance model performance, particularly when data was limited. Given the scarcity of datasets for intra-operative US images of brain

tumors, in Ch. 7, we developed detailed manual annotations for these images [21]. We believe that making our dataset publicly available to researchers will significantly advance the field, providing valuable resources for improving algorithms and techniques in brain tumor treatment and surgical interventions.

8.2 Future Work

In future work, expanding the scope of US image segmentation and classification is crucial, particularly given the challenges posed by limited data availability. In Chapter 2, we demonstrated the effectiveness of simulation data in addressing these challenges. Chapter 3 further explored the use of 2D models for 3D uterus scans to augment the amount of labeled data. A promising direction for overcoming the limitations of small datasets involves investigating more advanced techniques, such as generative adverserial nets (GANs) [73], diffusion models (DM) [210], and etc., in data augmentation and synthetic data generation. By developing more realistic and diverse simulated US images, we can enhance the pre-training process even further. Additionally, generative models can be employed to create synthetic US images that closely mimic real-world variations, thereby providing additional training data that can improve the robustness of deep learning models. Two promising generative models in computer vision are GANs and DM. Both GANs and DM have been recently applied to US imaging with success [6, 7, 76, 77]. GANs and DMs are particularly advantageous in medical imaging, where acquiring large amounts of labeled data remains a significant challenge.

Moreover, exploring semi-supervised and unsupervised methods could further optimize model performance [121]. These methods allow for the utilization of unlabeled data, which is often more accessible than labeled data, to refine the learning process. Semi-supervised learning can enhance the model's generalization ability by combining limited labeled data with a larger pool of unlabeled data [226]. In contrast, unsupervised methods can uncover patterns and features within the data without relying on labels.

Another vital avenue for future work is minimizing the challenges associated with small datasets by focusing on training smaller networks. This approach is particularly valuable in scenarios with limited computational resources or where extensive datasets are not feasible for training. In Chapter 2, we demonstrated how adjusting the receptive field size through proper network design can optimize small networks. Additionally, in Chapter 4, we explored KD as a method for training small models. Future work could involve refining KD frameworks by experimenting with various teacher-student architectures and exploring new KD pathways

and student model designs to optimize knowledge transfer from large, well-trained models to smaller, more efficient ones.

In Chapter 6, the proposed DeepSarc framework was employed to measure muscle thickness. An extension of this framework could involve incorporating additional biomarkers, such as hand grip strength, clinical frailty scales, or frailty indices, for a more comprehensive frailty assessment. This would introduce additional complexity to the training process, making the investigation of state-of-the-art deep models a potential future direction. Furthermore, in real clinical trials, especially with frail patients who face physical limitations, data acquisition is limited. In such cases, semi-supervised algorithms could prove to be invaluable for further exploration.

Additionally, by making our annotated dataset of iUS images of brain tumors, as proposed in Chapter 7, publicly available, we hope to inspire further innovation in the field. This dataset will enable the development of more effective deep-learning models for medical image analysis, ultimately contributing to improved clinical outcomes and advancing ultrasound-based diagnostics and interventions.

Finally, from an implementation perspective, all segmentation, classification, and regression models utilize an encoder that serves as the feature extractor. For the developed and proposed models, an intriguing avenue for future work would be to investigate the development of NLP-based encoders such as CLIP encoders [184]. These encoders could facilitate user interaction by integrating user requests with model performance, thereby enhancing the overall effectiveness and adaptability of the models. Additionally, integrating Segment Anything Models (SAM) [117] can be beneficial for segmentation tasks. A promising direction with significant applications in the medical US, particularly in clinical trials, involves combining CLIP with SAM models. For instance, when a clinician requests segmentation of a specific tissue, this interaction could dynamically adjust the model's application. A model trained on multiple instances could then respond to physician requests, tailoring its output to meet specific clinical needs and improving the relevance and accuracy of the segmentation process.

References

[1] Nabila Abraham and Naimul Mefraz Khan. A novel focal tversky loss function with improved attention u-net for lesion segmentation. In *2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019)*, pages 683–687. IEEE, 2019.

[2] Jonathan Afilalo, Karen P Alexander, Michael J Mack, Mathew S Maurer, Philip Green, Larry A Allen, Jeffrey J Popma, Luigi Ferrucci, and Daniel E Forman. Frailty assessment in the cardiovascular care of older adults. *Journal of the American College of Cardiology*, 63(8):747–762, 2014.

[3] Jonathan Afilalo, Aayushi Joshi, and Rita Mancini. If you cannot measure frailty, you cannot improve it, 2019.

[4] Mina Amiri, Rupert Brooks, Bahareh Behboodi, and Hassan Rivaz. Two-stage ultrasound image segmentation using u-net and test time augmentation. *International journal of computer assisted radiology and surgery*, 15(6):981–988, 2020.

[5] Emran Mohammad Abu Anas, Parvin Mousavi, and Purang Abolmaesumi. A deep learning approach for real time prostate segmentation in freehand ultrasound guided biopsy. *Medical image analysis*, 48:107–116, 2018.

[6] Hojat Asgariandehkordi, Sobhan Goudarzi, Adrian Basarab, and Hassan Rivaz. Deep ultrasound denoising using diffusion probabilistic models. In *2023 IEEE International Ultrasonics Symposium (IUS)*, pages 1–4. IEEE, 2023.

[7] Hojat Asgariandehkordi, Sobhan Goudarzi, Mostafa Sharifzadeh, Adrian Basarab, and Hassan Rivaz. Denoising plane wave ultrasound images using diffusion probabilistic models. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 2024.

[8] Jimmy Ba and Rich Caruana. Do deep nets really need to be deep? *Advances in neural information processing systems*, 27, 2014.

[9] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495, 2017.

[10] Sathiyabhama Balasubramaniam, Yuvarajan Velmurugan, Dhayanithi Jaganathan, and Seshathiri Dhanasekaran. A modified lenet cnn for breast cancer diagnosis in ultrasound images. *Diagnostics*, 13(17):2746, 2023.

[11] Dhiego Chaves De Almeida Bastos, Parikshit Juvekar, Yanmei Tie, Nick Jowkar, Steve Pieper, Willam M Wells, Wenya Linda Bi, Alexandra Golby, Sarah Frisken, and Tina Kapur. Challenges and opportunities of intraoperative 3d ultrasound with neuronavigation in relation to intraoperative mri. *Frontiers in Oncology*, 11:1463, 2021.

[12] Dhiego Chaves de Almeida Bastos, Parikshit Juvekar, Yanmei Tie, Nick Jowkar, Steve Pieper, William Mercer Wells, Wenya Linda Bi, Alexandra Golby, Sarah Frisken, and Tina Kapur. A clinical perspective on the use of intraoperative 3d ultrasound with neuronavigation and intraoperative mri. *Frontiers in Oncology*, 11:1463, 2021.

[13] Anton S Becker, Michael Mueller, Elina Stoffel, Magda Marcon, Soleen Ghafoor, and Andreas Boss. Classification of breast cancer in ultrasound imaging using a generic deep learning analysis software: a pilot study. *The British journal of radiology*, 91 (1083):20170576, 2018.

[14] Bahareh Behboodi and Hassan Rivaz. Ultrasound segmentation using u-net: learning from simulated data and testing on real data. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 6628–6631. IEEE, 2019.

[15] Bahareh Behboodi, Mina Amiri, Rupert Brooks, and Hassan Rivaz. Breast lesion segmentation in ultrasound images with limited annotated data. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pages 1834–1837. IEEE, 2020.

[16] Bahareh Behboodi, Maryse Fortin, Clyde J Belasso, Rupert Brooks, and Hassan Rivaz. Receptive field size as a key design parameter for ultrasound image segmentation with u-net. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 2117–2120. IEEE, 2020.

[17] Bahareh Behboodi, Hamze Rasaee, Ali K. Z. Tehrani, and Hassan Rivaz. Deep classification of breast cancer in ultrasound images: more classes, better results with

multi-task learning. In Brett C. Byram and Nicole V. Ruiter, editors, *Medical Imaging 2021: Ultrasonic Imaging and Tomography*, volume 11602, page 116020S. International Society for Optics and Photonics, SPIE, 2021. doi: 10.1117/12.2581930. URL https://doi.org/10.1117/12.2581930.

