# Receiver Channel Recovery in High-Frame-Rate Ultrasound Imaging using Branched Convolutional Neural Networks

by

William Michael Kovacs Pitman

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Master of Applied Science

in

Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2022

## Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

# Abstract

High frame rate ultrasound (HiFRUS) is an imaging paradigm that utilizes unfocused transmissions to perform acquisitions at kilohertz frame rates, and its high temporal resolution enables its use in tracking dynamic physiological events. Integration of HiFRUS techniques into compact ultrasound scanners could enable use of the paradigm in more remote and austere healthcare settings. While desirable, the high data rates and dedicated receiver electronics required for HiFRUS acquisitions make their implementation on compact systems difficult. Reduction of the number of receiving channels in a HiFRUS system can alleviate constraints related to data rate and receiver electronics, however, this reduction either limits the field of view of an ultrasound system or it leaves the system prone to spatial aliasing artifacts. To enable systems to operate with a reduced set of receiving channels, radiofrequency (RF) channel recovery frameworks have been proposed. While initial feasibility has been demonstrated for recovery of half of a system's receiving channels, higher degrees of recovery have minimal demonstrated viability. Higher degrees of recovery would allow for additional reduction in a system's receiving channels, enabling HiFRUS principles to be applied in systems with more appreciable reductions in cost and form factor.

The goal of this thesis is to devise a receiver channel recovery framework that is generalizable to multiple levels of channel-wise downsampling, and to evaluate its ability to recover RF channels after downsampling degrees of 2-times and higher. To facilitate channel recovery at multiple rates, novel branching encoder-decoder convolutional neural networks (CNNs) were developed. These CNNs were trained to recover omitted RF channels from angled plane wave acquisitions after channel-wise downsampling rates of 2-times, 3-times, and 4-times. To evaluate the utility of the trained CNNs, recovered RF data was used for ultrasound image formation using delay-and-sum beamforming, and for coherent compounding of beamformed images. When the trained recovery frameworks were applied to downsampled acquisitions of an *in vitro* point target, *in vivo* carotids, and an *in vivo* quadriceps muscle, inclusion of CNN-inferred RF data removed spatial aliasing artifacts from the beamformed images, recovering their underlying structure.

The proposed framework may be used to enable HiFRUS techniques on more compact and inexpensive systems. This can help extend the reach of HiFRUS, bringing technology that utilizes this paradigm into the hands of more users. Additionally, insights presented in this work can also be used

to guide further innovation, such as extension of channel recovery to 3D ultrasound or alternative transmissions schemes.

# Acknowledgements

Studying my master's degree has been an exciting, difficult, and fulfilling challenge, and I would like to express my gratitude to those who have helped me along the way.

I would first like to thank my supervisor, Prof. Alfred Yu, for providing me with the opportunity to pursue a MASc under his supervision. I am very grateful for his mentorship and his continued effort to help me develop my abilities as a researcher and communicator.

I would like to thank Adrian Chee, Billy Yiu, Di Xiao, and Hassan Nahas for all their guidance, encouragement, and patience throughout my studies. I am very appreciative of the time and effort you all spent providing me with valuable feedback on my research. I would also like to thank all my fellow colleagues at LITMUS for the insightful and cheerful conversations we shared over the past two years. It was a pleasure working alongside and learning from all of you.

Lastly, I would like to thank my parents for their unwavering love and support. Our conversations never fail to provide me with the encouragement, perspective, and emotional support I need.

# Table of Contents

## List of Figures

## List of Tables

## List of Abbreviations

| Abbreviations | |
|---|---|
| 2D | 2-Dimensional |
| 3D | 3-Dimensional |
| 2X | 2-Times |
| 3X | 3-Times |
| 4X | 4-Times |
| ADC | Analog-to-Digital Converter |
| CPU | Central Processing Unit |
| CNN | Convolutional Neural Network |
| CR | Contrast Ratio |
| DAS | Delay-and-Sum |
| DAQ | Digital acquisition |
| FWHM | Full Width at Half Maximum |
| GPU | Graphical Processing Unit |
| HiFRUS | High Frame Rate Ultrasound |
| FOV | Field of View |
| FPGA | Field Programmable Gate Arrays |
| L | Carotid Lumen (during contrast evaluation) |
| LNA | Low Noise Amplifier |
| PRF | Pulse Repetition Frequency |
| SSIM | Structural Similarity |
| T | Thyroid (during contrast evaluation) |
| W | Carotid Wall (during contrast evaluation) |

# List of Symbols

| Symbols | |
|---|---|
| $\alpha$ | Significance level for statistical hypothesis testing |
| $\alpha_{adjusted}$ | Adjusted significance level after Bonferroni correction |
| $\sigma$ | Activation operation in a neural network (chapter 2) |
| $\sigma_y$ | Variance of an image patch beamformed with all receiving channels (during SSIM calculation) |
| $\sigma_{\hat{y}}$ | Variance of an image patch beamformed with either a subset of receiving channels or a subset + CNN-inferred channels (during SSIM calculation) |
| $\sigma_{\hat{y}y}$ | Covariance between an image patch beamformed with all receiving channels and one beamformed with either a subset of receiving channels or a subset + CNN-inferred channels (during SSIM calculation) |
| $\mu_{tissue}$ | Mean of a tissue reference region in an ultrasound image |
| $\mu_{lumen}$ | Mean of a lumen reference region in an ultrasound image |
| $\mu_y$ | Mean of an image patch beamformed with all receiving channels (during SSIM calculation) |
| $\mu_{\hat{y}}$ | Mean of an image patch beamformed with either a subset of receiving channels or a subset + CNN-inferred channels (during SSIM calculation) |
| a | Activated output of a neuron |
| $A_d$ | Output feature map depth after a convolutional operation |
| $A_h$ | Output feature map height after a convolutional operation |
| $A_w$ | Output feature map width after a convolutional operation |
| b | Bias of a neuron |
| $c_0$ | Speed of sound |
| C | Number of channels in an RF image |
| $d_n$ | Distance from a scatterer to element i on a transducer |
| D | Degree of channel-wise downsampling |
| $F_d$ | Depth of a convolutional filter |
| $F_h$ | Height of a convolutional filter |
| $F_w$ | Width of a convolutional filter |

| | |
|---|---|
| $i$ | Index of a transducer element |
| $k$ | Number of inputs into a neuron |
| $K_1$ | Stability constant for SSIM calculations |
| $K_2$ | Stability constant for SSIM calculations |
| $\mathcal{L}$ | Machine learning model's loss |
| $M$ | Number of layers in a neural network |
| $n$ | Number of examples being examined in a machine learning loss function (chapter 2)<br>Number of RF examples being examined in an RF quality evaluation (chapter 3) |
| $N$ | Number of samples on a channel (column) of an RF image |
| $N_{scan}$ | Number of scanlines in an ultrasound image |
| $N_{trans}$ | Number of transmission angles in a compounded image |
| $P$ | Location in an imaging medium during the beamforming process |
| $p_{max}$ | Maximum amplitude of a point target |
| $R$ | Dynamic range of an ultrasound image |
| $t_n$ | Amount of time for a reflection from a scatterer to reach element n on a transducer |
| $w_1$ | Left side 6dB from maximum value in a FWHM calculation |
| $w_2$ | Right side 6dB from maximum value in a FWHM calculation |
| $X_d$ | Depth of an input feature map in a CNN |
| $X_h$ | Height of an input feature map in a CNN |
| $X_w$ | Width of an input feature map in a CNN |
| $y_i$ | A correct output for a machine learning model (Chapter 2)<br>A correct RF reference value from the recovery framework (Chapter 3) |
| $\hat{y_i}$ | A predicted output for a machine learning model (Chapter 2)<br>A predicted RF reference value from the channel recovery framework (Chapter 3) |
| $y_{max}$ | The maximum value for a reference region of RF data |
| $y_{min}$ | The minimum value for a reference region of RF data |

# Chapter 1
# Introduction

## 1.1 Channel Recovery in Compact High Frame Rate Ultrasound Systems

Ultrasound is currently undergoing innovation drives towards both 1) high frame rate ultrasound (HiFRUS) and 2) more compact and inexpensive scanners (Lanza, 2020). The HiFRUS paradigm allows acquisitions to be taken with sub-millisecond time resolution, enabling tracking of dynamic events in the human body. Meanwhile, the reduction in size and cost of ultrasound scanners allows them to be more accessible in remote and austere healthcare settings (Nelson & Sanghvi, 2016; Sippel, *et al*., 2011).

Implementation of HiFRUS techniques into compact scanners could extend the reach of the HiFRUS paradigm throughout the worldwide healthcare system; however, this integration is complicated by the large radiofrequency (RF) data volumes received during HiFRUS acquisitions. HiFRUS data rates can exceed 10GB/s, requiring high bandwidth connections for data transfer from system front-end to back-end. Furthermore, the electronics required to sample RF data on a full set of receiving channels during a HiFRUS acquisition increases the form factor and complexity of these systems. Both of these complications can be alleviated by reducing the number of receiving channels in a HiFRUS system, however, this channel reduction will either limit the field of view (FOV) of a system or introduce spatial aliasing artifacts due to an increased pitch between receiving channels/elements.

Direct receiver channel recovery can improve HiFRUS image quality when reduced receiver channel counts are used, while enabling a system to utilize lower data transfer rates and less receiving electronics. This recovery of received RF channels has been explored for HiFRUS acquisitions using compressed sensing recovery techniques (Ramkumar & Thittai, 2020; Anand & Thittai, 2021) and deep learning models (Xiao *et al.*, 2022; Kumar *et al.*, 2020). The initial feasibility demonstrated by these works showcase direct RF recovery as a promising tool to enable compact HiFRUS systems with low receiver channel counts.

## 1.2 Outline of Thesis Study

### 1.2.1 Motivation and Hypothesis

Given the advantages associated with receiver channel reduction (described further in section 2.3.2.1), it is of practical interest to reduce a HiFRUS system's receiving channel count by large degrees,

provided that the related performance tradeoffs (described further in section 2.3.2.2) can be mitigated. While higher degrees of receiver channel reduction translate to more appreciable improvements in system portability, bandwidth, and cost, there is limited demonstrated feasibility for receiver channel recovery after large degrees of channel-wise downsampling. Compressed sensing techniques show *in vitro* image degradation when inferring RF channels beyond a factor of 2-times (2X; Ramkumar & Thittai, 2020), and no *in vivo* feasibility has been demonstrated for higher degrees of channel-wise downsampling. Additionally, deep learning techniques have no displayed feasibility for HiFRUS RF channel recovery degrees beyond 2X, and currently explored deep learning frameworks (Xiao *et al.*, 2022; Kumar *et al.*, 2020) would require architecture changes to facilitate higher degrees of RF recovery. As such, there is a need for additional innovation in RF channel recovery to enable its implementation in systems with appreciable reduction in channel count. To drive this innovation, we plan to develop a novel inference framework that can infer missing RF data after various degrees of channel-wise downsampling. We hypothesize that similarities in the time-delayed reflections that each channel receives can be used to infer missing channel data, even when the subset of available RF data is smaller than ½ of the available channels. Accordingly, we posit that branching encoder-decoder CNNs can be used to perform this inference using deep learning principles; an encoder can learn compressed features from a received subset of RF channels, and branched decoders can use the compressed features to efficiently recover missing RF channels after multiple degrees of downsampling.

### 1.2.2 Research Objectives

The overall goal of this research work is to develop a framework that can recover missing RF channel data from uniformly downsampled subsets of channels. Specifically, this work aims to accomplish the following research objectives:

1) Develop a deep-learning-based framework that can recover RF channel data from plane wave acquisitions that have been downsampled by degrees of 2X, 3X, and 4X.
2) Evaluate the practical effectiveness of the framework by using recovered *in vitro* and *in vivo* RF data for ultrasound imaging, specifically delay-and-sum (DAS) beamforming and coherent plane wave compounding.

DAS beamforming and coherent plane wave compounding are two fundamental techniques used to produce HiFRUS images (to be described in sections 2.2.4 and 2.2.5). Therefore, an evaluation of the

effectiveness of inferred RF data for these operations can provide insight on the utility of the recovery framework in a compact HiFRUS system.

### 1.2.3 Thesis Organization

The remainder of this thesis is organized as follows:

- Chapter 2 provides relevant background on HiFRUS receiver channel reduction and machine learning fundamentals.
- Chapter 3 describes how the proposed RF channel recovery framework was developed, trained, and evaluated.
- Chapter 4 outlines the *in vitro* and *in vivo* experimental results from the recovery framework's evaluation.
- Chapter 5 interprets the results presented in chapter 4, and discusses advantages, limitations, and future directions for the RF channel recovery framework.

# Chapter 2
# Background: High Frame Rate Ultrasound Receiver Channel Reduction and Machine Learning Fundamentals

## 2.1 Chapter Overview

The purpose of this chapter is to provide background describing 1) how receiver channel recovery can be used to enable compact HiFRUS systems and 2) the machine learning principles that were used to develop the proposed recovery framework. First, the principles of HiFRUS operation and its advantages and benefits are described. Second, the components of an ultrasound system and the difficulties in downsizing HiFRUS systems are outlined; this is accompanied by an explanation of how receiver channel reduction can alleviate these difficulties, along with its associated challenges. Third, a literature review is provided on current techniques that can enable channel-wise downsampled ultrasound systems, along with a description of their limitations. The chapter is concluded with an explanation of the machine learning and convolutional neural network (CNN) fundamentals that the proposed recovery framework is built upon.

## 2.2 High Frame Rate Ultrasound Overview

### 2.2.1 High Frame Rate Ultrasound Advantages and Applications

HiFRUS operation is characterized by an unfocused transmission scheme, which allows acquisitions to be performed at frame rates as high as 10,000Hz. These kilohertz frame rates allow the monitoring of dynamic events that occur in the human body on a millisecond or sub-millisecond time scale; for example, the high temporal resolution provided by HiFRUS enables tracking of physiological processes such as arterial pulse waves (Couade *et. al*, 2011), heart contraction dynamics (Cikes *et al.*, 2014), complex blood flow dynamics (Bercoff *et. al*, 2011; Yiu & Yu, 2016), and shear waves that propagate through body tissues (Montaldo *et al.*, 2009). Tracking these physiological processes provides valuable information that can be used for medical diagnosis, prevention, and monitoring (Tanter & Fink, 2014). The following subsections describe the physical principles that enable HiFRUS to perform such acquisitions with high temporal resolution. First, a general introduction to ultrasound physics and image formation is provided, and then the HiFRUS paradigm is introduced.

### 2.2.2 Ultrasound Pulse Echo Sensing Principle

Generally, ultrasound systems operate according to a pulse-echo sensing principle. First, a transducer, typically made up of piezoelectric elements, is excited to insonify an imaging medium with pulses of ultrasonic waves. As the transmit wave propagates through the imaging medium, echoes are produced by reflection and backscattering events (described in section 2.2.3). These echoes are then returned to the transducer to be received on each of its individual elements. Received echoes are then converted to RF voltage signals with an analog-to-digital converter (ADC) and sampled on the ultrasound device's receiving channels. Once the data has been received and sampled, time of flight (ToF) principles can be used to determine the location of the medium's echoes, forming an ultrasound image through a process known as beamforming (described in section 2.2.4).

### 2.2.3 Echo Production: Reflection and Scattering

When an ultrasound wave that has been transmitted from a transducer encounters structures in the imaging medium, echoes produced by reflection and backscattering events are returned to the ultrasound probe. Reflection occurs when an ultrasound wave that is travelling through a region with a specific acoustic impedance encounters another region with a different acoustic impedance; this results in the reflection of a portion of the incident ultrasound wave (Humphrey, 2007). The degree of the signal that is reflected depends on the mismatch in impedances, with larger differences resulting in more reflection. Instances of reflection occur when the structure encountered is relatively large compared to the ultrasound wavelength. Conversely, when the structure is not large relative to the transmit wavelength, a portion of the incident ultrasound wave can scatter in multiple directions, with some signal returning towards the ultrasound probe (Powles *et al.*, 2018).