[18] Bahareh Behboodi, Hamze Rasaee, Ali KZ Tehrani, and Hassan Rivaz. Deep classification of breast cancer in ultrasound images: more classes, better results with multi-task learning. In *Medical Imaging 2021: Ultrasonic Imaging and Tomography*, volume 11602, page 116020S. International Society for Optics and Photonics, 2021.

[19] Bahareh Behboodi, Hassan Rivaz, Susan Lalondrelle, and Emma Harris. Automatic 3d ultrasound segmentation of uterus using deep learning. In *2021 IEEE International Ultrasonics Symposium (IUS)*, pages 1–4, 2021. doi: 10.1109/IUS52206.2021.9593671.

[20] Bahareh Behboodi, Rupert Brooks, and Hassan Rivaz. Optimizing knowledge distillation for efficient breast ultrasound image segmentation: Insights and performance enhancement. *Artificial Intelligence in Health*, 2024.

[21] Bahareh Behboodi, Francois-Xavier Carton, Matthieu Chabanas, Sandrine de Ribaupierre, Ole Solheim, Bodil K. R. Munkvold, Hassan Rivaz, Yiming Xiao, and Ingerid Reinertsen. Open access segmentations of intra-operative brain tumor ultrasound images. *Medical Physics*, 2024. ISSN 2076-3417. doi: 10.1002/mp.17317. URL https://aapm.onlinelibrary.wiley.com/doi/full/10.1002/mp.17317.

[22] Bahareh Behboodi, Jeremy Obrand, Jonathan Afilalo, and Hassan Rivaz. Deepsarc-us: A deep learning framework for assessing sarcopenia using ultrasound images. *Applied Sciences*, 14(15), 2024. ISSN 2076-3417. doi: 10.3390/app14156726. URL https://www.mdpi.com/2076-3417/14/15/6726.

[23] Clyde J Belasso, Bahareh Behboodi, Habib Benali, Mathieu Boily, Hassan Rivaz, and Maryse Fortin. Luminous database: lumbar multifidus muscle segmentation from ultrasound images. *BMC Musculoskeletal Disorders*, 21(1):1–11, 2020.

[24] Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning*, pages 41–48, 2009.

[25] Peng Bian, Xiyu Zhang, Ruihong Liu, Huijie Li, Qingqing Zhang, and Baoling Dai. Deep-learning-based color doppler ultrasound image feature in the diagnosis of elderly

patients with chronic heart failure complicated with sarcopenia. *Journal of Healthcare Engineering*, 2021, 2021.

[26] Lior Bibas, Eli Saleh, Samah Al-Kharji, Jessica Chetrit, Louis Mullie, Marcelo Cantarovich, Renzo Cecere, Nadia Giannetti, and Jonathan Afilalo. Muscle mass and mortality after cardiac transplantation. *Transplantation*, 102(12):2101–2107, 2018.

[27] Paul Blanc-Durand, J-B Schiratti, Kathryn Schutte, Paul Jehanno, Paul Herent, Frédéric Pigneur, Olivier Lucidarme, Y Benaceur, Alexandre Sadate, Alain Luciani, et al. Abdominal musculature segmentation and surface prediction from ct using deep learning for sarcopenia assessment. *Diagnostic and Interventional Imaging*, 101(12): 789–794, 2020.

[28] Alexander Buslaev, Vladimir I Iglovikov, Eugene Khvedchenya, Alex Parinov, Mikhail Druzhinin, and Alexandr A Kalinin. Albumentations: fast and flexible image augmentations. *Information*, 11(2):125, 2020.

[29] Michal Byra, Michael Galperin, Haydee Ojeda-Fournier, Linda Olson, Mary O'Boyle, Christopher Comstock, and Michael Andre. Breast mass classification in sonography with transfer learning using a deep convolutional neural network and color conversion. *Medical physics*, 46(2):746–755, 2019.

[30] Luca Canalini, Jan Klein, Dorothea Miller, and Ron Kikinis. Segmentation-based registration of ultrasound volumes for glioma resection in image-guided neurosurgery. *International journal of computer assisted radiology and surgery*, 14(10):1697–1713, 2019.

[31] Luca Canalini, Jan Klein, Dorothea Miller, and Ron Kikinis. Enhanced registration of ultrasound volumes by segmentation of resection cavity in neurosurgical procedures. *International Journal of Computer Assisted Radiology and Surgery*, 15(12):1963–1974, 2020.

[32] Zhantao Cao, Guowu Yang, Qin Chen, Xiaolong Chen, and Fengmao Lv. Breast tumor classification through learning from noisy labeled ultrasound images. *Medical Physics*, 47(3):1048–1057, 2020.

[33] François-Xavier Carton, Matthieu Chabanas, Florian Le Lann, and Jack H. Noble. Automatic segmentation of brain tumor resections in intraoperative ultrasound images

using U-Net. *Journal of Medical Imaging*, 7(3):1 – 15, 2020. doi: 10.1117/1.JMI.7.3.031503.

[34] Ruey-Feng Chang, Kuang-Che Chang-Chien, Hao-Jen Chen, Dar-Ren Chen, Etsuo Takada, and Woo Kyung Moon. Whole breast computer-aided screening using free-hand ultrasound. In *International Congress Series*, volume 1281, pages 1075–1080. Elsevier, 2005.

[35] Hila Chefer, Shir Gur, and Lior Wolf. Transformer interpretability beyond attention visualization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 782–791, 2021.

[36] Changhao Chen, Bing Wang, Chris Xiaoxuan Lu, Niki Trigoni, and Andrew Markham. A survey on deep learning for localization and mapping: Towards the age of spatial machine intelligence. *arXiv preprint arXiv:2006.12567*, 2020.

[37] Chung-Ming Chen, Yi-Hong Chou, Ko-Chung Han, Guo-Shian Hung, Chui-Mei Tiu, Hong-Jen Chiou, and See-Ying Chiou. Breast lesions on sonograms: computer-aided diagnosis with nearly setting-independent features and artificial neural networks. *Radiology*, 226(2):504–514, 2003.

[38] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*, 2017.

[39] Zhen Chen, Xiaoqing Guo, Peter YM Woo, and Yixuan Yuan. Super-resolution enhanced medical image diagnosis with sample affinity interaction. *IEEE Transactions on Medical Imaging*, 40(5):1377–1389, 2021.

[40] Heng-Da Cheng, Juan Shan, Wen Ju, Yanhui Guo, and Ling Zhang. Automated breast cancer detection and classification using ultrasound images: A survey. *Pattern recognition*, 43(1):299–317, 2010.

[41] Yu Cheng, Duo Wang, Pan Zhou, and Tao Zhang. A survey of model compression and acceleration for deep neural networks. *arXiv preprint arXiv:1710.09282*, 2017.

[42] Tsung-Chen Chiang, Yao-Sian Huang, Rong-Tai Chen, Chiun-Sheng Huang, and Ruey-Feng Chang. Tumor detection in automated breast ultrasound using 3-d cnn and

prioritized candidate aggregation. *IEEE transactions on medical imaging*, 38(1):240–249, 2019.

[43] Jui-Ying Chiao, Kuan-Yung Chen, Ken Ying-Kai Liao, Po-Hsin Hsieh, Geoffrey Zhang, and Tzung-Chi Huang. Detection and classification the breast tumors using mask r-cnn on sonograms. *Medicine*, 98(19), 2019.

[44] Kyunghyun Cho. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.

[45] Sophie Church, Emily Rogers, Kenneth Rockwood, and Olga Theou. A scoping review of the clinical frailty scale. *BMC geriatrics*, 20:1–18, 2020.

[46] Paul A Cohen, Anjua Jhingran, Ana Oaknin, and Lynette Denny. Cervical cancer. *The Lancet*, 393(10167):169–182, 2019.

[47] Marly Guimarães Fernandes Costa, João Paulo Mendes Campos, Gustavo de Aquino e Aquino, Wagner Coelho de Albuquerque Pereira, and Cícero Ferreira Fernandes Costa Filho. Evaluating the performance of convolutional neural networks with direct acyclic graph architectures in automatic segmentation of breast lesion in us images. *BMC Medical Imaging*, 19(1):1–13, 2019.

[48] AJ Cruz-Jentoft and J-P Michel. Sarcopenia: a useful paradigm for physical frailty. *European Geriatric Medicine*, 4(2):102–105, 2013.

[49] Alfonso J Cruz-Jentoft, Gülistan Bahat, Jürgen Bauer, Yves Boirie, Olivier Bruyère, Tommy Cederholm, Cyrus Cooper, Francesco Landi, Yves Rolland, Avan Aihie Sayer, et al. Sarcopenia: revised european consensus on definition and diagnosis. *Age and ageing*, 48(1):16–31, 2019.

[50] Abdulla A Damluji, Daniel E Forman, Sean Van Diepen, Karen P Alexander, Robert L Page, Scott L Hummel, Venu Menon, Jason N Katz, Nancy M Albert, Jonathan Afilalo, et al. Older adults in the cardiac intensive care unit: factoring geriatric syndromes in the management, prognosis, and process of care: a scientific statement from the american heart association. *Circulation*, 141(2):e6–e32, 2020.

[51] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[52] Jacob Devlin. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[53] KT Dilna and D Jude Hemanth. Fibroid detection in ultrasound uterus images using image processing. In *International Conference on Innovative Computing and Communications*, pages 173–179. Springer, 2020.

[54] Jianrui Ding, Heng-Da Cheng, Jianhua Huang, Jiafeng Liu, and Yingtao Zhang. Breast ultrasound image classification based on multiple-instance learning. *Journal of digital imaging*, 25(5):620–627, 2012.

[55] Wanli Ding, Heye Zhang, Shuxin Zhuang, Zhemin Zhuang, and Zhifan Gao. Multi-view stereoscopic attention network for 3d tumor classification in automated breast ultrasound. *Expert Systems with Applications*, 234:120969, 2023.

[56] Xuehai Ding, Yanting Liu, Junjuan Zhao, Ren Wang, Chengfan Li, Quanyong Luo, and Chentian Shen. A novel wavelet-transform-based convolution classification network for cervical lymph node metastasis of papillary thyroid carcinoma in ultrasound images. *Computerized Medical Imaging and Graphics*, 109:102298, 2023.

[57] Richard Dodds and Avan Aihie Sayer. Sarcopenia and frailty: new challenges for clinical practice. *Clinical medicine*, 16(5):455, 2016.

[58] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[59] Qi Dou, Quande Liu, Pheng Ann Heng, and Ben Glocker. Unpaired multi-modal segmentation via knowledge distillation. *IEEE transactions on medical imaging*, 39 (7):2415–2425, 2020.

[60] J Egger, T Kapur, A Fedorov, S Pieper, JV Miller, H Veeraraghavan, B Freisleben, AJ Golby, C Nimsky, and R Kikinis. Gbm volumetry using the 3d slicer medical image computing platform. sci rep 3: 1364, 2013.

[61] Jianan Fan, Dongnan Liu, Hang Chang, Heng Huang, Mei Chen, and Weidong Cai. Taxonomy adaptive cross-domain adaptation in medical imaging via optimization trajectory distillation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 21174–21184, 2023.

[62] Andriy Fedorov, Reinhard Beichel, Jayashree Kalpathy-Cramer, Julien Finet, Jean-Christophe Fillion-Robin, Sonia Pujol, Christian Bauer, Dominique Jennings, Fiona Fennessy, Milan Sonka, et al. 3d slicer as an image computing platform for the quantitative imaging network. *Magnetic resonance imaging*, 30(9):1323–1341, 2012.

[63] Rosie Fountotos, Haroon Munir, Michael Goldfarb, Sandra Lauck, Dae Kim, Louis Perrault, Rakesh Arora, Emmanuel Moss, Lawrence G Rudski, Melissa Bendayan, et al. Prognostic value of handgrip strength in older adults undergoing cardiac surgery. *Canadian Journal of Cardiology*, 2021.

[64] Jun Fu, Jing Liu, Haijie Tian, Yong Li, Yongjun Bao, Zhiwei Fang, and Hanqing Lu. Dual attention network for scene segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3146–3154, 2019.

[65] Tomoyuki Fujioka, Kazunori Kubota, Jen Feng Hsu, Ruey Feng Chang, Terumasa Sawada, Yoshimi Ide, Kanae Taruno, Meishi Hankyo, Tomoko Kurita, Seigo Nakamura, et al. Examining the effectiveness of a deep learning-based computer-aided breast cancer detection system for breast ultrasound. *Journal of Medical Ultrasonics*, 50(4): 511–520, 2023.

[66] Chengling Gao, Hailiang Ye, Feilong Cao, Chenglin Wen, Qinghua Zhang, and Feng Zhang. Multiscale fused network with additive channel–spatial attention for image segmentation. *Knowledge-Based Systems*, 214:106754, 2021.

[67] Zhifan Gao, Jonathan Chung, Mohamed Abdelrazek, Stephanie Leung, William Kongto Hau, Zhanchao Xian, Heye Zhang, and Shuo Li. Privileged modality distillation for vessel border detection in intracoronary imaging. *IEEE transactions on medical imaging*, 39(5):1524–1534, 2019.

[68] Ian J. Gerard, Marta Kersten-Oertel, Jeffery A. Hall, Denis Sirhan, and D. Louis Collins. Brain shift in neuronavigation of brain tumors: An updated review of intra-operative ultrasound applications. *Frontiers in Oncology*, 10, 2021. ISSN 2234-943X. doi: 10.3389/fonc.2020.618837. URL https://www.frontiersin.org/articles/10.3389/fonc.2020.618837.

[69] Behnaz Gheflati and Hassan Rivaz. Vision transformers for classification of breast ultrasound images. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 480–483. IEEE, 2022.

[70] Dipannita Ghosh, Amish Kumar, Palash Ghosal, Tamal Chowdhury, Anup Sadhu, and Debashis Nandi. Breast lesion segmentation in ultrasound images using deep convolutional neural networks. In *2020 IEEE Calcutta Conference (CALCON)*, pages 318–322. IEEE, 2020.

[71] Walter Gómez, Wagner Coelho Albuquerque Pereira, and Antonio Fernando C Infantosi. Analysis of co-occurrence texture statistics as a function of gray-level quantization for classifying breast ultrasound. *IEEE transactions on medical imaging*, 31(10):1889–1899, 2012.

[72] Yunchao Gong, Liu Liu, Ming Yang, and Lubomir Bourdev. Compressing deep convolutional networks using vector quantization. *arXiv preprint arXiv:1412.6115*, 2014.

[73] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

[74] Jianping Gou, Baosheng Yu, Stephen J Maybank, and Dacheng Tao. Knowledge distillation: A survey. *International Journal of Computer Vision*, 129(6):1789–1819, 2021.

[75] Sobhan Goudarzi and Hassan Rivaz. Deep reconstruction of high-quality ultrasound images from raw plane-wave data: A simulation and in vivo study. *Ultrasonics*, 125: 106778, 2022.

[76] Sobhan Goudarzi, Amir Asif, and Hassan Rivaz. Multi-focus ultrasound imaging using generative adversarial networks. In *2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019)*, pages 1118–1121. IEEE, 2019.

[77] Sobhan Goudarzi, Amir Asif, and Hassan Rivaz. Fast multi-focus ultrasound image recovery using generative adversarial networks. *IEEE Transactions on Computational Imaging*, 6:1272–1284, 2020.

[78] Sobhan Goudarzi, Adrian Basarab, and Hassan Rivaz. Inverse problem of ultrasound beamforming with denoising-based regularized solutions. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 69(10):2906–2916, 2022.

[79] Sobhan Goudarzi, Adrian Basarab, and Hassan Rivaz. A unifying approach to inverse problems of ultrasound beamforming and deconvolution. *IEEE Transactions on Computational Imaging*, 9:197–209, 2023.

[80] Sobhan Goudarzi, Jesse Whyte, Mathieu Boily, Anna Towers, Robert D Kilgour, and Hassan Rivaz. Segmentation of arm ultrasound images in breast cancer-related lymphedema: A database and deep learning algorithm. *IEEE Transactions on Biomedical Engineering*, 70(9):2552–2563, 2023.