### 2.2.4 Ultrasound Receive Beamforming

Ultrasound image formation is typically performed through a process called delay-and-sum (DAS) beamforming, as described in Figure 2.1. Given an excited scatterer positioned at point P in an imaging medium (shown in Figure 2.1 (a)), different channels in an ultrasound array should receive an echo signal from the scatterer, albeit at different times. To determine the delay that is experienced before the echo signals reach each element, ToF principles can be used. That is, using the distance $d_i$ between point P and an element in question, the ToF for an echo to reach the i'th element is

$$t_i = {}^{d_i}/_{c_0} \tag{2.1}$$

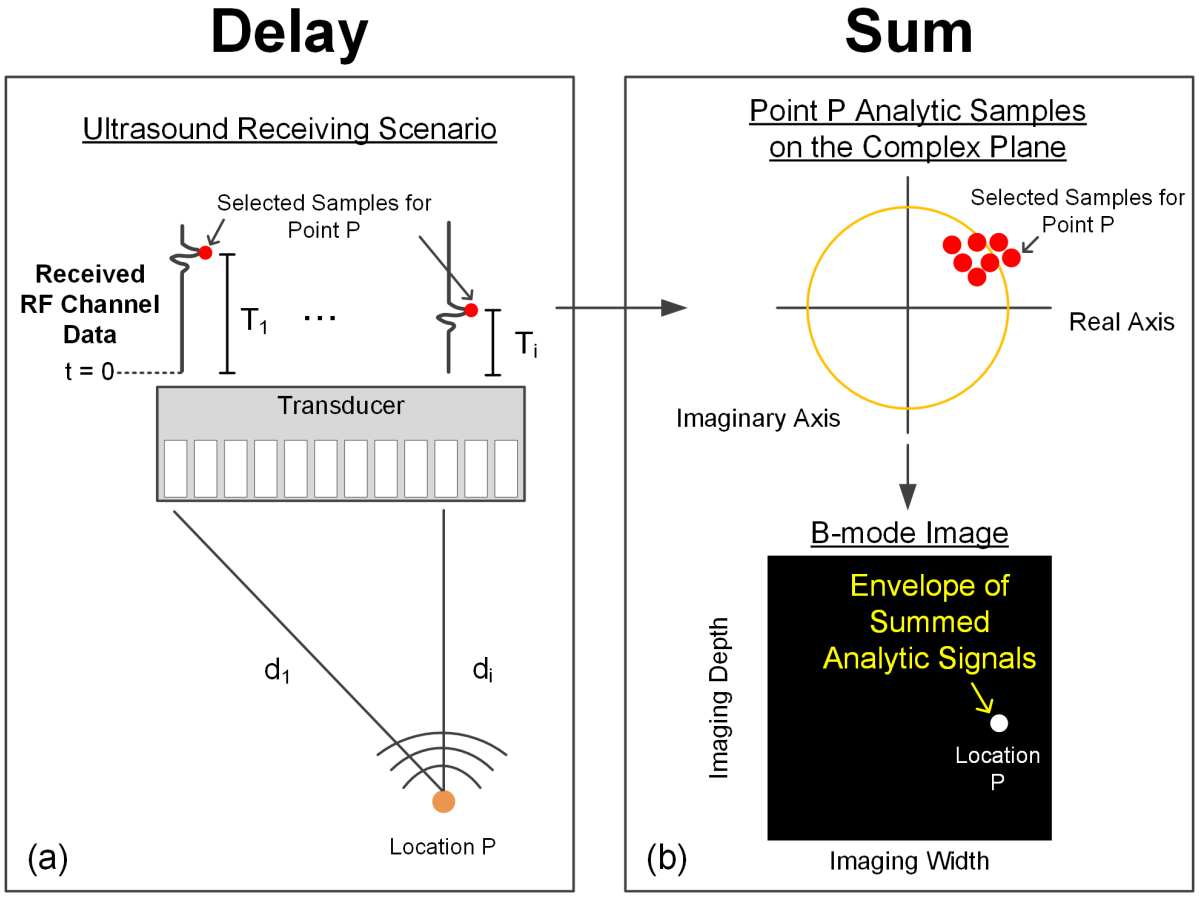


Figure 2.1. Delay and Sum beamforming process. (a) Delay: Shows the process of receiving ultrasound echoes and selecting samples for a given point based on ToF principles. (b) Sum: summation of selected samples to determine an imaging point's amplitude.

where $c_0$ is the speed of sound in the medium and it is usually assumed to be 1540m/s while imaging tissue. This delay can be added to the time that it should take before a transmit wave would arrive at point P (which can be calculated with similar principles based on one's transmission scheme); summing these two values yields the total delay $T_i$ needed to find samples that correspond to point P's reflections.

Once all the delays for each channel have been calculated, the corresponding samples can then be summed, as shown in Figure 2.1 (b). Prior to summation, the raw RF signals are converted to analytic form through a process such as the Hilbert transform (Perrot *et al.*, 2021). The corresponding analytic

samples for a given position can then be summed in the complex domain. The amplitude of the resulting complex signal is then taken to give the ultrasound image value for location P. Lastly, these image values are typically logarithmically scaled to compress the image's overall dynamic range. This process can be repeated for each pixel location in an imaging medium to form an ultrasound brightness-mode (B-mode) image.

### 2.2.5 High Frame Rate Ultrasound Acquisition Scheme

Traditional ultrasound and HiFRUS acquisition schemes both operate using pulse-echo sensing and beamforming, but their main difference is in the transmission and receiving schemes performed during an acquisition. These differences are highlighted in Figure 2.2, where (a) shows the traditional focused scan-line imaging scheme, and (b) shows an unfocused HiFRUS imaging scheme. In the traditional imaging scheme, focused beams are used to insonify specific regions of the imaging medium, and echoes from this region are DAS beamformed into a line of image data. This forms one scanline of an image, and this process is repeated $N_{scan}$ times to form $N_{scan}$ scanlines that make up the final ultrasound

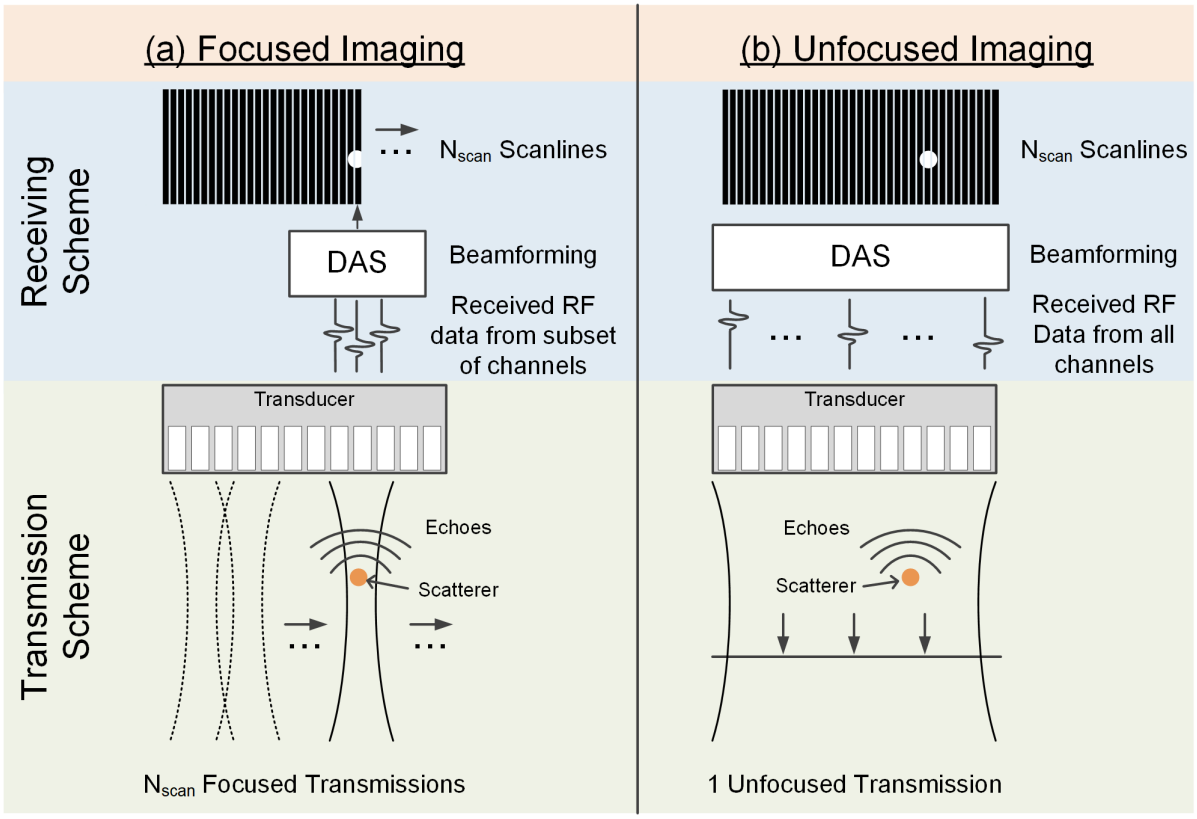


Figure 2.2. Ultrasound (a) focused and (b) unfocused imaging scenarios.

image. Conversely, for HiFRUS acquisitions, the entire imaging medium is insonified with a single unfocused transmission. All the scanlines in the image are then beamformed together using reflections from the individual transmission. By enabling ultrasound image formation from a single transmission, the frame rate of an ultrasound system can be significantly increased; typical frame rates for scanline-based imaging range are ~25Hz whereas HiFRUS can achieve rates as high as 10,000Hz. To achieve unfocused transmissions, plane wave (Montaldo *et al.*, 2009) or diverging wave (Jensen *et al.*, 2006) transmission schemes can be used, where Figure 2.2 (b) shows a plane wave transmission propagating through the medium.

#### 2.2.5.1 HiFRUS Image Compounding

While the unfocused nature of HiFRUS enables a much higher frame rate to be achieved, the lack of transmission focus results in contrast and resolution degradation in beamformed images. To mitigate

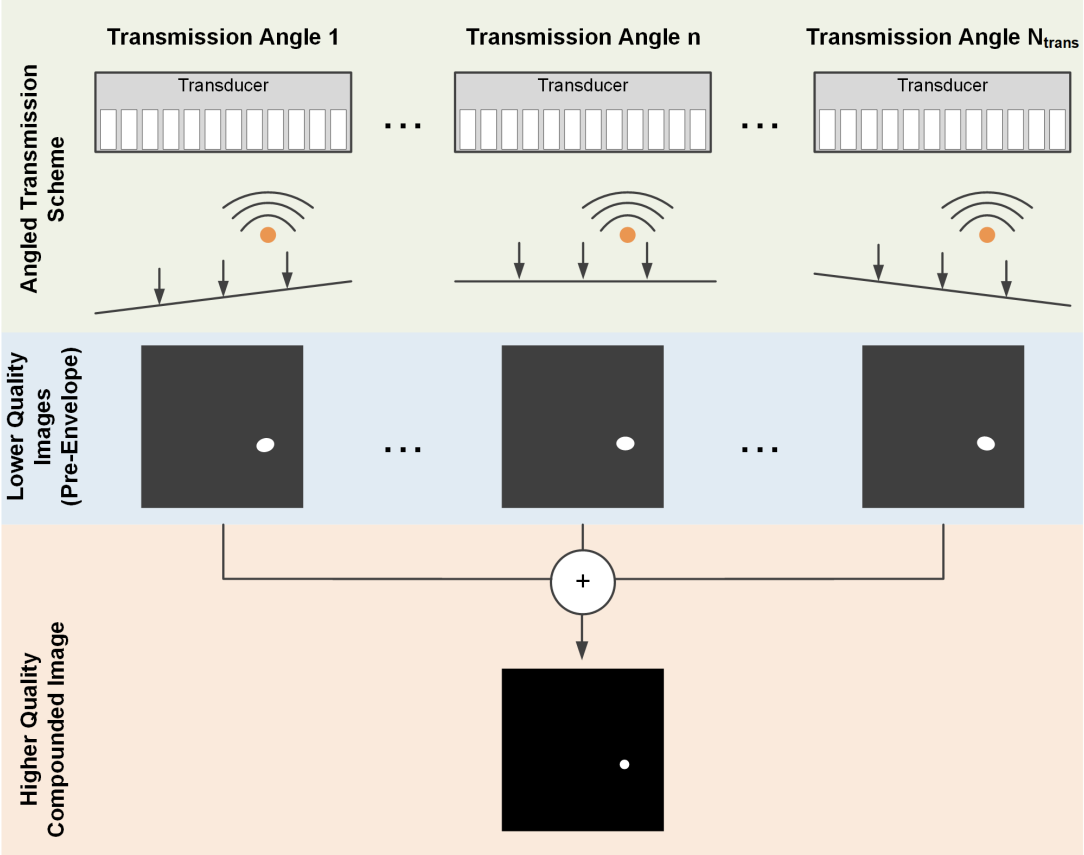


Figure 2.3. Plane wave image compounding process. Beamformed images from multiple steered plane waves are added together to form a higher quality compounded image.

this loss in image quality, coherent compounding of multiple HiFRUS images can be used (Montaldo *et al.*, 2009). As shown in Figure 2.3, image compounding is achieved by adding together the analytic beamformed (pre-envelope) images from multiple unfocused transmissions. For plane wave imaging, the set of images can be from multiple transmit steering angles. The compounding of multiple lower-quality images results in a higher quality final image, at the cost of a lower frame rate. Despite the reduction in frame rate, compounding HiFRUS images can still result in high quality images formed at frame rates above 1000Hz, significantly surpassing frame rates achieved with traditional scanline-based imaging.

## 2.3 Integrating High Frame Rate Ultrasound into Compact Ultrasound Systems

Advances in transducer manufacturing, transmit/receive circuitry, and signal processing algorithms have paved the way for the development of more inexpensive and compact ultrasound scanners (Baran & Webster, 2009). This has led to increased uptake of ultrasound not only within hospital settings, but also in prehospital, austere, and remote environments (Nelson & Sanghvi, 2016; Sippel *et al.*, 2011). A successful integration of HiFRUS techniques into compact ultrasound scanners should result in a broader reach of new HiFRUS applications within the worldwide healthcare system. This section outlines the specific difficulties involved in integrating HiFRUS into more compact scanners and discusses how receiver channel reduction is a possible solution. This is followed by a description of the challenges associated with receiver channel reduction.

### 2.3.1 Ultrasound System Hardware Overview and HiFRUS System Constraints

An ultrasound system needs to be capable of the following operations: 1) generating transmission events to insonify a medium, 2) receiving and sampling RF echoes from the medium, 3) subsequently processing received/sampled RF echoes, and 4) displaying system outputs to the operator in real time. In this thesis, the focus will be on ultrasound scanners that perform the RF processing in a software-based system back-end. Providing raw received RF channel data to a back-end processing unit provides a system with the flexibility to implement numerous different ultrasound techniques that require raw channel data as inputs (Boni *et al.*, 2018; Van Sloun *et al.*, 2020). The components required in these software-based ultrasound scanners are displayed in Figure 2.4 and described in the following subsection.

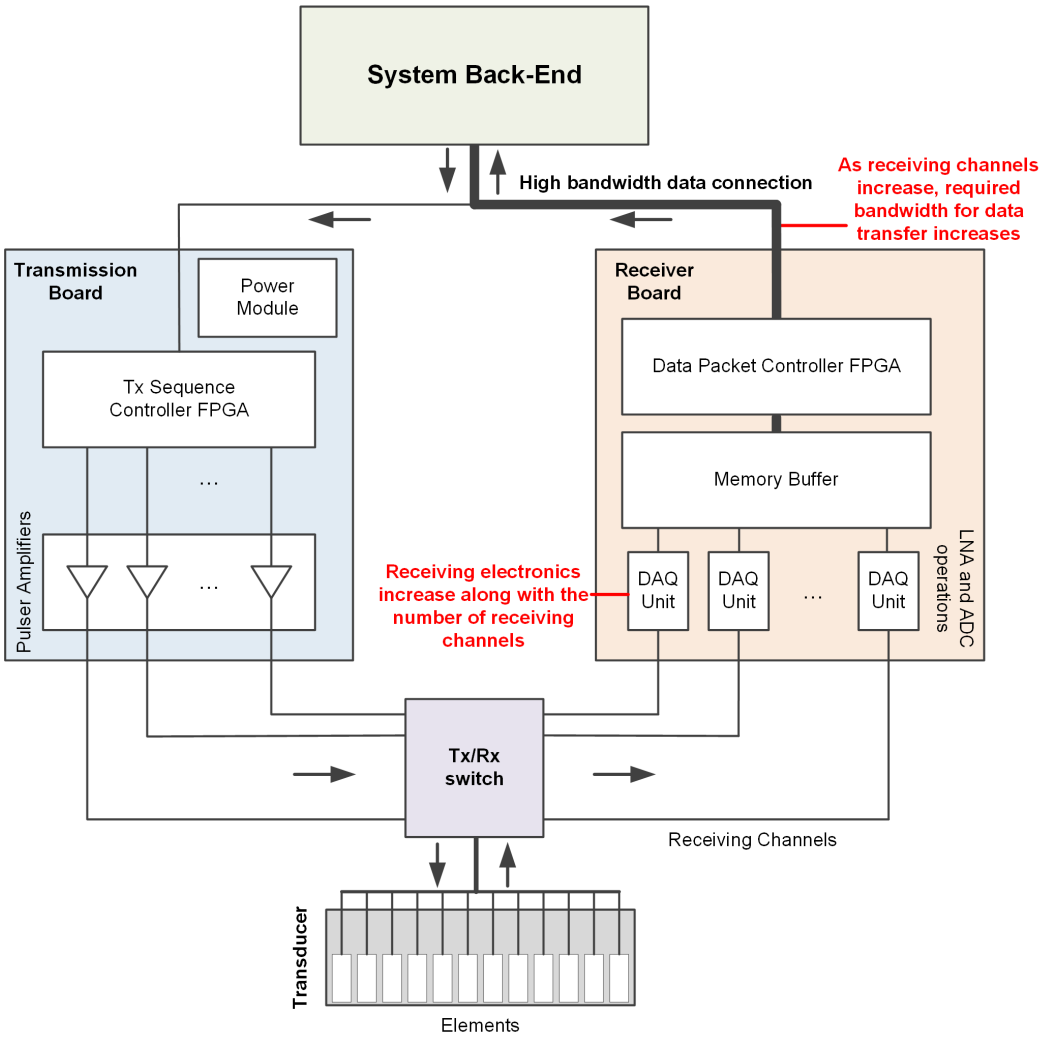


Figure 2.4. Components of a software-based ultrasound scanner. Transmission and receiving events each have their own dedicated electronics, and they both interact with the system back-end. Adapted from Fig. 2 in (Boni *et al.*, 2018) under CC 4.0.

#### 2.3.1.1 Software-Based Ultrasound Scanner System Overview

The back end of an ultrasound system interacts with the transducer for transmission and receiving events, each having its own dedicated electronics (Boni, Yu, Freear, Jensen & Tortoli, 2018). For transmissions, a field programmable gate array (FPGA) controls the transmission sequence, and the FPGA's output signals are amplified and sent to the transducer to insonify the medium. Following a transmission, switches are first used to connect receiving electronics to the transducer. Next, echo signals from the imaging medium are amplified and sampled by digital acquisition (DAQ) units, where

low noise amplifiers (LNAs) and ADCs are often employed to execute this stage. After being processed by a memory buffer and a data packet controller FPGA, the received RF channel data is sent to the system back-end through a high-bandwidth data streaming connection. Once transferred to the system back-end, RF processing such as DAS beamforming may be performed by central processing units (CPUs) and graphical processing units (GPUs).

#### 2.3.1.2 System Constraints Imposed by HiFRUS

The acquisition scheme performed during HiFRUS places a burden on a system's electronics and makes it difficult to implement this paradigm in compact systems. The difficulties associated with HiFRUS are highlighted in red in the system diagram of Figure 2.4. First, receiving RF signals on each element/channel at very high pulse repetition frequencies (PRFs) results in a large amount of data being produced during imaging. HiFRUS data rates can reach ~75MB/s for a single channel (or almost 10GB/s in a 128-channel system), requiring very high bandwidth data links between a system's front-end and back-end (Boni *et al.*, 2018). Secondly, the requirement to receive RF data on each element/channel on each HiFRUS acquisition imposes the need for dedicated receiving electronics for the full set of elements/channels. These constraints that are imposed by the HiFRUS paradigm serve as an obstacle when trying to reduce size and cost in an ultrasound system.

### 2.3.2 Receiver Channel Reduction: A Means of Reducing System Complexity

#### 2.3.2.1 Advantages of Receiver Channel Reduction

To decrease system complexity and to alleviate data transfer bandwidth requirements, an attractive option is to decrease the number of receiving channels in a system. As channels are decreased, fewer receiving electronics are required, allowing a smaller system form factor. Additionally, the total volume of received RF data is decreased when less channels receive data, allowing for less taxing data transfer bandwidth requirements. Higher degrees of system simplification can be achieved as more receiver channels are removed, since fewer receiving electronics are required, and lower volumes of data are produced with each acquisition. As shown in Figure 2.5, a tangible reduction in form factor can be observed in the US4R-Lite system (us4us Ltd., Warsaw, Poland), which contains 4X less receiving (and transmitting) channels compared to its less compact counterpart, the US4R (us4us Ltd., Warsaw, Poland). This channel reduction results in a compact system with 32 receiving channels and 128 transmitting channels. The simplified US4R-Lite has an associated ~4X reduction in total volume

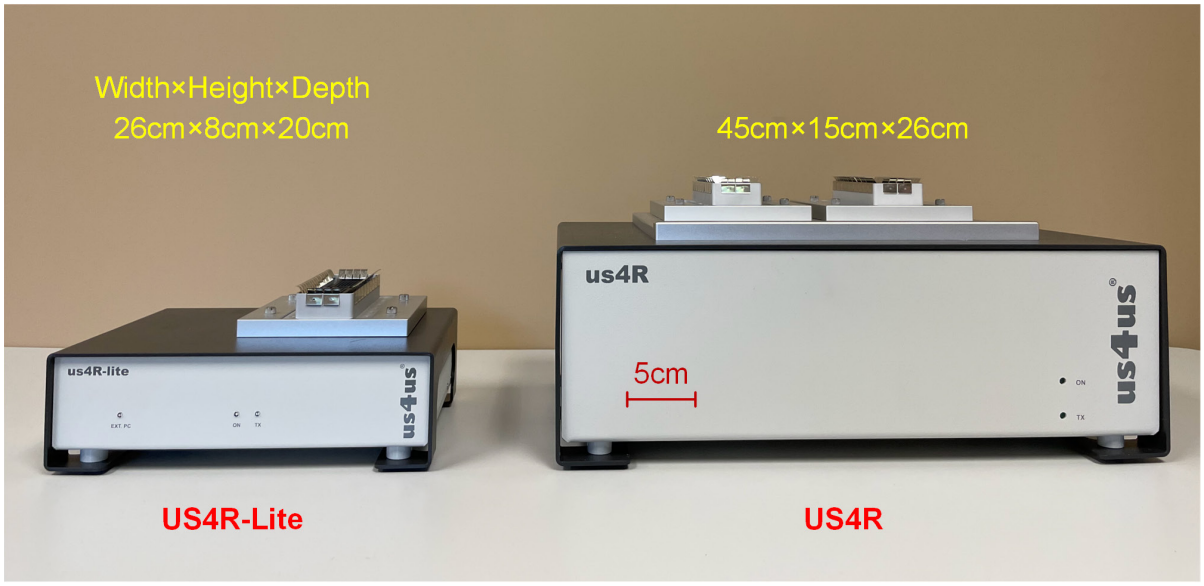

Figure 2.5. Size comparison between the US4R-Lite and US4R ultrasound systems.

(4160cm$^3$ vs. 17550cm$^3$) that is coupled with a 4X reduction in required data bandwidth when transferring received RF data to a system back-end.

#### 2.3.2.2 Difficulties Associated with Receiver Channel Reduction

While receiver channel reduction is an easily implementable method of reducing ultrasound system complexity, there are associated performance tradeoffs. Given a fixed pitch between receiving elements/channels, enough channels need to receive RF data to cover the desired imaging FOV. If the FOV of a system is to be preserved, this means that the pitch between receiving elements must be increased. Systems that have a large pitch relative to their transmit wavelength are susceptible to spatial aliasing artifacts that obscure beamformed images. An example of spatial aliasing is given in Figure 2.6 (a) and (b), where a simulated (Jensen, 1996; Jensen & Svendsen, 1992) point target image is obscured with artifacts when only the even numbered channels are used to receive data. The spatial aliasing phenomenon is one of the main obstacles that must be overcome when designing low channel count systems, and its origin will be explained in the following subsection.

#### 2.3.2.3 Spatial Aliasing

Spatial aliasing is a consequence of the pulse-echo nature of ultrasound imaging, since echoes from one location can contribute to the RF samples taken while beamforming another location. The potential for echo ambiguity is illustrated in Figure 2.6, where received RF channels that contain echoes from the

point target shown in (a) are placed into the columns of an image (referred to as an RF image) in (c). When arranged in the RF image form, echoes from the point target manifest themselves in a hyperbola along the received channels. To beamform pixel locations 1 and 2 in (b), the samples that are selected based on ToF principles are highlighted as yellow points in the RF image of (c). As emphasized in the red box in (c), the samples that are beamformed for locations 1 and 2 both contain echoes from the point target hyperbola. It is desired to sum these echo signals for location 1, as it is the true location of the point target. Conversely, location 2 does not contain any object, and it is not desired for these samples to coherently sum. Despite this, it is apparent in (b) that when only the even channels are used on receive there is some coherence in the summation, as an artifact appears in the beamformed image.

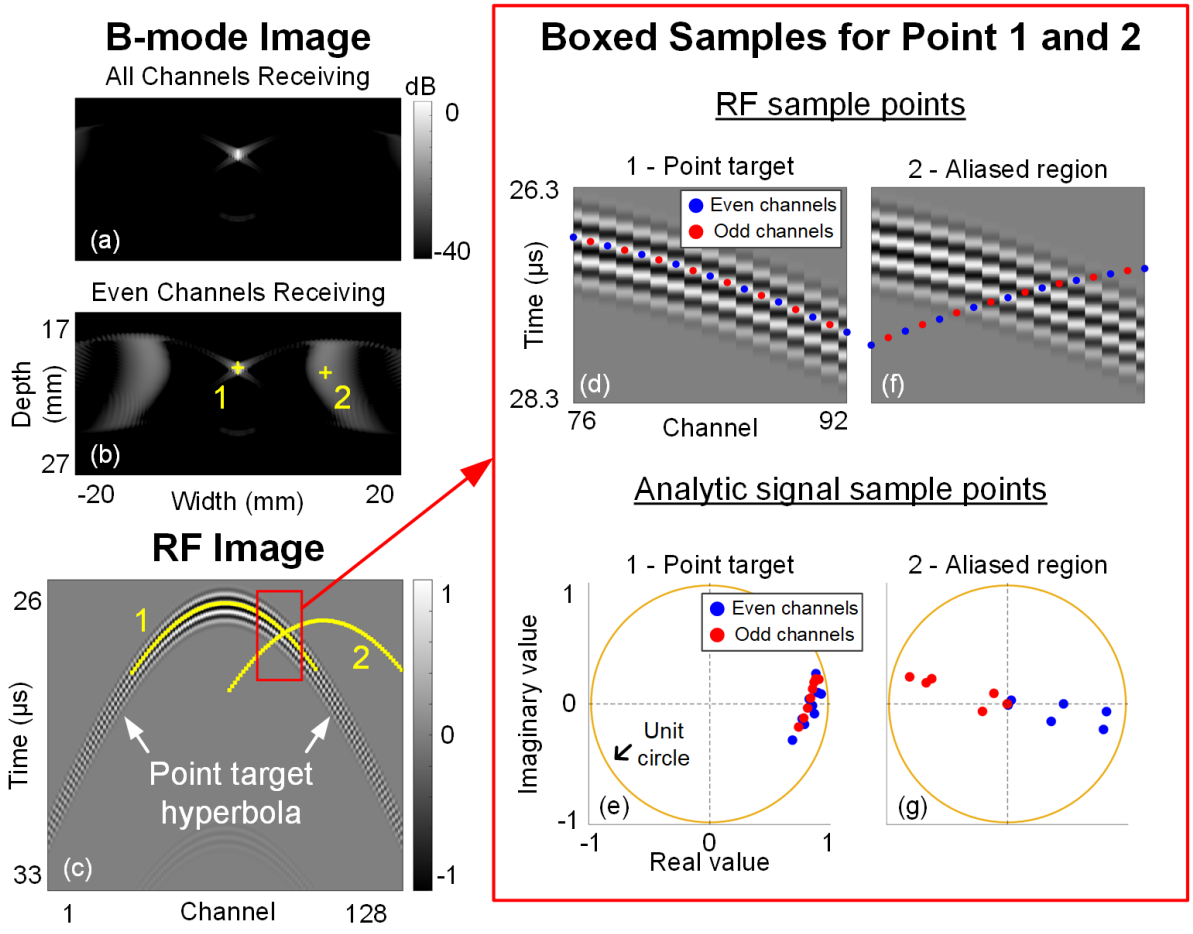


Figure 2.6. Spatial aliasing artifact explanation. (a) Beamformed point target with all channels receiving RF data. (b) Aliased point target image when only even channels are receiving. Point target and aliased points of interest highlighted in yellow. (c) Corresponding RF image for the beamformed point target. Samples beamformed for the points of interest in (b) are highlighted in yellow. Samples in the red box are examined more closely in (d)-(g). (d) Beamformed samples for point 1 in (b), (e) corresponding analytic samples. (f) beamformed samples for point 2 in (b), (g) corresponding analytic samples.