[81] Holly Gwyther, Rachel Shaw, Eva-Amparo Jaime Dauden, Barbara D'Avanzo, Donata Kurpas, Maria Bujnowska-Fedak, Tomasz Kujawa, Maura Marcucci, Antonio Cano, and Carol Holland. Understanding frailty: a qualitative study of european healthcare policy-makers' approaches to frailty screening and management. *BMJ open*, 8(1): e018653, 2018.

[82] M Halliwell. A tutorial on ultrasonic physics and imaging techniques. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, 224(2):127–142, 2010.

[83] Seokmin Han, Ho-Kyung Kang, Ja-Yeon Jeong, Moon-Ho Park, Wonsik Kim, Won-Chul Bang, and Yeong-Kyeong Seong. A deep learning framework for supporting the classification of breast lesions in ultrasound images. *Physics in Medicine & Biology*, 62(19):7714, 2017.

[84] Stephen Hanson and Lorien Pratt. Comparing biases for minimal network construction with back-propagation. *Advances in neural information processing systems*, 1, 1988.

[85] Hoda S Hashemi, Stefanie Fallone, Mathieu Boily, Anna Towers, Robert D Kilgour, and Hassan Rivaz. Assessment of mechanical properties of tissue in breast cancer-related lymphedema using ultrasound elastography. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 66(3):541–550, 2018.

[86] Mohamed A Hassanien, Vivek Kumar Singh, Domenec Puig, and Mohamed Abdel-Nasser. Predicting breast tumor malignancy using deep convnext radiomics and quality-based score pooling in ultrasound sequences. *Diagnostics*, 12(5):1053, 2022.

[87] Babak Hassibi and David Stork. Second order derivatives for network pruning: Optimal brain surgeon. *Advances in neural information processing systems*, 5, 1992.

[88] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015.

[89] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[90] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.

[91] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16000–16009, 2022.

[92] Tong He, Chunhua Shen, Zhi Tian, Dong Gong, Changming Sun, and Youliang Yan. Knowledge adaptation for efficient semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 578–587, 2019.

[93] Mohammad Hesam Hesamian, Wenjing Jia, Xiangjian He, and Paul Kennedy. Deep learning techniques for medical image segmentation: achievements and challenges. *Journal of digital imaging*, 32(4):582–596, 2019.

[94] Geoffrey Hinton, Oriol Vinyals, Jeff Dean, et al. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2(7), 2015.

[95] Thi Kieu Khanh Ho and Jeonghwan Gwak. Utilizing knowledge distillation in deep learning for classification of chest x-ray abnormalities. *IEEE Access*, 8:160749–160761, 2020.

[96] Emiel O Hoogendijk, Jonathan Afilalo, Kristine E Ensrud, Paul Kowal, Graziano Onder, and Linda P Fried. Frailty: implications for clinical practice and public health. *The Lancet*, 394(10206):1365–1375, 2019.

[97] Andrew Howard, Andrey Zhmoginov, Liang-Chieh Chen, Mark Sandler, and Menglong Zhu. Inverted residuals and linear bottlenecks: Mobile networks for classification, detection and segmentation. In *Proc. CVPR*, pages 4510–4520, 2018.

[98] Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, et al. Searching for

mobilenetv3. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1314–1324, 2019.

[99] Yi-Hsuan Hsiao, Shun-Fa Yang, Ya-Hui Chen, Tze-Ho Chen, Horng-Der Tsai, Ming-Chih Chou, and Pang-Hsin Chou. Updated applications of ultrasound in uterine cervical cancer. *Journal of Cancer*, 12(8):2181, 2021.

[100] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018.

[101] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.

[102] Haibo Huang, Haobo Chen, Haohao Xu, Ying Chen, Qihui Yu, Yehua Cai, and Qi Zhang. Cross-tissue/organ transfer learning for the segmentation of ultrasound images using deep residual u-net. *Journal of Medical and Biological Engineering*, 41 (2):137–145, 2021.

[103] Yunzhi Huang, Luyi Han, Haoran Dou, Honghao Luo, Zhen Yuan, Qi Liu, Jiang Zhang, and Guangfu Yin. Two-stage cnns for computerized bi-rads categorization in breast ultrasound images. *Biomedical engineering online*, 18(1):1–18, 2019.

[104] Sumaira Hussain, Xiaoming Xi, Inam Ullah, Yongjian Wu, Chunxiao Ren, Zhao Lianzheng, Cuihuan Tian, and Yilong Yin. Contextual level-set method for breast tumor segmentation. *IEEE Access*, 8:189343–189353, 2020.

[105] Pavel Iakubovskii. Segmentation models pytorch. `https://github.com/qubvel/segmentation_models.pytorch`, 2019.

[106] Eddy Ilg, Nikolaus Mayer, Tonmoy Saikia, Margret Keuper, Alexey Dosovitskiy, and Thomas Brox. Flownet 2.0: Evolution of optical flow estimation with deep networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2462–2470, 2017.

[107] Elisee Ilunga-Mbuyamba, Juan Gabriel Avina-Cervantes, Dirk Lindner, Felix Arlt, Jean Fulbert Ituna-Yudonago, and Claire Chalopin. Patient-specific model-based segmentation of brain tumors in 3d intraoperative ultrasound images. *International journal of computer assisted radiology and surgery*, 13(3):331–342, 2018.

[108] Fabian Isensee, Jens Petersen, Andre Klein, David Zimmerer, Paul F Jaeger, Simon Kohl, Jakob Wasserthal, Gregor Koehler, Tobias Norajitra, Sebastian Wirkert, et al. nnu-net: Self-adapting framework for u-net-based medical image segmentation. *arXiv preprint arXiv:1809.10486*, 2018.

[109] Fabian Isensee, Paul F Jaeger, Simon AA Kohl, Jens Petersen, and Klaus H Maier-Hein. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, 18(2):203–211, 2021.

[110] Noushin Jafarpisheh, Timothy J Hall, Hassan Rivaz, and Ivan M Rosado-Mendez. Analytic global regularized backscatter quantitative ultrasound. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 68(5):1605–1617, 2020.

[111] Jørgen Arendt Jensen. Field: A program for simulating ultrasound systems. In *10TH Nordicbaltic Conference on Biomedical Imaging, VOL. 4, Supplement 1, Part 1: 351–353*. Citeseer, 1996.

[112] Jørgen Arendt Jensen and Niels Bruun Svendsen. Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 39(2):262–267, 1992.

[113] Aayushi Joshi, Rita Mancini, Stephan Probst, Gad Abikhzer, Yves Langlois, Jean-Francois Morin, Lawrence G Rudski, and Jonathan Afilalo. Sarcopenia in cardiac surgery: Dual x-ray absorptiometry study from the mcgill frailty registry. *American Heart Journal*, 2021.

[114] Parikshit Juvekar, Reuben Dorent, Fryderyk Kögl, Erickson Torio, Colton Barr, Laura Rigolo, Colin Galvin, Nick Jowkar, Anees Kazi, Nazim Haouchine, et al. Remind: The brain resection multimodal imaging database. *medRxiv*, 2023.

[115] Kyungsang Kim, Fabiola Macruz, Dufan Wu, Christopher Bridge, Suzannah McKinney, Ahad Alhassan Al Saud, Elshaimaa Sharaf, Adam Pely, Paul Danset, Tom Duffy, et al. Point-of-care ai-assisted stepwise ultrasound pneumothorax diagnosis. *Physics in Medicine & Biology*, 68(20):205013, 2023.

[116] DP Kingma and J Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[117] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4015–4026, 2023.

[118] Gotaro Kojima, Steve Iliffe, and Kate Walters. Frailty index as a predictor of mortality: a systematic review and meta-analysis. *Age and ageing*, 47(2):193–200, 2018.

[119] Terry K Koo and Mae Y Li. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine*, 15(2):155–163, 2016.

[120] Viksit Kumar, Jeremy M Webb, Adriana Gregory, Max Denis, Duane D Meixner, Mahdi Bayat, Dana H Whaley, Mostafa Fatemi, and Azra Alizad. Automated and real-time segmentation of suspicious breast masses using convolutional neural network. *PloS one*, 13(5):e0195816, 2018.