When RF echoes are received at an inadequate spatial sampling rate, they can incorrectly appear to be coming from multiple locations in the imaging medium. For the beamformed samples in Figure 2.6 (c), the red-boxed samples are further examined in (d)-(g): (d) and (f) show a close-up of the RF samples beamformed for locations 1 and 2, respectively; and (e) and (g) display the corresponding analytic signals plotted on the complex plane. For location 1, when samples are removed from the beamforming summation (due to removal of channels), the analytic sum will still be coherent, albeit at a lower resulting amplitude. Conversely, for location 2 the coherence of the analytic summation is dependent on the samples/channels used. Shown in (f) and (g), omission of either odd or even channels will result in a partially coherent summation, causing an aliasing artifact to appear in the beamformed image. Due to the insufficient spatial sampling rate taken on receive, it incorrectly appears that there is a reflector originating at location 2. This spatial aliasing phenomenon may occur at multiple points in an image when an insufficient channel pitch is used for beamforming, resulting in the aliased clouds in Figure 2.6 (b). Additionally, as more receiving channels are removed, the spatial aliasing phenomenon will worsen, as is shown in Figure 2.7. As the pitch of the receiving array expands further beyond the transmit aperture's fundamental wavelength λ, the spatial aliasing artifacts become more prominent in the beamformed image. It should be emphasized that a more echogenic reflector will result in a more prominent spatial aliasing artifact, since higher amplitude samples will coherently sum at the artifact

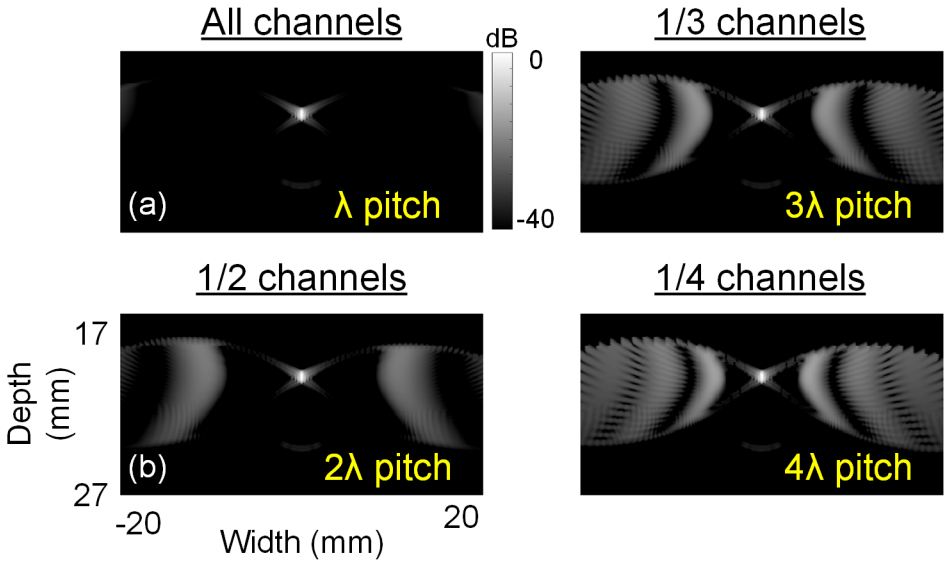


Figure 2.7. Worsening spatial aliasing artifacts as a linear array's receiving pitch is increased. (a) All channels receiving, λ pitch. (b) 1/2 channels receiving, 2λ pitch. (c) 1/3 channels receiving, 3λ pitch. (d) 1/4 channels receiving, 4λ pitch.

location. Consequently, the most obstructive aliasing artifacts in an imaging medium will tend to be caused by insufficient sampling of the medium's most echogenic reflections.

## 2.4 Current Research to Enable Low Channel Count HiFRUS Systems

To improve HiFRUS system performance while enabling a lower receiving channel count, several techniques are available. As a hardware-based approach, channel multiplexing can be used to gradually receive a full set of channel data over successive transmissions (Yu *et al.*, 2020; Carpenter *et al.,* 2016). This method effectively simplifies a system's electronics; however, it requires multiple transmissions to receive a full set of RF data. This reduces a system's effective frame rate and leaves a system vulnerable to motion artifacts caused by object movement between successive transmissions. Alternatively, microbeamforming can be performed at the probe's head and grouped data from multiple elements can be transferred together on one channel (Larson, 1993). This method can alleviate the data-transfer bandwidth imposed by HiFRUS, but it requires additional electronics in the ultrasound probe's head, and the partially beamformed RF data allows less flexibility for subsequent RF processing. Lastly, sparse arrays with optimized layouts can be used to minimize spatial aliasing artifacts in beamformed images (Lockwood *et al.*, 1996). This method can be used to reduce the spatial aliasing artifacts present in beamformed images, but it comes with the cost of increased sidelobe amplitude in beamformed images (Diarra *et al.*, 2013).

To enable low-receiver-count HiFRUS systems without requiring substantial hardware modifications, software-based approaches have been explored. Specialized beamformers have been proposed to directly recover ultrasound images from subsampled RF channels: compressed sensing techniques (Donoho, 2006) have been used to directly recover ultrasound images from subsets of channels (Besson *et al.*, 2016; David *et al.*, 2015); deep learning techniques can be used to learn optimal channel subsampling schemes and a corresponding direct recovery scheme for ultrasound B-mode or doppler images (Hujiben *et al.*, 2020); and a specialized convolution-based nonlinear beamformer has been developed to improve ultrasound image quality when a subset of channels are used for image formation (Cohen & Eldar, 2018). Additionally, post-beamforming deep learning methods can be utilized to suppress spatial aliasing artifacts in ultrasound images (Perdios *et al.*, 2020). While all these techniques can improve beamformed image quality while operating with a reduced channel count, they all bypass the recovery of a full set of RF channel data for a given transmission. This omission of RF channel recovery restricts a system's ability to implement signal processing techniques that require

access to a full set of RF data for a given frame. This includes methods for speed of sound mapping (Feigin *et al.*, 2020) and image segmentation (Nair *et al.*, 2020), as well as any specialized beamforming algorithms (Synnevåg *et al.*, 2009; Matrone *et al.*, 2015; Cheng & Lu, 2006; Garcia *et al.,* 2013). An alternative approach to enable low-receiver-count HiFRUS systems is to directly recover a full set of RF data from a subset of channels. This approach gives a higher degree of flexibility in a system for subsequent image formation and RF analysis.

## 2.5 Machine Learning and CNN Fundamentals

To achieve RF recovery from multiple levels of downsampling, a CNN-based model was developed (described in chapter 3). This subsection describes the machine learning and CNN fundamentals that the RF recovery framework is built upon.

### 2.5.1 Supervised Learning Overview

The basic idea behind supervised learning is that by feeding a machine learning model with examples that have known inputs and outputs, that model can be trained to learn the underlying function describing the relationship between the provided inputs and outputs. If the model can successfully learn the underlying function that describes that data, then new examples with unknown outputs can be fed into the machine learning model and it should correctly predict the example's outputs. This supervised learning process is outlined in Figure 2.8. Training is facilitated by calculating a loss that takes the machine learning model output and the correct training example output as inputs. The machine learning model parameters are modified to try and minimize the loss calculations as examples are passed into the model. Once suitable parameters have been found, the trained model can be used to predict the output from additional inputs that do not have known outputs.

When determining what type of model to use for a machine learning problem, the inputs that will be fed into the model need to be considered. With traditional machine learning models, inputs are often manually produced/handcrafted features. This is compared to deep learning approaches that, due to their higher complexity, can learn abstract features from the raw data itself (Dargan *et al.*, 2019). To enable inputs of raw received RF data into a recovery framework, a deep learning approach was taken in this research work.

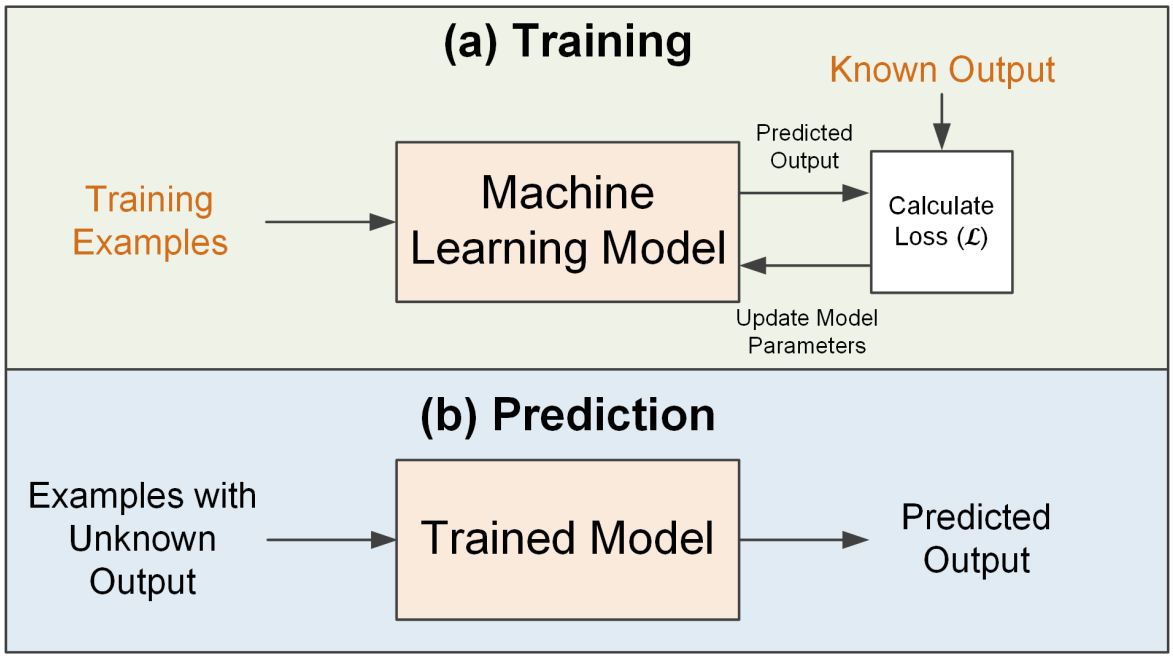


Figure 2.8. Supervised learning outline. (a) outlines training of a machine learning model. Examples with known outputs are given to a model, and parameters are updated based on a loss function. (b) trained models predict outputs from examples where the output is unknown.

### 2.5.2 CNN Overview

CNNs are a type of deep learning model that is inspired by the visual perception of animals (Gu *et al.*, 2018). In CNNs, collections of neurons perform operations on local groups of data from a larger gridded input, and each local operation's output is placed on a grid to be fed into another collection of neurons. This subsection details the different components and operations that are used to build up a CNN.

#### 2.5.2.1 Neurons

The most basic component of a CNN model is a single neuron. As shown in Figure 2.9, neurons perform multiplication, addition, and activation operations. When a set of inputs ($x_1$, $x_2$, … $x_k$) are fed into a single neuron, each input is multiplied by a unique weight ($w_1$, $w_2$, … $w_k$). Each of these products is then summed together along with an additional bias term b. The sum of these terms is then fed into an activation function σ, which gives the neural network's output a. The activation function is often used to introduce nonlinearity into the machine learning model, allowing it to approximate more complex functions. 1D examples of commonly used activation functions are given in Figure 2.10. The overall operation of a neuron that is shown in Figure 2.9 can be summarized by the following equation:

$$a = \sigma\left(\sum_{i=1}^{k}[w_i x_i] + b\right) \tag{2.2}$$

Adjusting the weights or bias of a neuron scales or shifts the inputs to its activation function. The goal when training a neural network is to find the optimal weights and biases that result in accurate predictions at the output of the network.

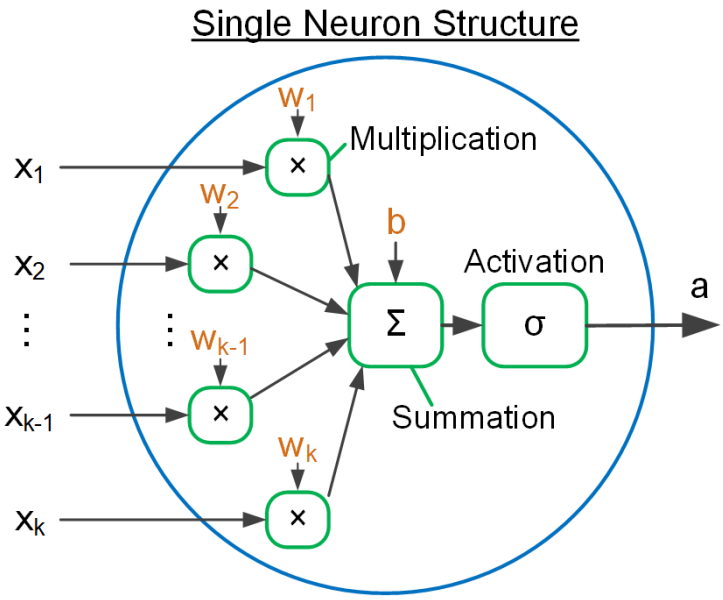


Figure 2.9. Overview of a single neuron. Inputs are multiplied by unique weights and then summed. The summation output is lastly passed into an activation function.

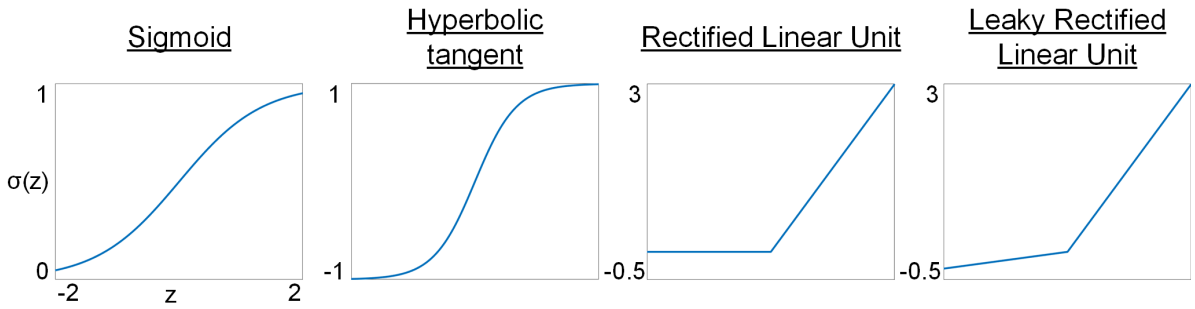


Figure 2.10. Common activation functions used at the output of a neuron.

#### 2.5.2.2 Convolutional Neurons

In a traditional fully connected neural network, all terms of a particular input are simultaneously fed into a neuron to produce a single output. This scenario is shown in Figure 2.11 (a), where each input of a 5×5 image is fed into a neuron to be multiplied by a unique weight. This arrangement is suitable when there are a low number of inputs, but when inputs are scaled up it becomes difficult to implement. For example, if a 512×512 image was to be fed into a single fully connected neuron, this would require the neuron to contain 512×512 = 262144 individual weights. Typical neural networks have multiple layers that contain several neurons, and this can quickly result in an unmanageable number of parameters that need to be optimized during training.

To decrease the number of parameters in a neural network when the size of an input is large, CNNs can be used. The basic building block of a CNN is still a neuron, but the way the neuron processes inputs becomes different. The method that CNNs use for processing inputs is shown in Figure 2.11 (b), where local groups of inputs are individually processed, and a matrix of outputs are given instead of a single output. This matrix of activated outputs holds the local features extracted from the neuron input, and it is often referred to as a feature map. It is important to note that the weights used for each operation in (b) do not change and they are kept in the same orientation; this allows the neuron to contain only 4 unique weights. Processing inputs in this way allows substantial reduction in the number of parameters required in a network. Additionally, processing local patches of an input allows a CNN to learn features that may be present in multiple locations of an input (Albawi *et al.*, 2017).

It can be observed that this alternative method of processing inputs is equivalent to orienting the weights of the neuron into a 2-dimensional (2D) filter, performing a 2D cross-correlation of the input with the filter, adding a bias, and passing each output through an activation function. This interpretation of the operation is given in Figure 2.11 (c), where the cross-correlation with the filter is used to express the same operation as is shown in (b). This overall cross-correlation operation can be described by the following equation:

$$a_{ij} = \sigma\left(\sum_{f=1}^{F_h}\sum_{g=1}^{F_w}\left[\left[w_{fg}x_{(i+f)(j+g)}\right]\right] + b\right) \tag{2.3}$$

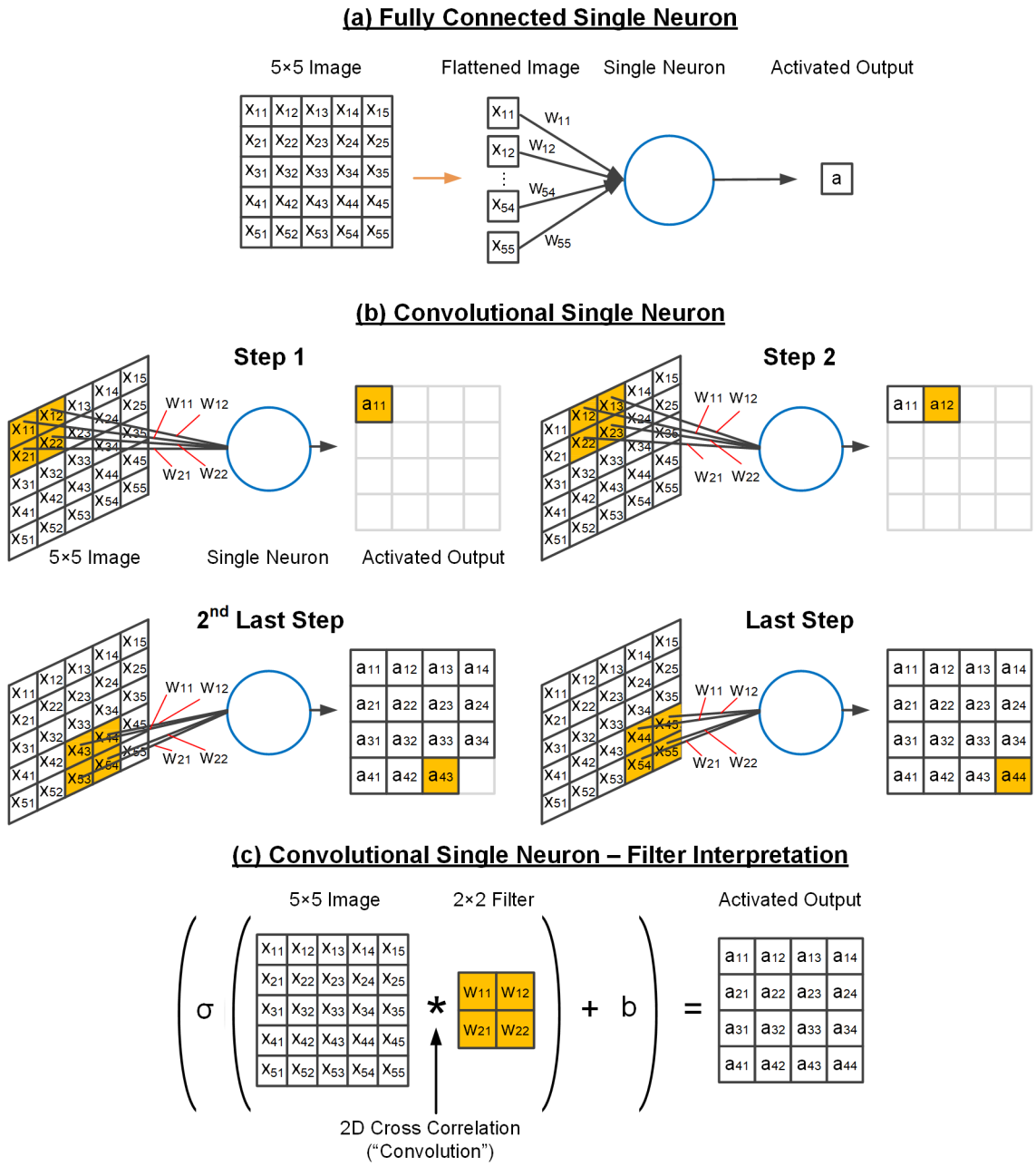


Figure 2.11. Comparison between a fully connected and a convolutional neuron. (a) fully connected neuron scenario. Each input of an image is given to a neuron, requiring a unique weight for multiplication with each input. (b) Convolutional neuron scenario. Patches of an input image are processed by the same set of weights, resulting in a feature map of activated outputs. (c) Convolutional filter interpretation of the scenario presented in (b). Neuron weights are arranged in a 2D filter, and 2D cross correlation is performed on the input. A bias is then added, and outputs are passed through an activation function.

where $F_h$ and $F_w$ are the height and width of the filter, respectively. The double sum operation in the center of equation 2.3 is equivalent to a Frobenius inner product (denoted by ":") between the input patch and the filter, resulting in the following expression:

$$a_{ij} = \sigma\left(\left[W : X_{ij}\right] + b\right) \tag{2.4}$$

where W corresponds to the 2D weight filter being used for the cross-correlation, and $X_{ij}$ corresponds to the 2D image patch that provides output $a_{ij}$. This operation is repeated until each patch of the inputs are processed. It should be noted that the cross-correlation operation can be replaced by a convolution if the weights in the filter are flipped prior to processing. This is not necessarily performed in practice, but conventionally the operation described in Figure 2.11 (b) and (c) is referred to as a convolution (Goodfellow *et al.*, 2016). Following this convention, this operation will be referred to as a convolution in this work as well.

### 2.5.2.3 Specialized Convolutional Operations

Modifications can be made to the basic convolutional operation to alter its provided outputs. The modifications that are utilized in this work are shown in Figure 2.12, where the first, second, and last step of each operation are given to highlight their differences. Only the convolutional operation is shown here for simplicity, and the addition of a bias and the application of an activation are implied.

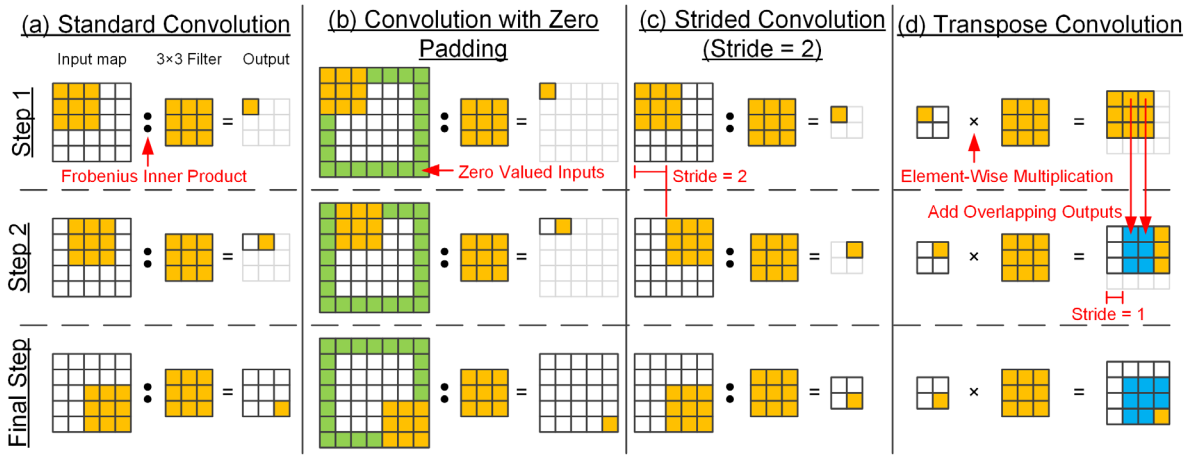


Figure 2.12. Specialized convolutional operations. (a) is the standard convolutional operation described in Figure 2.11 (b) and (c). (b) is a convolution with zero padding, yielding outputs with the same height/width as its input. (c) is a strided convolution, where steps greater than 1 are used when selecting input patches of an input matrix. (d) is a transpose convolution with stride = 1, used to upsample inputs.

A standard convolutional operation (shown in (a)) will result in outputs with reduced size compared to the input; there will be an $F_h$-1 reduction in height and a $F_w$-1 reduction in width. If this property is undesired, the inputs can be zero-padded prior to convolution (shown in (b)). By increasing the input dimensions, the output dimensions can be increased to be the original input's size.

Convolutional operations can additionally be modified to downsample or upsample the input feature map. Stride corresponds to the distance traversed between each patch of an input map that is processed, where the default stride value is 1. Convolutional operations with increased stride (shown in Figure 2.12 (c)) can be used to learn an effective downsampling scheme in a CNN (Springenberg *et al.*, 2015). Conversely, transpose convolutions (Dumoulin & Visin, 2018) can be used to learn an effective upsampling scheme in a CNN. Shown in Figure 2.12 (d), a single step of a transpose convolution consists of taking a single input value, performing an element-wise multiplication with a convolutional filter, and placing each output on a grid. When gridded outputs overlap, they are then added together. The stride of the transpose convolutional operation corresponds to the distance between placements of the output groups, where a larger stride will result in a larger output feature map. This operation can be used to increase the dimensions of an output feature map.

#### 2.5.2.4 Building up a CNN

CNNs are typically made up of several layers that consist of multiple neurons. The way that this is scaled is shown in Figure 2.13. (a) shows the single neuron case, where a single 2D filter is applied to a 2D input, providing a single 2D output. As additional neurons are added to a layer, this can be interpreted as the addition of new filters that each perform their own 2D convolutional operations on the input. For a 2D input (shown in (b)), the 2D output from each 2D filter is stacked together to form a 3D output with depth $A_d$, where $A_d$ is the number of filters/neurons in the layer.

When 3D inputs must be handled by a convolutional layer, the depth of filters will increase to match the depth of inputs. For a layer with a single neuron/filter (shown in (c)), the filter depth will increase to $X_d$, where $X_d$ is the depth of the input. Each 2D slice of the 3D input will be 2D-convolved with its corresponding 2D slice of the filter, and then each output is summed along the third dimension, resulting in a single 2D output for a single neuron/filter. Similarly, when a 3D input with depth $X_d$ is passed into a layer that contains $A_d$ total filters/neurons (shown in (d)), each filter will have a depth of $X_d$ to process the 3D input, and the output will have a depth of $A_d$.

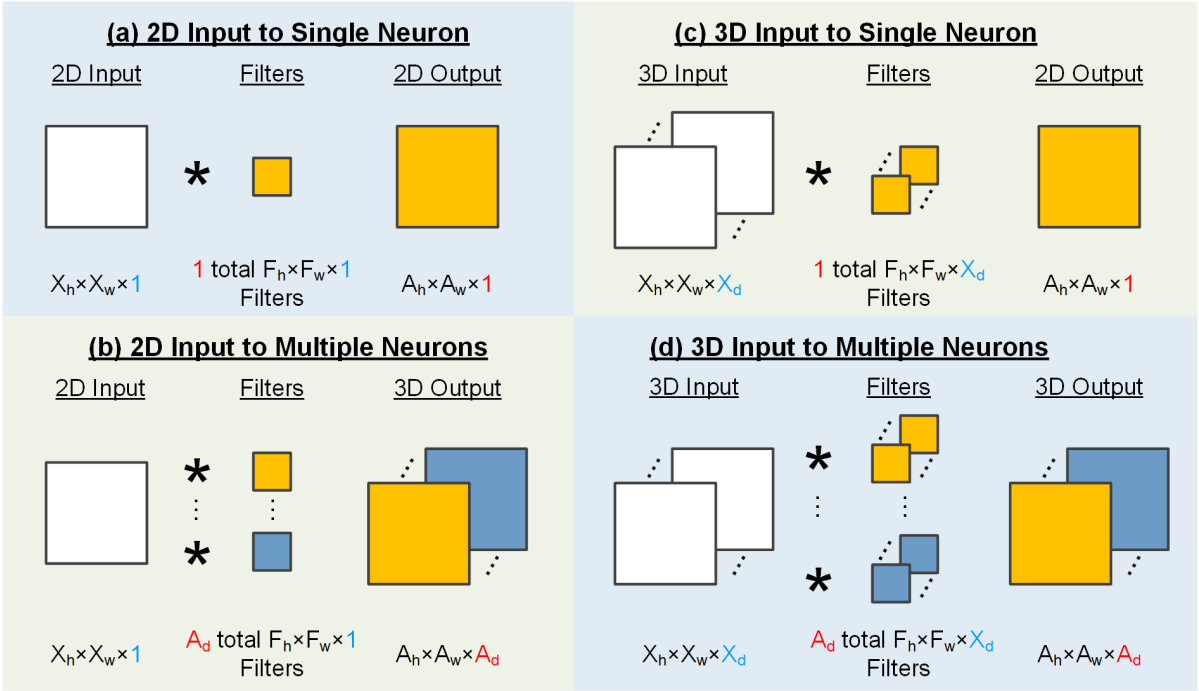


Figure 2.13. Scenarios for scaling up CNNs. (a) 2D input into a single neuron/filter, producing a 2D output. (b) 2D input to multiple neurons/filters, producing a 3D output. (c) 3D input to a single neuron/filter, producing a 2D output. (d) 3D input to multiple neurons/filters, producing a 3D output.

The general architecture of a CNN is comprised of several convolutional layers that each include multiple filters/neurons, where feature map outputs from each layer are passed forward to successive convolutional layers for subsequent processing. As the number of layers in a network is increased, the network can often yield better performance (Simonyan & Zisserman, 2015; Szegedy *et al.,* 2015). Features extracted by a CNN become increasingly more complex and nonlinear as the number of layers is increased, allowing complex underlying patterns in data to be learned during training. Additionally, the receptive field of a CNN (Araujo *et al.*, 2019) becomes increased through successive convolutional operations. A CNN's receptive field corresponds to the size of the input region that is used to produce an output feature. This concept is visualized in Figure 2.14, where two successive 3×3 convolutions are performed (indicated by arrows), and the input/output feature maps are shown. Since each feature in the 2$^{nd}$ layer is calculated using a 3×3 input in the first layer, the output from a 3×3 convolution on the second layer will have used a larger window of values in the first layer. As the number of layers in a CNN are increased and the resulting receptive field grows, outputs are provided with more input information during inference.

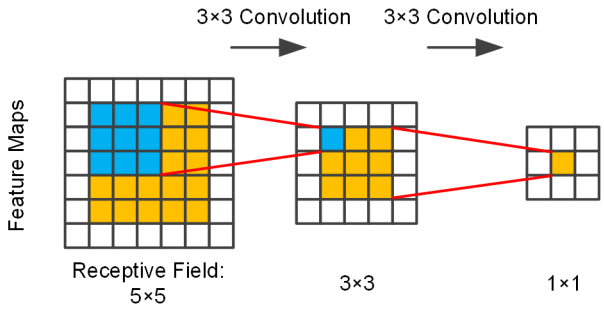


Figure 2.14. Receptive field of a CNN.

### 2.5.3 Training Neural Networks

Once a CNN architecture has been determined, the optimal parameters in the network need to be calculated. Each convolutional layer is made up of several neurons/filters that contain weights and biases. The process of finding the optimal weights and biases in a CNN is called training. The two fundamental techniques used to find the optimal parameters in a CNN are gradient descent and backpropagation.

#### 2.5.3.1 Gradient Descent: Training a CNN

To try and find the optimal parameters in a CNN, first a metric that can be used to grade a model's performance needs to be determined. This metric is known as loss $\mathcal{L}$, and the goal of training is to find parameters that minimize the model's loss. The loss function is determined with reference to a set of examples with known outputs. An example of a typical loss function is the mean squared error (MSE), given as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{2.5}$$

where $y_i$ is a known output for a given input $x_i$, $\hat{y}_i$ is the machine learning model's output when $x_i$ is input, and n corresponds to the number of examples in the full set being examined (for example in the set of training examples). If a model has a relatively high MSE, that means that its prediction error is relatively high, which is undesired. An example of two linear regression models (also interpreted as a single neuron with a linear activation function) with differing MSE losses is shown in Figure 2.15. (a)

shows a linear regression fit that does not model the data well, while (b) shows a fit that models the data well. The corresponding MSE values for each of these fits are given in (c) and (d), where the MSE value for the poor fit is comparatively higher.