[121] Ali KZ Tehrani, Morteza Mirzaei, and Hassan Rivaz. Semi-supervised training of optical flow convolutional neural networks in ultrasound elastography. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 504–513. Springer, 2020.

[122] Sarah Leclerc, Erik Smistad, Joao Pedrosa, Andreas Østvik, Frederic Cervenansky, Florian Espinosa, Torvald Espeland, Erik Andreas Rye Berg, Pierre-Marc Jodoin, Thomas Grenier, et al. Deep learning for segmentation using an open large-scale dataset in 2d echocardiography. *IEEE transactions on medical imaging*, 38(9):2198–2210, 2019.

[123] Yann LeCun, John Denker, and Sara Solla. Optimal brain damage. *Advances in neural information processing systems*, 2, 1989.

[124] Haeyun Lee, Jinhyoung Park, and Jae Youn Hwang. Channel attention module with multiscale grid average pooling for breast cancer segmentation in an ultrasound image. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 67(7):1344–1353, 2020.

[125] Kyungsu Lee, Haeyun Lee, Georges El Fakhri, Jonghye Woo, and Jae Youn Hwang. Self-supervised domain adaptive segmentation of breast cancer via test-time fine-tuning. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 539–550. Springer, 2023.

[126] Seung Hoo Lee and Hyun Sik Gong. Measurement and interpretation of handgrip strength for research on sarcopenia and osteoporosis. *Journal of bone metabolism*, 27 (2):85, 2020.

[127] Jun Li, Junyu Chen, Yucheng Tang, Ce Wang, Bennett A Landman, and S Kevin Zhou. Transforming medical imaging with transformers? a comparative review of key properties, current progresses, and future perspectives. *Medical Image Analysis*, page 102762, 2023.

[128] Kang Li, Lequan Yu, Shujun Wang, and Pheng-Ann Heng. Towards cross-modality medical image segmentation with online mutual knowledge distillation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 775–783, 2020.

[129] Lele Li, Ziling Wu, Juan Liu, Lang Wang, Yu Jin, Peng Jiang, Jing Feng, and Meng Wu. Cross-attention based multi-scale feature fusion vision transformer for breast ultrasound image classification. In *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 1616–1619. IEEE, 2022.

[130] Zeju Li, Konstantinos Kamnitsas, and Ben Glocker. Overfitting of neural nets under class imbalance: Analysis and improvements for segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 402–410. Springer, 2019.

[131] Yuanhao Liang, Ran He, Yongshuai Li, and Zhili Wang. Simultaneous segmentation and classification of breast lesions from ultrasound images using mask r-cnn. In *2019 IEEE International Ultrasonics Symposium (IUS)*, pages 1470–1472. IEEE, 2019.

[132] David Liljequist, Britt Elfving, and Kirsti Skavberg Roaldsen. Intraclass correlation–a discussion and demonstration of basic features. *PloS one*, 14(7):e0219854, 2019.

[133] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017.

[134] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, and Clara I Sánchez. A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88, 2017.

[135] Xilun Liu and Mohamed Almekkawy. Ultrasound super resolution using vision transformer with convolution projection operation. In *2022 IEEE International Ultrasonics Symposium (IUS)*, pages 1–4. IEEE, 2022.

[136] Yifan Liu, Ke Chen, Chris Liu, Zengchang Qin, Zhenbo Luo, and Jingdong Wang. Structured knowledge distillation for semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2604–2613, 2019.

[137] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021.

[138] Ze Liu, Han Hu, Yutong Lin, Zhuliang Yao, Zhenda Xie, Yixuan Wei, Jia Ning, Yue Cao, Zheng Zhang, Li Dong, et al. Swin transformer v2: Scaling up capacity and resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12009–12019, 2022.

[139] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11976–11986, 2022.

[140] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[141] Jonathan L Long, Ning Zhang, and Trevor Darrell. Do convnets learn correspondence? In *Advances in neural information processing systems*, pages 1601–1609, 2014.

[142] Ian Loram, Abdul Siddique, María B Sánchez, Pete Harding, Monty Silverdale, Christopher Kobylecki, and Ryan Cunningham. Objective analysis of neck muscle boundaries for cervical dystonia using ultrasound imaging and deep learning. *IEEE journal of biomedical and health informatics*, 24(4):1016–1027, 2020.

[143] Meng Lou, Jie Meng, Yunliang Qi, Xiaorong Li, and Yide Ma. Mcrnet: Multi-level context refinement network for semantic segmentation in breast ultrasound imaging. *Neurocomputing*, 470:154–169, 2022.

[144] David N Louis, Arie Perry, Pieter Wesseling, Daniel J Brat, Ian A Cree, Dominique Figarella-Branger, Cynthia Hawkins, H K Ng, Stefan M Pfister, Guido Reifenberger, Riccardo Soffietti, Andreas von Deimling, and David W Ellison. The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. *Neuro-Oncology*, 23(8):1231–1251, 06 2021. ISSN 1522-8517. doi: 10.1093/neuonc/noab106. URL https://doi.org/10.1093/neuonc/noab106.

[145] Wenjie Luo, Yujia Li, Raquel Urtasun, and Richard Zemel. Understanding the effective receptive field in deep convolutional neural networks. In *Advances in neural information processing systems*, pages 4898–4906, 2016.

[146] Karttikeya Mangalam and Mathieu Salzamann. On compressing u-net using knowledge distillation. *arXiv preprint arXiv:1812.00249*, 2018.

[147] Francesco Marzola, Nens van Alfen, Jonne Doorduin, and Kristen M Meiburger. Deep learning segmentation of transverse musculoskeletal ultrasound images for neuromuscular disease assessment. *Computers in Biology and Medicine*, 135:104623, 2021.

[148] Sarah A Mason, Ingrid M White, Susan Lalondrelle, Jeffrey C Bamber, and Emma J Harris. The stacked-ellipse algorithm: an ultrasound-based 3-d uterine segmentation tool for enabling adaptive radiotherapy for uterine cervix cancer. *Ultrasound in medicine & biology*, 46(4):1040–1052, 2020.

[149] Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging*, 34(10):1993–2024, 2014.

[150] Laurence Mercier, Rolando F Del Maestro, Kevin Petrecca, David Araujo, Claire Haegelen, and D Louis Collins. Online database of clinical mr and ultrasound images of brain tumors. *Medical physics*, 39(6Part1):3253–3261, 2012.

[151] Oleg V Michailovich and Allen Tannenbaum. Despeckling of medical ultrasound images. *ieee transactions on ultrasonics, ferroelectrics, and frequency control*, 53(1):64–78, 2006.

[152] Fausto Milletari, Seyed-Ahmad Ahmadi, Christine Kroll, Annika Plate, Verena E. Rozanski, Juliana Maiostre, Johannes Levin, Olaf Dietrich, Birgit Ertl-Wagner, Kai

Bötzel, and Nassir Navab. Hough-cnn: Deep learning for segmentation of deep brain regions in MRI and ultrasound. *CoRR*, abs/1601.07014, 2016. URL http://arxiv.org/abs/1601.07014.

[153] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)*, pages 565–571. IEEE, 2016.

[154] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 Fourth International Conference on 3D Vision (3DV)*, pages 565–571. IEEE, 2016.

[155] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.

[156] Woo Kyung Moon, Yan-Wei Lee, Hao-Hsiang Ke, Su Hyun Lee, Chiun-Sheng Huang, and Ruey-Feng Chang. Computer-aided diagnosis of breast ultrasound images using ensemble learning from convolutional neural networks. *Computer methods and programs in biomedicine*, 190:105361, 2020.

[157] Christopher L Moore and Joshua A Copel. Point-of-care ultrasonography. *New England Journal of Medicine*, 364(8):749–757, 2011.

[158] John E Morley, Bruno Vellas, G Abellan Van Kan, Stefan D Anker, Juergen M Bauer, Roberto Bernabei, Matteo Cesari, WC Chumlea, Wolfram Doehner, Jonathan Evans, et al. Frailty consensus: a call to action. *Journal of the American Medical Directors Association*, 14(6):392–397, 2013.