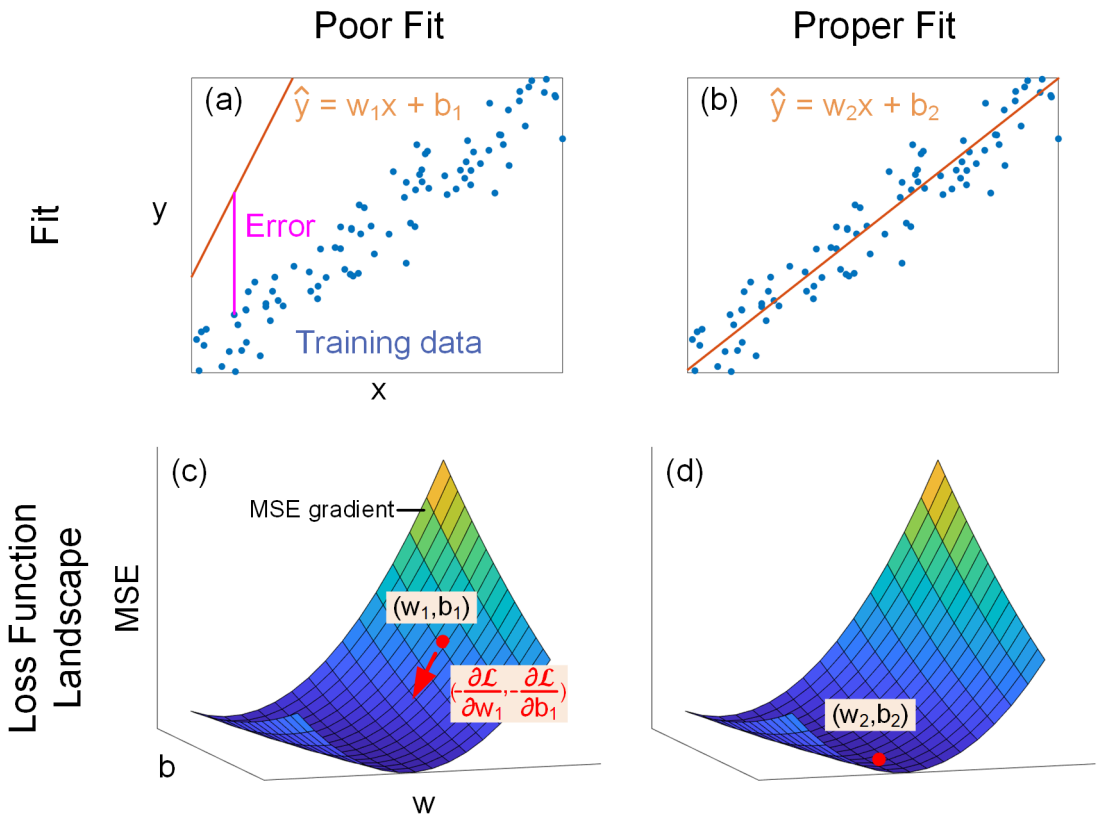


Figure 2.15. Linear regression fits and their corresponding position on an MSE loss landscape. (a) A poor linear regression fit. (b) A relatively better linear regression fit. (c) MSE value for the fit in (a). Surrounding MSE values for different values of w and b are plotted. Direction of the negative gradient at position ($w_1$, $b_1$) given by the red arrow. (d) MSE value for the fit in (b).

Once a proper loss function has been determined, the loss is minimized on a training set of data through a process called gradient descent. First, the gradient of the loss function $\nabla\mathcal{L}$, that is its derivative with respect to each parameter in the neural network, is calculated. Then, steps down the negative gradient are taken to minimize the model's loss and find a model with optimal fit. The gradient of the MSE loss for the linear regression fits in Figure 2.15 (a) and (b) are visualized in (c) and (d). To bring the model's fit from that in (a) to (b), the loss gradient terms $(\frac{\partial \mathcal{L}}{\partial w_1}, \frac{\partial \mathcal{L}}{\partial b_1})$ for the model need to be calculated at the model's current parameters ($w_1$ and $b_1$), and then the model needs to alter the

parameters by stepping down the gradient (shown by the red arrow in (c)). This process can be used to find a machine learning model's optimal fit to a given set of data. In more complex models with a larger number of trainable parameters, this process is the same, only the gradient will be made up of a larger number of terms (with a term for each parameter in the model).

To speed up the process of gradient descent and machine learning model training, the gradient of a loss function for a training set is often approximated over a batch of the training set. This allows faster calculation of the batch gradient, enabling a quicker overall training process. Furthermore, loss functions are generally defined in a summation over a set of training examples, allowing the batch gradient to be calculated by adding the loss's gradient for each individual example in the batch. This process is repeated until all the batches in a training set have been used in gradient calculations, and this constitutes completion of an epoch. Epochs are repeated until a model that produces a suitably low loss value is found.

#### 2.5.3.2 Backpropagation: Finding the CNN's Gradient

To calculate an individual training example's gradient in a neural network, the derivative of that example's loss function needs to be calculated with respect to each parameter (weight and bias) in the network. This calculation is predominantly complicated by the feedforward nature of a neural network, as network parameters from early layers may be embedded in many operations within the overall neural network's output expression. For example, the output of an M-layer single neuron chain (no convolutions) can be described through a recursive implementation of equation 2.2 (k = 1), where the layer # is given in bracketed superscripts:

$$\hat{y} = a^{(M)} = \sigma\left(w_1^{(M)}\left\{\sigma\left(w_1^{(M-1)}\{\cdots\} + b_1^{(M-1)}\right)\right\} + b_1^{(M)}\right). \tag{2.6}$$

Additional expressions of equation 2.2 would be placed in the "…", depending on the depth M of the network. When the output of equation 2.6 is given as an input to a loss function, derivation of the loss function expression with respect to a weight or bias in an early layer would require extensive use of the chain rule. To implement these calculations practically, backpropagation (Hecht-Nielsen, 1989) calculates gradient terms in a systematic fashion from the final layer towards the beginning of a network. This flow of calculation is shown for a single neuron chain in Figure 2.16. After an initial

forward pass of a particular input example (shown in (a)), the loss can be calculated with the network's output and the correct training label (shown in (b)). To find the loss gradient values for each parameter ($\frac{\partial \mathcal{L}}{\partial w}, \frac{\partial \mathcal{L}}{\partial b}$), the derivative of the loss with respect to the predicted output ($\frac{\partial \mathcal{L}}{\partial \hat{y}}$) is calculated, and this begins the flow of gradient term calculations back through the network (shown in (c)).

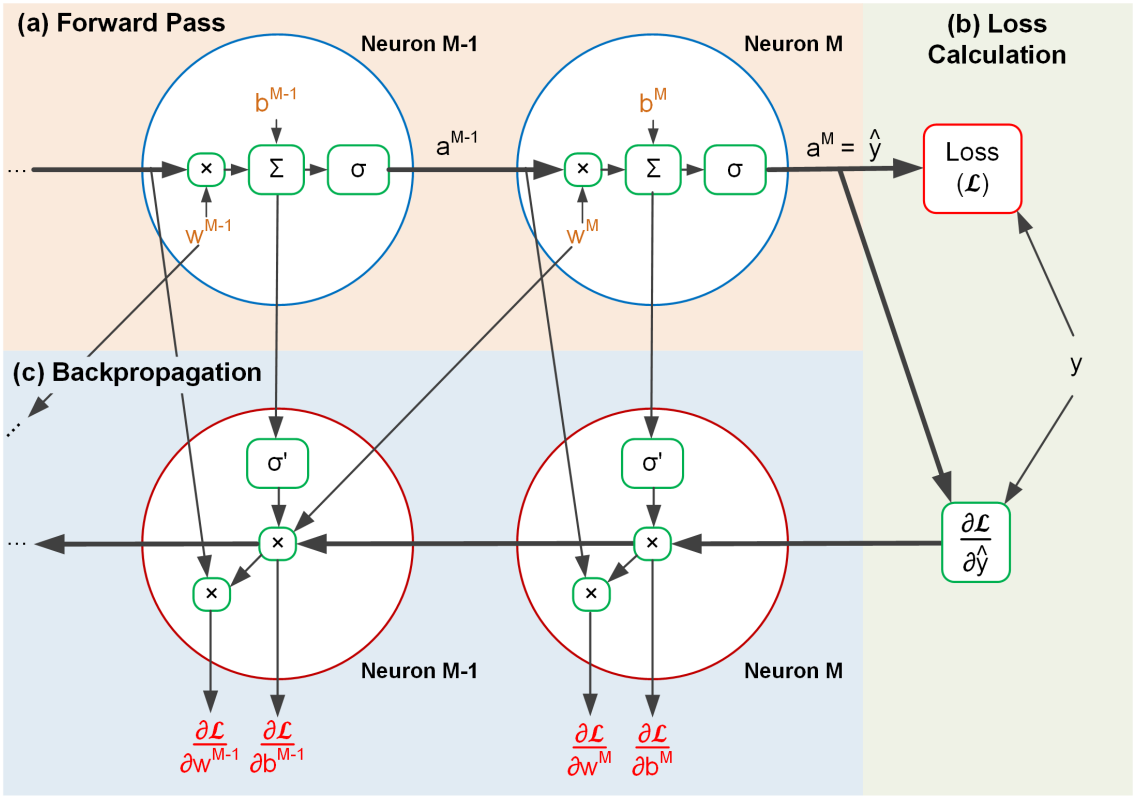


Figure 2.16. Backpropagation for a single neuron chain. (a) Neuron operations during a forward pass. (b) Loss function calculations. (c) Backpropagation operations for the corresponding neurons in (a).

When calculating the gradient for a more complex neural network such as a CNN, the same principles of backpropagation are used. The gradient term calculations become more complicated due to the convolutional nature of the network, but the backwards propagation of calculation is still used to determine the overall gradient (Bouvrie, 2006). After using backpropagation to determine the gradient terms of a CNN, gradient descent can be used to optimize the network's weights and biases.

# Chapter 3
# CNN-Based Recovery of RF Channels

## 3.1 Chapter Overview

The purpose of this chapter is to introduce the proposed RF recovery framework alongside the physical principles that enable its operation. First, the RF structure redundancies that can be exploited to infer additional channels are explained. Second, the overall RF inference framework and the details of the branching encoder-decoder CNN architecture are outlined. Third, the specific details used to create, train, and evaluate CNNs for downsampling levels of 2X, 3X, and 4X are described.

## 3.2 RF Redundancy: Shared Reflections in Received Channels

Even in high degrees of channel-wise downsampling, redundancies in received RF images provide echo location information that can be exploited to produced additional RF data. After a medium is insonified, the echo signals from a given scatterer should be received on multiple channels at different points in time (Xiao *et al.*, 2022). When received channels of RF data are stacked together into an RF image, the similar time-delayed echo signals may become distinguishable in the overall image. As shown in Figure 3.1 (a) and (b), reflections from a point target manifest themselves in a hyperbola due to the ToF differences for them to reach each channel (the RF image is logarithmically scaled here to make this hyperbola more prominent). This sharing of information amongst channels in RF images is the main phenomenon that will be exploited to infer RF data from missing channels. Figure 3.1 shows that even though images beamformed with a subset of channels will contain spatial aliasing artifacts (as explained in section 2.3.2.3), the corresponding RF images still contain discernable hyperbolas, even at downsampling rates as high as 4X. Therefore, by matching similar signals on neighboring channels and leveraging the distance between elements/channels, one should be able to infer the RF data for the missing channels in this RF image.

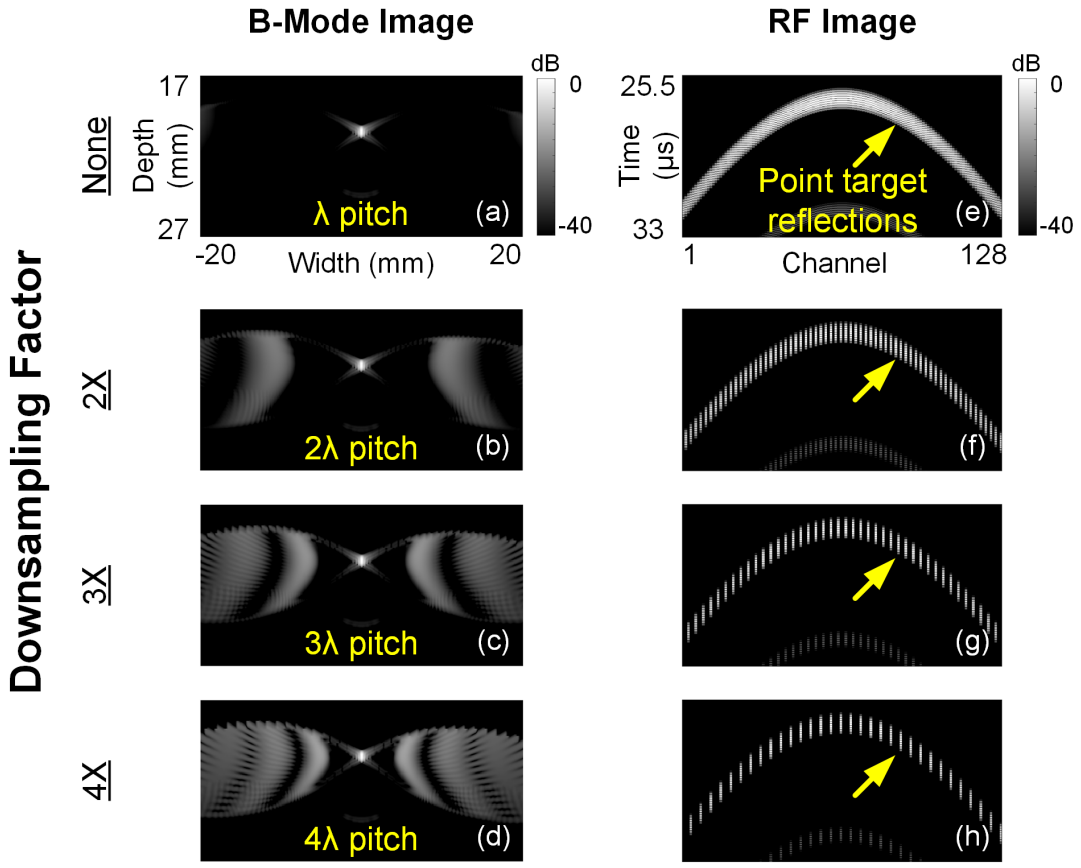


Figure 3.1. Hyperbolas in downsampled RF images. (a)-(d): B-mode images of point targets beamformed with a linear array (λ pitch) after 1X (none), 2X, 3X, and 4X channel-wise downsampling. (e)-(g): corresponding logarithmically scaled RF images. Missing channels are given values of 0 (filled in black).

### 3.3 Branched Encoder-Decoder CNNs for RF Channel Inference

A machine learning approach to RF recovery was taken to account for potential inconsistencies in each channel's received signals. While hyperbolas from a given echo event may still be present in substantially downsampled RF images, acoustic nonlinearities, anisotropic scattering, directivity of transducer elements, and interfering signals from other scatterers may cause differences in the signals received on each channel. These differences are expected to be magnified at higher levels of downsampling, since further distances are traveled for the echo signal to reach each channel. By providing examples of received RF data from realistic imaging scenarios, machine learning can be leveraged to determine the underlying patterns of the RF data, allowing inference of additional RF channels.

To facilitate RF inference from multiple degrees of channel-wise downsampling, novel branching encoder-decoder CNN architectures were developed. CNNs are capable of learning from raw spatiotemporal data such as RF data (Wang *et al.*, 2020), and our group has previously shown that an encoder-decoder architecture can effectively produce omitted RF channel data from half of a received subset (Xiao *et al.*, 2022). To facilitate larger degrees of RF inference, novel CNN architectures that contain a more complex encoding stage followed by a branched decoding segment that produces multiple outputs were developed. The details of the overall recovery framework and the CNN architecture are described in the following subsections.

### 3.3.1 Overall RF Recovery Framework

A branched recovery scheme was constructed to facilitate RF recovery from multiple levels of uniform downsampling. The overall framework is shown in Figure 3.2. First, a set of uniformly downsampled RF data from a steered plane wave transmission is placed into an N×(C/D)×1 RF image. In the RF image, N corresponds to the number of samples received on each channel, C is the number of channels in the full receiver array, and D is the degree of channel-wise downsampling. After the preprocessing step, the input RF image is passed into an encoder-decoder CNN that branches to provide sets of output RF images, where each output is a set of RF data with equal dimensions to the original input. The physical meaning of these output sets is shown by the colour-coded channel outputs in Figure 3.2; each output corresponds to an offset set of channel data that has the same pitch as the input array. The number of branches is dependent on D, where D-1 branches are needed to recover a full set of RF data. The framework's number of branches can be easily adjusted to accommodate different degrees of uniform

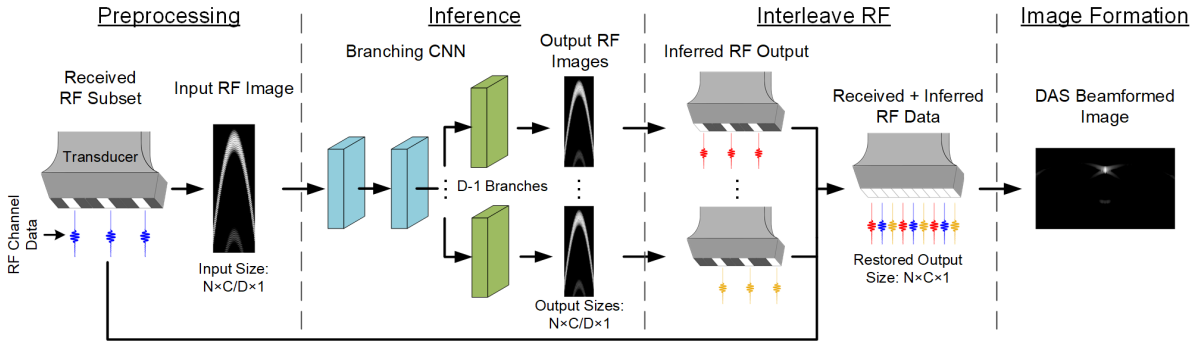


Figure 3.2. The proposed recovery framework. Downsampled RF subsets are placed into an RF image and fed into a CNN. Outputs from the network correspond to offset subsets of RF data, which are interleaved with the network input to recover a full set of RF data. This recovered set of RF data can then be beamformed using standard DAS beamforming. In this Figure the RF images and the DAS beamformed image are logarithmically scaled for visualization purposes.

downsampling, where the only requirement is that the full number of channels C is divisible by the desired downsampling degree D. When this requirement is not met, extra channels can be omitted from the recovery process and optionally added back during the framework's interleave step. After inference, the branched output sets of RF data can be interleaved together to produce the full set of RF data. From here, the RF data can be used for any desired image formation or analysis. In this work, the focus is on image formation with DAS beamforming, and coherent plane wave compounding.

### 3.3.2 Branched Encoder-Decoder CNNs

By performing the RF inference step, the branching encoder-decoder CNN is the key component that enables operation of the RF recovery framework. A detailed diagram of the CNN architecture is given in Figure 3.3 and key features of the architecture are described in the following subsections.

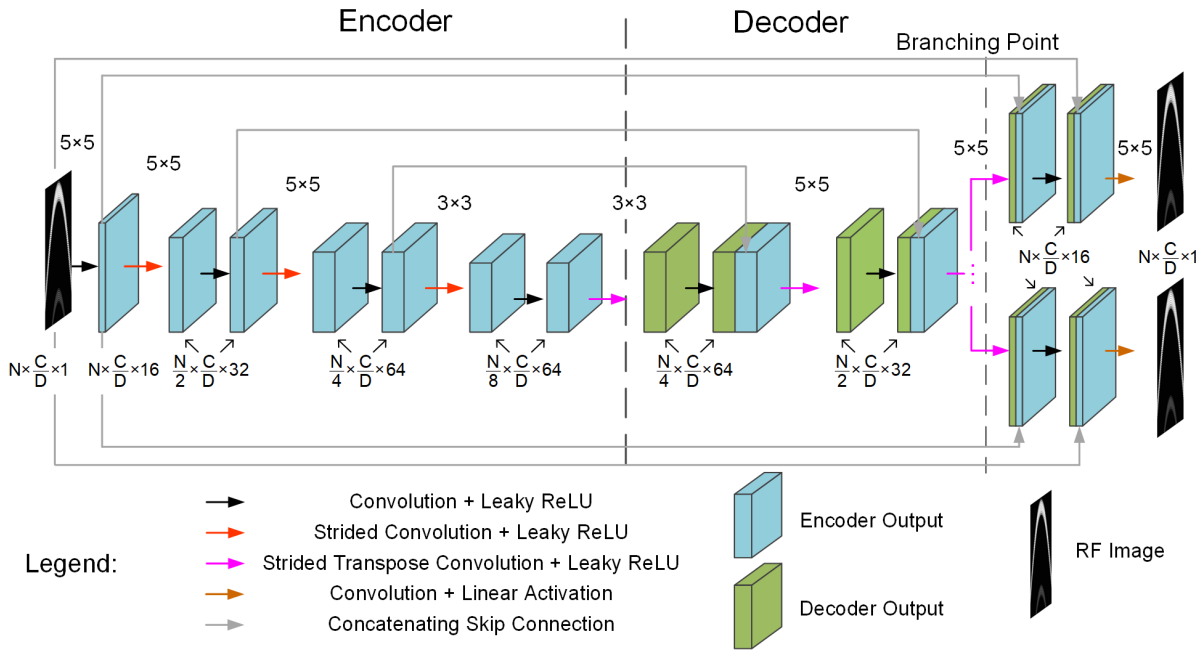


Figure 3.3. Encoder-Decoder architecture used for RF inference. Convolutional operations, activations, and concatenations are shown by the color-coded arrows. Filter sizes used for convolutions are given above the operation. If there is no filter size given, the filter size from the previous operation is used. Feature maps are represented by blocks, with their dimensions indicated below each block (with an exception in the branched region where the label is between branches). Feature map dimensions for concatenated maps refer to the convolutional output only. Note that the RF Images are logarithmically scaled for visualization purposes.

#### 3.3.2.1 Encoder-Decoder Structure

An encoder-decoder CNN is organized into two sections: an encoder that extracts compressed features from input data; and a decoder that produces an output based on these compressed features. This section-based approach to inference has yielded success when predicting missing sections of natural images, also known as inpainting (Elharrouss *et al.*, 2020). To infer missing channel data from downsampled RF images, similar principles were adopted: the proposed networks employ strided convolutions to encode the received RF subset into a compact feature representation, and then strided transpose convolutions decode this compressed representation into additional sets of RF data.

#### 3.3.2.2 Network Input: Downsampled RF Image

The RF image input to the CNN only contains the received RF channels, stacked side by side. If downsampled RF channels are selected with a uniform sampling scheme, each column is separated by an equal pitch and there is no need to encode information on missing channel location with additional columns in the RF image.

#### 3.3.2.3 Network Depth, Filters, and Activations

The depth, filter sizes, and filter depths of the architecture were chosen to balance computational cost and network complexity. The encoder segment for a network employs 7 layers, with strided convolutions used to compress the features in the vertical direction (along each channel). The earlier layers use a filter size of 5×5 to capture a large receptive field early, and later layers use a filter size of 3×3 to reduce computational cost. This results in a lateral receptive field of 25 channels at the output of each network's encoder. If RF data is being inferred for a 128-element probe, this means that the compressed features output by an encoder segment cover 38% (49/128 elements) of the transducer width for 2X downsampling, 58% (74/128 elements) for 3X downsampling, and 77% (99/128 elements) for 4X downsampling. The expanding coverage of the receptive field acts to bolster a network's inference ability in the face of larger degrees of downsampling. To increase the complexity of features learned throughout the networks, nonlinear activations are used after convolutional operations, employing leaky rectified linear unit (leaky ReLU) activations (Maas *et al.*, 2013) with a negative slope of 0.01. Feature depth is grown from 1 to 64 throughout the encoder segment to gradually increase the number of learned features alongside their complexity. After encoding, the decoder segments of each network use an additional 7 layers to infer RF sets from these encoded features. Filter sizes/depths are

constructed in a pattern that mirrors the encoder, and strided transpose convolutions are used to upsample the network in the vertical direction.

#### 3.3.2.4 Network Outputs, Branching Decoders, and Feature Sharing

A branching scheme is used in the decoder segment to 1) output RF sets with the same dimensions as the input and to 2) provide each output RF set with specialized upsampling and inference filters. With the branching scheme, the path from the input RF set to an individual branch's output RF set forms a symmetrical encoder-decoder CNN (Ronneberger *et al.*, 2015; Mao *et al.*, 2016; Liu *et al.*, 2018). Orientation of the CNN in this manner enables the inputs to the CNN to only include the downsampled channels (with advantages explained in section 3.3.2.2), while also enabling symmetrical sharing of encoder/decoder features via concatenating skip connections. This feature sharing restores some of the information lost in the encoding scheme of the network, and it also promotes more stable training by enabling direct pathways for the gradient back through the network (Mao *et al.*, 2016). The network's branching point is placed midway through the decoder segment to enable feature sharing between outputs during the first half of the decoder while still allowing each individual output RF set to be decoded with its own set of specialized upsampling/inference filters. In accordance with the overall framework, the number of branches is dependent on the downsampling degree and is given by D-1. At the end of each branch, the output sets of RF data are attained through a final convolution followed by a linear activation, allowing positive and negative RF values as outputs.

## 3.4 Training Dataset Acquisition, Cleaning, and Preprocessing

To facilitate the training of the recovery framework in a supervised learning fashion, a dataset of input/output RF pairs first needed to be acquired. This subsection describes the process taken to acquire, clean, and prepare a training dataset for the RF recovery framework.

### 3.4.1 Dataset Acquisition

A dataset of in vivo carotid artery scans from a SonixTouch research scanner (SonixTouch; Analogic Ultrasound; Peabody, MA, USA) was used to train the RF recovery architectures. This dataset consisted of steered plane wave acquisitions (-15° to 15° range, 1° separation) of 7 volunteers' (age: 25.9 ± 4.9) carotid arteries. The acquisition orientations for the training dataset are shown in Figure 3.4, where both the short and long axis of the carotid were scanned. The research scanner was programmed to acquire 31 steered angles at a time, and a total of 67301 separate RF frames (range: -2048 and 2047 with 12-

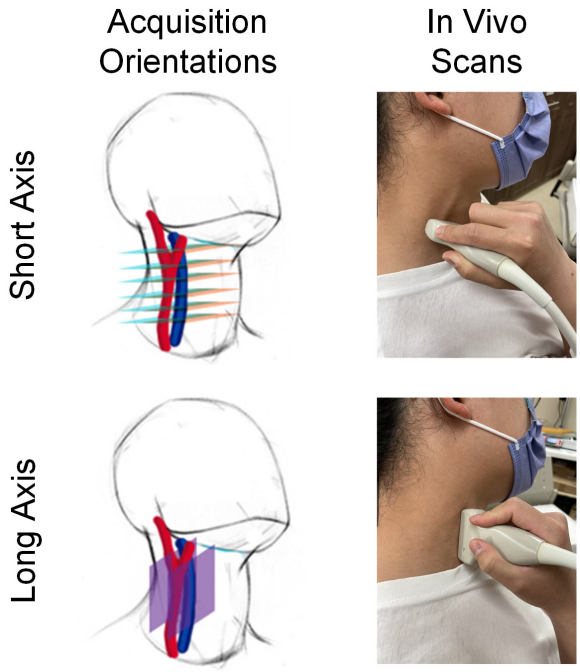


Figure 3.4. Training dataset acquisition area. Steered plane wave acquisitions were taken of volunteer's necks. Both short axis and long axis scans were acquired. Parameters used in acquisitions are summarized in Table 3.1.

# Table 3.1

**Training Data Acquisition Parameters**

| Parameter | Details |
|---|---|
| Ultrasound Scanner | SonixTouch |
| Ultrasound Probe | L14-5 |
| RF Data Range and Resolution | -2048 to 2047, 12 bits |
| Number of Tx/Rx Channels | 128 |
| Array Pitch | 0.3048mm |
| Transmit Frequency | 5MHz |
| Transmit Angle Range | -15° to 15°, 1° separation |
| Sampling Rate | 20MHz |
| Imaging Depth | 60mm |
| Pulse Repetition Frequency | 10kHz |

bit resolution) were acquired to train the networks. The acquisition was performed with a 128-element L14-5 probe; the system operated with a 5MHz transmission frequency, 2-pulse transmissions, a 10kHz pulse repetition frequency, and a 20MHz sampling rate. A summary of acquisition parameters can be found in Table 3.1. Acquired data was transferred to a computer server (SYS-4028-TRT; Super Micro, San Jose, CA, USA) with a Xeon E5-2620 central processing unit (Intel, Santa Clara, CA, USA) to be preprocessed in MATLAB (ver. 2020b; MathWorks, Natick, MA, USA).

### 3.4.2 Dataset Cleaning

The acquired dataset was cleaned and preprocessed to facilitate a stable and effective training process, with steps shown in Figure 3.5. First, the initial 196 samples of each channel were removed from RF images to eliminate the amount of near field reflections present in the training RF images. This resulted in N = 1304 samples per channel. Second, the dataset was cleaned by removing RF frames that had an excess of clipped samples. Training frames with more than 50 samples valued at 2047 (the maximum output value of the Sonixtouch's ADC) were removed from the training dataset. The removal of excessively clipped frames was meant to prevent the network from prioritizing inference of high-amplitude clipped regions. This removal resulted in 59864 training frames, down from 67301 (89% of the original set). The cleaned subset of training data was then normalized to be between -0.5 and 0.5 prior to being input into the network.

### 3.4.3 Dataset Preprocessing

CNN inputs and outputs were formed by selecting uniformly spaced channels from an RF frame and placing them into smaller RF images. For 2X downsampled data, odd numbered channels were selected as inputs with even channels selected as network outputs (each sized N×C/D = 1304×64). For 3X downsampled data, 3 groups of uniformly spaced channels with 2-element separation were formed (N×C/D = 1304×42); this resulted in one group corresponding to the input and two groups for the 2 branched framework outputs. The first and last channels of the full RF set were discarded to ensure a uniform downsampling scheme where inputs and outputs to the network all had the same size (C = 126 for 3X downsampling). Lastly, for 4X downsampled data, 4 groups of uniformly spaced channels with 3-element separation were formed (N×C/D = 1304×32), denoting the input and 3 output branches of the network.