[159] Rand Muhtaseb and Mohammad Yaqub. Echocotr: Estimation of the left ventricular ejection fraction from spatiotemporal echocardiography. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part IV*, pages 370–379. Springer, 2022.

[160] Bodil Karoline Ravn Munkvold, Hans Kristian Bø, Asgeir Store Jakola, Ingerid Reinertsen, Erik Magnus Berntsen, Geirmund Unsgård, Sverre Helge Torp, and Ole Solheim. Tumor volume assessment in low-grade gliomas: A comparison of preoperative magnetic resonance imaging to c oregistered intraoperative 3-dimensional ultrasound recordings. *Neurosurgery*, 83(2):288–296, 2018. doi: 10.1093/neuros/nyx392.

[161] Arya Nabavi, Peter McL. Black, David T Gering, Carl-Fredrik Westin, Vivek Mehta, Richard S Pergolizzi Jr, Mathieu Ferrant, Simon K Warfield, Nobuhiko Hata, Richard B Schwartz, et al. Serial intraoperative magnetic resonance imaging of brain shift. *Neurosurgery*, 48(4):787–798, 2001.

[162] Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814, 2010.

[163] Ana I.L. Namburete, Weidi Xie, Mohammad Yaqub, Andrew Zisserman, and J. Alison Noble. Fully-automated alignment of 3d fetal brain ultrasound to a canonical reference space using multi-task learning. *Medical Image Analysis*, 46:1–14, 2018. ISSN 1361-8415. doi: https://doi.org/10.1016/j.media.2018.02.006. URL https://www.sciencedirect.com/science/article/pii/S1361841518300306.

[164] Sumanth Nandamuri, Debarghya China, Pabitra Mitra, and Debdoot Sheet. Sumnet: Fully convolutional model for fast segmentation of anatomical structures in ultrasound volumes. *arXiv preprint arXiv:1901.06920*, 2019.

[165] James O' Neill. An overview of neural network compression. *arXiv preprint arXiv:2006.03669*, 2020.

[166] Zhenyuan Ning, Ke Wang, Shengzhou Zhong, Qianjin Feng, and Yu Zhang. Cf2-net: Coarse-to-fine fusion convolutional network for breast ultrasound image segmentation. *arXiv preprint arXiv:2003.10144*, 2020.

[167] Alison Noble and Djamal Boukerroui. Ultrasound image segmentation: a survey. *IEEE Transactions on medical imaging*, 25(8):987–1010, 2006.

[168] J Alison Noble and Djamal Boukerroui. Ultrasound image segmentation: a survey. *IEEE Transactions on medical imaging*, 25(8):987–1010, 2006.

[169] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018.

[170] Antonio Omuro and Lisa M DeAngelis. Glioblastoma and other malignant gliomas: a clinical review. *Jama*, 310(17):1842–1850, 2013.

[171] Sonia H Contreras Ortiz, Tsuicheng Chiu, and Martin D Fox. Ultrasound image enhancement: A review. *Biomedical Signal Processing and Control*, 7(5):419–428, 2012.

[172] Abdeldjalil Ouahabi and Abdelmalik Taleb-Ahmed. Deep learning for real-time semantic segmentation: Application in ultrasound imaging. *Pattern Recognition Letters*, 144:27–34, 2021.

[173] Julia P Owen, Marian Blazes, Niranchana Manivannan, Gary C Lee, Sophia Yu, Mary K Durbin, Aditya Nair, Rishi P Singh, Katherine E Talcott, Alline G Melo, et al. Student becomes teacher: training faster deep learning lightweight networks for automated identification of optical coherence tomography b-scans of interest using a student-teacher framework. *Biomedical optics express*, 12(9):5387–5399, 2021.

[174] Megha J Padghamod and Jayanand P Gawande. Classification of ultrasonic uterine images. *Adv Res Electr Electron Eng*, 1(3):89–92, 2014.

[175] Chao Peng, Xiangyu Zhang, Gang Yu, Guiming Luo, and Jian Sun. Large kernel matters–improve semantic segmentation by global convolutional network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4353–4361, 2017.

[176] Emmanuel Pintelas, Ioannis E Livieris, Nikolaos Barotsis, George Panayiotakis, and Panagiotis Pintelas. An autoencoder convolutional neural network framework for sarcopenia detection based on multi-frame ultrasound image slices. In *IFIP International Conference on Artificial Intelligence Applications and Innovations*, pages 209–219. Springer, 2021.

[177] Tobias Pohlen, Alexander Hermans, Markus Mathias, and Bastian Leibe. Full-resolution residual networks for semantic segmentation in street scenes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4151–4160, 2017.

[178] Xiaofeng Qi, Lei Zhang, Yao Chen, Yong Pi, Yi Chen, Qing Lv, and Zhang Yi. Automated diagnosis of breast ultrasonography images using deep neural networks. *Medical image analysis*, 52:185–198, 2019.

[179] Dian Qin, Jia-Jun Bu, Zhe Liu, Xin Shen, Sheng Zhou, Jing-Jun Gu, Zhi-Hua Wang, Lei Wu, and Hui-Fen Dai. Efficient medical image segmentation based on knowledge distillation. *IEEE Transactions on Medical Imaging*, 40(12):3820–3831, 2021.

[180] Xiaolei Qu, Yao Shi, Yaxin Hou, and Jue Jiang. An attention-supervised full-resolution residual network for the segmentation of breast ultrasound images. *Medical physics*, 47(11):5702–5714, 2020.

[181] Xiaolei Qu, Hongyan Lu, Wenzhong Tang, Shuai Wang, Dezhi Zheng, Yaxin Hou, and Jue Jiang. A vgg attention vision transformer network for benign and malignant classification of breast ultrasound images. *Medical Physics*, 49(9):5787–5798, 2022.

[182] Reza Moradi Rad, Parvaneh Saeedi, Jason Au, and Jon Havelock. Trophectoderm segmentation in human embryo images via inceptioned u-net. *Medical Image Analysis*, page 101612, 2020.

[183] A Radford. Improving language understanding by generative pre-training. 2018.

[184] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR, 2021.

[185] Ingerid Reinertsen, D Louis Collins, and Simon Drouin. The essential role of open data and software for the future of ultrasound-based neuronavigation. *Frontiers in Oncology*, 10:3219, 2021.

[186] Adriana Romero, Nicolas Ballas, Samira Ebrahimi Kahou, Antoine Chassang, Carlo Gatta, and Yoshua Bengio. Fitnets: Hints for thin deep nets. *arXiv preprint arXiv:1412.6550*, 2014.

[187] O Ronneberger, P Fischer, and T Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-assisted Intervention*, pages 234–241. Springer, 2015.

[188] Olaf Ronneberger, Philipp Fischer, and Thomas 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*, pages 234–241. Springer, 2015.

[189] JM Rubin, M Mirfakhraee, EE Duda, GJ Dohrmann, and F Brown. Intraoperative ultrasound examination of the brain. *Radiology*, 137(3):831–832, 1980.

[190] Sebastian Ruder. An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv:1706.05098*, 2017.

[191] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015. doi: 10.1007/s11263-015-0816-y.

[192] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4510–4520, 2018.

[193] Rahul Sastry, Wenya Linda Bi, Steve Pieper, Sarah Frisken, Tina Kapur, William Wells III, and Alexandra J Golby. Applications of ultrasound in the resection of brain tumors. *Journal of Neuroimaging*, 27(1):5–15, 2017.

[194] Caroline A Schneider, Wayne S Rasband, and Kevin W Eliceiri. Nih image to imagej: 25 years of image analysis. *Nature methods*, 9(7):671, 2012.

[195] Thomas Schneider, Christian Mawrin, Cordula Scherlach, Martin Skalej, and Raimund Firsching. Gliomas in adults. *Deutsches Ärzteblatt International*, 107(45):799, 2010.

[196] T. Selbekk, A.S. Jakola, O. Solheim, T. F. Johansen, F. Lindseth, I. Reinertsen, and G. Unsgard. Ultrasound imaging in neurosurgery: approaches to minimize surgically induced image artefacts for improved resection control. *Acta Neurochir*, 155, 2013. doi: https://doi.org/10.1007/s00701-013-1647-7.

[197] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626, 2017.

[198] Bryar Shareef, Alex Vakanski, Min Xian, and Phoebe E Freer. Estan: Enhanced small tumor-aware network for breast ultrasound image segmentation. *arXiv preprint arXiv:2009.12894*, 2020.

[199] Bryar Shareef, Min Xian, and Aleksandar Vakanski. Stan: Small tumor-aware network for breast ultrasound image segmentation. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pages 1–5. IEEE, 2020.

[200] Mostafa Sharifzadeh, Habib Benali, and Hassan Rivaz. Phase aberration correction: A convolutional neural network approach. *IEEE Access*, 8:162252–162260, 2020.

[201] Mostafa Sharifzadeh, Sobhan Goudarzi, An Tang, Habib Benali, and Hassan Rivaz. Mitigating aberration-induced noise: A deep learning-based aberration-to-aberration approach. *IEEE Transactions on Medical Imaging*, 2024.

[202] Dinggang Shen, Guorong Wu, and Heung-Il Suk. Deep learning in medical image analysis. *Annual review of biomedical engineering*, 19:221–248, 2017.

[203] Xiaoyan Shen, Liangyu Wang, Yu Zhao, Ruibo Liu, Wei Qian, and He Ma. Dilated transformer: residual axial attention for breast ultrasound image segmentation. *Quantitative Imaging in Medicine and Surgery*, 12(9):4512, 2022.

[204] Jun Shi, Shichong Zhou, Xiao Liu, Qi Zhang, Minhua Lu, and Tianfu Wang. Stacked deep polynomial network based representation learning for tumor classification with small ultrasound image dataset. *Neurocomputing*, 194:87–94, 2016.

[205] Seung Yeon Shin, Soochahn Lee, Il Dong Yun, Sun Mi Kim, and Kyoung Mu Lee. Joint weakly and semi-supervised deep learning for localization and classification of masses in breast ultrasound images. *IEEE transactions on medical imaging*, 38(3):762–774, 2018.

[206] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[207] Vivek Kumar Singh, Mohamed Abdel-Nasser, Farhan Akram, Hatem A Rashwan, Md Mostafa Kamal Sarker, Nidhi Pandey, Santiago Romani, and Domenec Puig. Breast tumor segmentation in ultrasound images using contextual-information-aware deep adversarial learning framework. *Expert Systems with Applications*, 162:113870, 2020.

[208] Leslie N Smith. Cyclical learning rates for training neural networks. In *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 464–472. IEEE, 2017.

[209] Carlos Sobral, José Silvestre Silva, Alexandra André, and Jaime B Santos. Sarcopenia diagnosis: Deep transfer learning versus traditional machine learning.

[210] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*, pages 2256–2265. PMLR, 2015.

[211] Praotasna Sombune, Phongphan Phienphanich, Sutanya Phuechpanpaisal, Sombat Muengtaweepongsa, Anuchit Ruamthanthong, and Charturong Tantibundhit. Automated embolic signal detection using deep convolutional neural network. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 3365–3368. IEEE, 2017.

[212] Hannah Spitzer, Kai Kiwitz, Katrin Amunts, Stefan Harmeling, and Timo Dickscheid. Improving cytoarchitectonic segmentation of human brain areas with self-supervised siamese networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 663–671. Springer, 2018.

[213] Andrej Šteňo, Ján Buvala, Veronika Babková, Adrián Kiss, David Toma, and Alexander Lysak. Current limitations of intraoperative ultrasound in brain tumor surgery. *Frontiers in Oncology*, 11:851, 2021.

[214] Matt S Stock and Brennan J Thompson. Echo intensity as an indicator of skeletal muscle quality: applications, methodology, and future directions. *European journal of applied physiology*, 121:369–380, 2021.

[215] Howard J Stringer and Daisy Wilson. The role of ultrasound as a diagnostic tool for sarcopenia. *The Journal of frailty & aging*, 7(4):258–261, 2018.

[216] Run Su, Deyun Zhang, Jinhuai Liu, and Chuandong Cheng. Msu-net: Multi-scale u-net for 2d medical image segmentation. *Frontiers in Genetics*, 12:140, 2021.

[217] Deqing Sun, Xiaodong Yang, Ming-Yu Liu, and Jan Kautz. Models matter, so does training: An empirical study of cnns for optical flow estimation. *IEEE transactions on pattern analysis and machine intelligence*, 42(6):1408–1423, 2019.

[218] Jiawei Sun, Bobo Wu, Tong Zhao, Liugang Gao, Kai Xie, Tao Lin, Jianfeng Sui, Xiaoqin Li, Xiaojin Wu, and Xinye Ni. Classification for thyroid nodule using vit with

contrastive learning in ultrasound images. *Computers in Biology and Medicine*, 152: 106444, 2023.

[219] Nima Tajbakhsh, Laura Jeyaseelan, Qian Li, Jeffrey N Chiang, Zhihao Wu, and Xiaowei Ding. Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. *Medical Image Analysis*, 63:101693, 2020.

[220] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning*, pages 6105–6114. PMLR, 2019.

[221] Ali KZ Tehrani, Mina Amiri, Ivan M Rosado-Mendez, Timothy J Hall, and Hassan Rivaz. Ultrasound scatterer density classification using convolutional neural networks and patch statistics. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 2021.

[222] Klaus D Toennies. *Guide to medical image analysis*. Springer, 2017.

[223] Helena R Torres, Pedro Morais, Bruno Oliveira, Cahit Birdir, Mario Rüdiger, Jaime C Fonseca, and João L Vilaça. A review of image processing methods for fetal head and brain analysis in ultrasound images. *Computer methods and programs in biomedicine*, 215:106629, 2022.

[224] Frederick Tung and Greg Mori. Similarity-preserving knowledge distillation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1365–1374, 2019.

[225] Nishant Uniyal, Hani Eskandari, Purang Abolmaesumi, Samira Sojoudi, Paula Gordon, Linda Warren, Robert N Rohling, Septimiu E Salcudean, and Mehdi Moradi. Ultrasound rf time series for classification of breast lesions. *IEEE transactions on medical imaging*, 34(2):652–661, 2014.

[226] Jesper E Van Engelen and Holger H Hoos. A survey on semi-supervised learning. *Machine learning*, 109(2):373–440, 2020.

[227] Vincent Vanhoucke, Andrew Senior, Mark Z Mao, et al. Improving the speed of neural networks on cpus. In *Proc. deep learning and unsupervised feature learning NIPS workshop*, volume 1, page 4, 2011.

[228] Sagar Vaze, Weidi Xie, and Ana IL Namburete. Low-memory cnns enabling real-time ultrasound segmentation towards mobile deployment. *IEEE Journal of Biomedical and Health Informatics*, 24(4):1059–1069, 2020.

[229] Guotai Wang, Wenqi Li, Sébastien Ourselin, and Tom Vercauteren. Automatic brain tumor segmentation using convolutional neural networks with test-time augmentation. In *International MICCAI Brainlesion Workshop*, pages 61–72. Springer, 2018.

[230] Heng Wang, Donghao Zhang, Yang Song, Siqi Liu, Yue Wang, Dagan Feng, Hanchuan Peng, and Weidong Cai. Segmenting neuronal structure in 3d optical microscope images via knowledge distillation with teacher-student network. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 228–231. IEEE, 2019.

[231] Lin Wang and Kuk-Jin Yoon. Knowledge distillation and student-teacher learning for visual intelligence: A review and new outlooks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

[232] Liyuan Wang, Xingxing Zhang, Hang Su, and Jun Zhu. A comprehensive survey of continual learning: theory, method and application. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

[233] Yi Wang, Zijun Deng, Xiaowei Hu, Lei Zhu, Xin Yang, Xuemiao Xu, Pheng-Ann Heng, and Dong Ni. Deep attentional features for prostate segmentation in ultrasound. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 523–530. Springer, 2018.

[234] Juerd Wijntjes and Nens van Alfen. Muscle ultrasound: Present state and future opportunities. *Muscle & nerve*, 63(4):455–466, 2021.

[235] Jiaxiang Wu, Cong Leng, Yuhang Wang, Qinghao Hu, and Jian Cheng. Quantized convolutional neural networks for mobile devices. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4820–4828, 2016.

[236] Kaizhi Wu, Xi Chen, and Mingyue Ding. Deep learning based classification of focal liver lesions with contrast-enhanced ultrasound. *Optik*, 125(15):4057–4063, 2014.

[237] C Xia, J Li, X Chen, A Zheng, and Y Zhang. What is and what is not a salient object? learning salient object detector by ensembling linear exemplar regressors. In

Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 4142–4150, 2017.

[238] Menghua Xia, Hongbo Yang, Yanan Qu, Yi Guo, Guohui Zhou, Feng Zhang, and Yuanyuan Wang. Multilevel structure-preserved gan for domain adaptation in intravascular ultrasound analysis. *Medical Image Analysis*, 82:102614, 2022.

[239] Min Xian, Yingtao Zhang, Heng-Da Cheng, Fei Xu, Boyu Zhang, and Jianrui Ding. Automatic breast ultrasound image segmentation: A survey. *Pattern Recognition*, 79: 340–355, 2018.

[240] Yiming Xiao, Maryse Fortin, Geirmund Unsgård, Hassan Rivaz, and Ingerid Reinertsen. Re trospective evaluation of cerebral tumors (resect): A clinical database of pre-operative mri and intra-operative ultrasound in low-grade glioma surgeries. *Medical physics*, 44(7):3875–3882, 2017.

[241] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. Aggregated residual transformations for deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1492–1500, 2017.

[242] Bing Xu, Naiyan Wang, Tianqi Chen, and Mu Li. Empirical evaluation of rectified activations in convolutional network. *arXiv preprint arXiv:1505.00853*, 2015.

[243] Chunbo Xu, Yunliang Qi, Yiming Wang, Meng Lou, Jiande Pi, and Yide Ma. Arf-net: An adaptive receptive field network for breast mass segmentation in whole mammograms and ultrasound images. *Biomedical Signal Processing and Control*, 71:103178, 2022.

[244] Kunran Xu, Lai Rui, Yishi Li, and Lin Gu. Feature normalized knowledge distillation for image classification. In *European Conference on Computer Vision*, pages 664–680. Springer, 2020.

[245] Meng Xu, Kuan Huang, Qiuxiao Chen, and Xiaojun Qi. Mssa-net: Multi-scale self-attention network for breast ultrasound image segmentation. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 827–831. IEEE, 2021.

[246] Yuan Xu, Yuxin Wang, Jie Yuan, Qian Cheng, Xueding Wang, and Paul L Carson. Medical breast ultrasound image segmentation by machine learning. *Ultrasonics*, 91: 1–9, 2019.

[247] Jian Yang, Haoyang Cai, Zhi-Xiong Xiao, Hangyu Wang, and Ping Yang. Effect of radiotherapy on the survival of cervical cancer patients: an analysis based on seer database. *Medicine*, 98(30), 2019.

[248] Kaiwen Yang, Aiga Suzuki, Jiaxing Ye, Hirokazu Nosato, Ayumi Izumori, and Hidenori Sakanashi. Ctg-net: Cross-task guided network for breast ultrasound diagnosis. *PloS one*, 17(8):e0271106, 2022.

[249] Min-Chun Yang, Woo Kyung Moon, Yu-Chiang Frank Wang, Min Sun Bae, Chiun-Sheng Huang, Jeon-Hor Chen, and Ruey-Feng Chang. Robust texture analysis using multi-resolution gray-scale invariant features for breast sonographic tumor diagnosis. *IEEE Transactions on Medical Imaging*, 32(12):2262–2273, 2013.

[250] Moi Hoon Yap, Gerard Pons, Joan Martí, Sergi Ganau, Melcior Sentís, Reyer Zwiggelaar, Adrian K Davison, and Robert Marti. Automated breast ultrasound lesions detection using convolutional neural networks. *IEEE journal of biomedical and health informatics*, 22(4):1218–1226, 2017.

[251] Moi Hoon Yap, Manu Goyal, Fatima Osman, Ezak Ahmad, Robert Martí, Erika Denton, Arne Juette, and Reyer Zwiggelaar. End-to-end breast ultrasound lesions recognition with a deep learning approach. In *Medical imaging 2018: biomedical applications in molecular, structural, and functional imaging*, volume 10578, page 1057819. International Society for Optics and Photonics, 2018.

[252] Ujwal Yeole, Vikas Singh, Ajit Mishra, Salman Shaikh, Prakash Shetty, and Aliasgar Moiyadi. Navigated intraoperative ultrasonography for brain tumors: a pictorial essay on the technique, its utility, and its benefits in neuro-oncology. *Ultrasonography*, 39 (4):394, 2020.

[253] Michael Yeung, Evis Sala, Carola-Bibiane Schönlieb, and Leonardo Rundo. Unified focal loss: Generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation. *Computerized Medical Imaging and Graphics*, 95:102026, 2022.

[254] Yongjian Yu and Scott T Acton. Edge detection in ultrasound imagery using the instantaneous coefficient of variation. *IEEE Transactions on Image processing*, 13(12): 1640–1655, 2004.

[255] Paul A. Yushkevich, Joseph Piven, Heather Cody Hazlett, Rachel Gimpel Smith, Sean Ho, James C. Gee, and Guido Gerig. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *Neuroimage*, 31(3):1116–1128, 2006.

[256] Sergey Zagoruyko and Nikos Komodakis. Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer. *arXiv preprint arXiv:1612.03928*, 2016.

[257] Jiansong Zhang, Yongjian Chen, and Peizhong Liu. Automatic recognition of standard liver sections based on vision-transformer. In *2022 IEEE 16th International Conference on Anti-counterfeiting, Security, and Identification (ASID)*, pages 1–4. IEEE, 2022.

[258] Jiashu Zhang, Xiaolei Chen, Yan Zhao, Fei Wang, Fangye Li, and Bainan Xu. Impact of intraoperative magnetic resonance imaging and functional neuronavigation on surgical outcome in patients with gliomas involving language areas. *Neurosurgical review*, 38 (2):319–330, 2015.

[259] Yu Zhang, Yuanyuan Wang, Weiqi Wang, and Bin Liu. Doppler ultrasound signal denoising based on wavelet frames. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 48(3):709–716, 2001.

[260] Feng Zhao and Xianghua Xie. An overview of interactive medical image segmentation. *Annals of the BMVA*, 2013(7):1–22, 2013.

[261] Hang Zhou and Hassan Rivaz. Registration of pre-and postresection ultrasound volumes with noncorresponding regions in neurosurgery. *IEEE journal of biomedical and health informatics*, 20(5):1240–1249, 2016.

[262] Kevin Zhou. *Medical image recognition, segmentation and parsing: machine learning and multiple object approaches*. Academic Press, 2015.

[263] Lichen Zhou, Chuang Zhang, and Ming Wu. D-linknet: Linknet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction. In *CVPR Workshops*, pages 182–186, 2018.

[264] Qikui Zhu, Bo Du, Baris Turkbey, Peter L Choyke, and Pingkun Yan. Deeply-supervised cnn for prostate segmentation. In *2017 International Joint Conference on Neural Networks (Ijcnn)*, pages 178–184. IEEE, 2017.

[265] Zhemin Zhuang, Nan Li, Alex Noel Joseph Raj, Vijayalakshmi GV Mahesh, and Shunmin Qiu. An rdau-net model for lesion segmentation in breast ultrasound images. *PloS one*, 14(8):e0221535, 2019.

[266] Laurent Zieleskiewicz, Laurent Muller, Karim Lakhal, Zoe Meresse, Charlotte Arbelot, Pierre-Marie Bertrand, Belaid Bouhemad, Bernard Cholley, Didier Demory, Serge Duperret, et al. Point-of-care ultrasound in intensive care units: assessment of 1073 procedures in a multicentric, prospective, observational study. *Intensive care medicine*, 41(9):1638–1647, 2015.

[267] Jesse Zuckerman, Matthew Ades, Louis Mullie, Amanda Trnkus, Jean-Francois Morin, Yves Langlois, Felix Ma, Mark Levental, José A Morais, and Jonathan Afilalo. Psoas muscle area and length of stay in older adults undergoing cardiac operations. *The Annals of thoracic surgery*, 103(5):1498–1504, 2017.