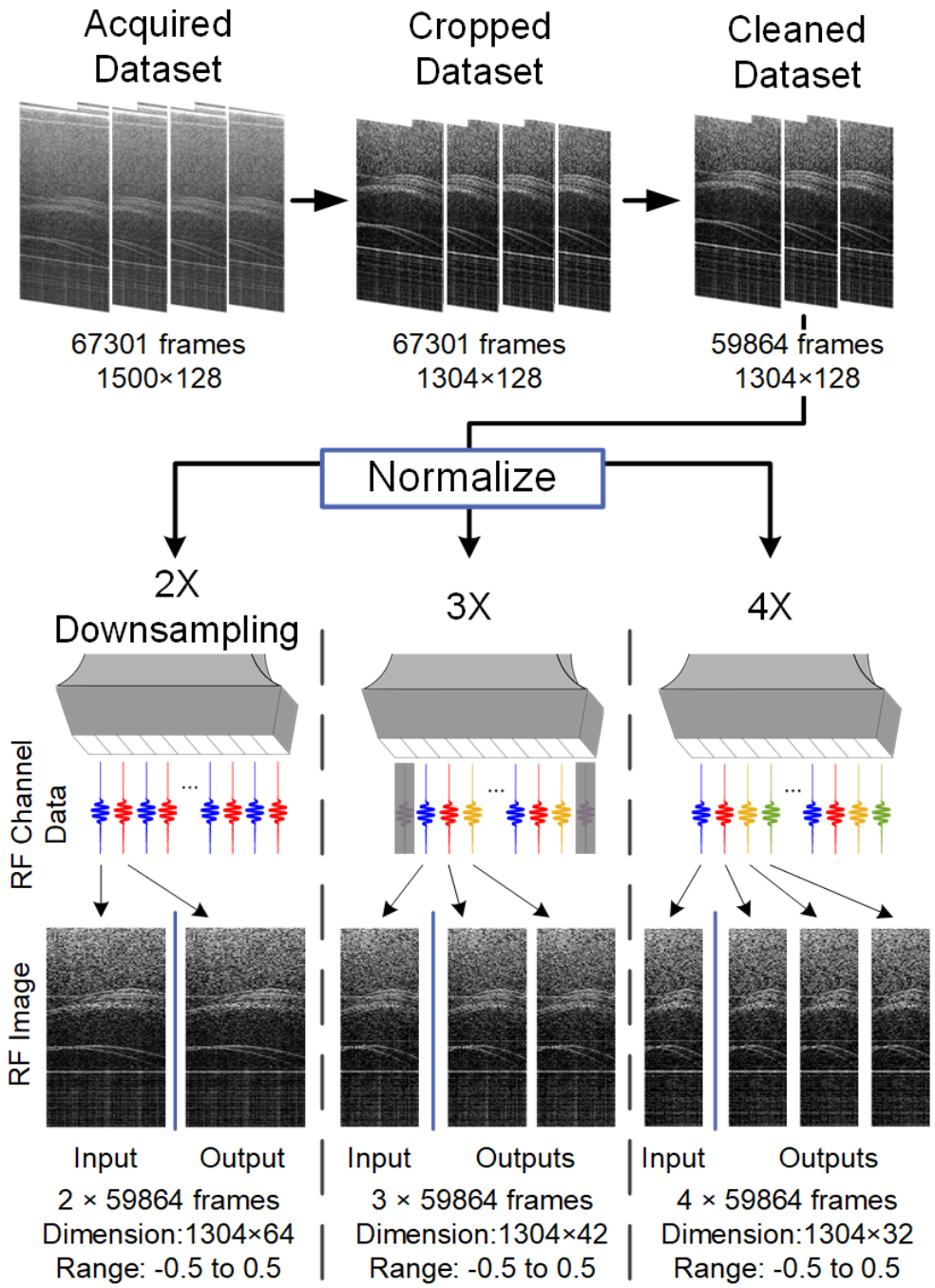


Figure 3.5. CNN training dataset preprocessing pipeline. Data was cropped, cleaned to remove heavily clipped frames, normalized, and parsed by channel for different downsampling scenarios. Note that the RF Images are logarithmically scaled for visualization purposes.

## 3.5 CNN Creation

Networks for each downsampling level were developed and trained in Python (ver. 3.6.7), utilizing Tensorflow-GPU (ver. 1.12.0) with the Keras (ver. 2.1.6) application-user-interface. The networks (shown in Figure 3.6) were created to conform with the overall architecture outlined in Figure 3.3: for 2X downsampling there was 1 branch, for 3X downsampling there were 2 branches; for 4X downsampling there were 3 branches. The dimensions of the input to each CNN conformed to the training dataset created for each degree of downsampling (shown at the bottom of Figure 3.5). The encoder section of each network is the same and it is omitted from the figure.

## 3.6 CNN Training

The training of the networks was facilitated using an RTX-1080 GPU (Nvidia, Santa Clara, CA, USA). Each layer's weights were initialized according to a zero-mean uniform distribution with 2/k variance, where k is the number of inputs to the layer (He, Zhang, Ren & Sun, 2015), and then the Adam optimization algorithm (Kingma & Ba, 2014) was used for gradient-based training; each network employed a learning rate of 0.001, a batch size of 32, and 50 total epochs. The mean-absolute-error (MAE) from each branch's output was added together with equal weighting to form the overall loss function. 90% of the cleaned dataset was used for training with the remaining 10% used for validation of each network. The loss over the training and validation sets over the 50 epochs are plotted in Figure 3.7.

It can be seen in Figure 3.7 that the overall losses on the training and validation data followed similar trajectories over the training process, with each of them plateauing as 50 epochs were approached. For the MAE loss at the output of each branch (which are added to form the overall loss), similar trajectories were also observed, although branch 2 experienced a higher error compared to branch 1 and 3 for the 4X downsampling case. This is expected since the outputs predicted from branch 2 are not adjacent to any input channels to the network (as is the case for branch 1 and 3). This results in a more difficult inference problem, as there is a larger offset between the predicted channels from branch 2 and the input channels to the network. Nonetheless, the loss is progressively minimized over the training process for this branch, albeit with a higher overall MAE.

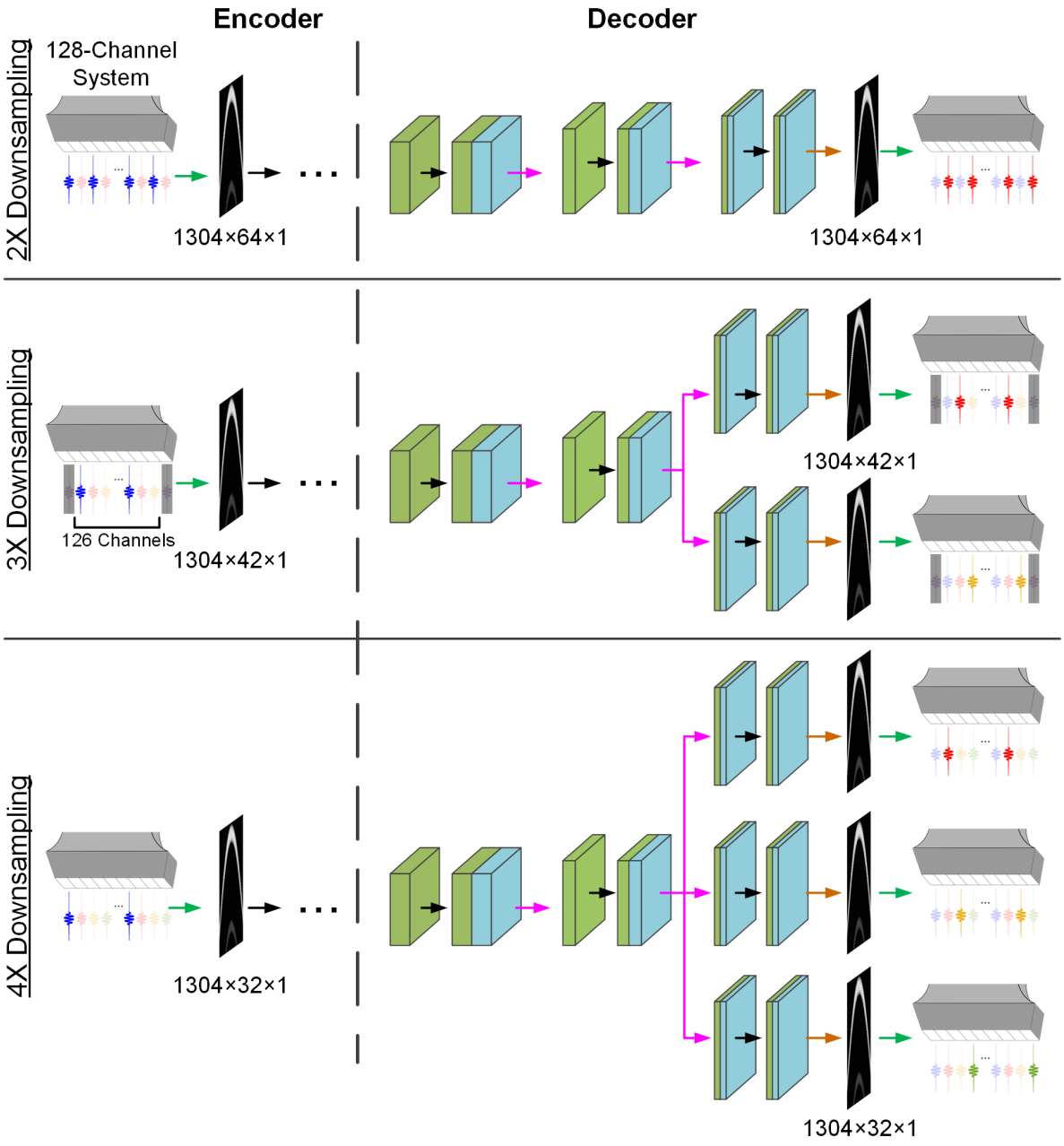


Figure 3.6. Different CNNs for different levels of downsampling. CNN architecture configurations and dimensions are the same as outlined in Figure 3.3. Arrow colours also correspond to the legend in Figure 3.3, except for the green arrow, which refers to arranging RF channel data into an RF image.

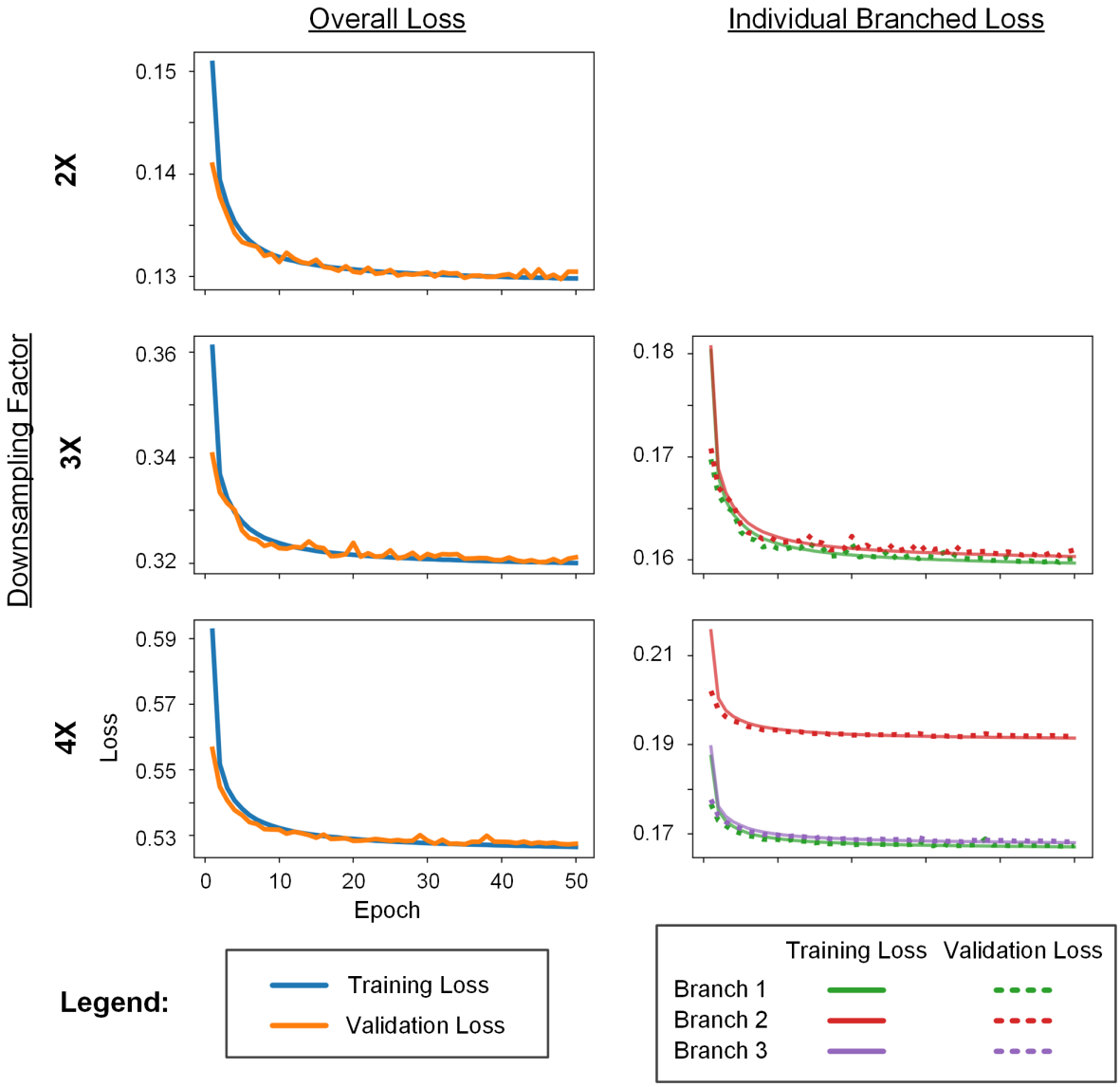


Figure 3.7. Training process for each of the CNNs displayed in Figure 3.6. Individual branched losses in the right column are added to form the overall loss shown in the left column.

## 3.7 Recovery Framework Performance Evaluation

The effectiveness of the RF recovery framework was assessed using both raw RF signal analysis and quality comparisons of DAS beamformed images. For image analysis, images were beamformed with 1) only the downsampled subset of RF channels, 2) the downsampled RF channels + CNN-inferred RF channels, and 3) the full original set of RF channels. Additional image quality analysis was performed on compounded images, where angled transmissions from each of the 3 groups were compounded together. Evaluation of metrics on individual RF data and beamformed images was performed in MATLAB (ver. 2020b) with additional statistical testing performed in R (ver. 4.0.5).

### 3.7.1 Evaluation Scenarios

Additional RF datasets were acquired from several different imaging scenarios to evaluate the recovery framework's success. Firstly, 9 additional carotid scans from separate volunteers were acquired for evaluating RF reconstruction success and spatial aliasing artifact reduction on subjects apart from the training set. These acquisitions were performed using the same imaging parameters as the training data (outlined in Table 3.1). Secondly, point target phantom scans were acquired for evaluating RF inference success and spatial aliasing artifact reduction *in vitro*. The acquisitions were performed using the parameters in Table 3.1, except for the transmission angles, which were taken from -14.75°:0.5°:14.75° to test the framework's generalizability to unseen transmission angles. Lastly, an additional 3-cycle pulse acquisition of a volunteer's quadriceps muscle was performed for evaluating spatial aliasing artifact reduction in acquisitions from a different *in vivo* region. This acquisition was performed using the imaging parameters in Table 3.1, except for a 3-cycle pulse transmission, which was chosen to evaluate the framework's generalizability to acquisitions from a different transmission scheme. All scenarios were evaluated at downsampling/recovery levels of 2X, 3X and 4X.

### 3.7.2 Evaluation Data Preparation

Additional RF acquisitions were cropped to have a length of 1304 samples, where the cropping range was chosen to retain samples for each acquisition's desired imaging depth. RF data was normalized and partitioned in the same manner as the training data, and downsampled RF images were then fed into the trained networks to recover a full set of RF data for each level of downsampling. For subsequent image analysis, RF data was first bandpass filtered between 3-7MHz and converted to an analytic signal with the Hilbert transform. The analytic data was then DAS beamformed with an F number of 1mm and rectangular apodization. Key parameters from the beamforming process are summarized in Table

3.2. After beamforming, the final step in the ultrasound image formation procedure was to logarithmically-scale the beamformed pixel values to compress the image's dynamic range.

**Table 3.2**

**Image Beamforming Parameters**

| Parameter | Details |
|---|---|
| Prefilter Passband | 3-7 MHz |
| Filter Design Method | Equiripple (30$^{th}$ Order) |
| Apodization | Rectangular |
| Focusing | Constant $F_{mm}$ = 1 |
| Image Resolution (Axial and Lateral) | 0.1mm/pixel |

### 3.7.3 Beamformed Image Quality Evaluation

#### 3.7.3.1 Full Reference Image Quality

To assess the overall image quality restoration due to RF recovery, the structural similarity measure (SSIM; Wang *et al.*, 2004) was evaluated on the displayed ultrasound images. For SSIM calculation, pixel values from the DAS operation were logarithmically scaled and clipped to values within the displayed dynamic range. Images beamformed with a full set of receiving RF data were used as reference. The total quantity for the SSIM is taken from the average of several windowed SSIM calculations, with the windowed calculation as follows:

$$SSIM = \frac{\left(2\mu_{\hat{y}}\mu_y + (K_1R)^2\right)\left(2\sigma_{\hat{y}y} + (K_2R)^2\right)}{\left(\mu_{\hat{y}}^2 + \mu_y^2 + (K_1R)^2\right)\left(\sigma_{\hat{y}}^2 + \sigma_y^2 + (K_2R)^2\right)} \tag{3.1}$$

For the SSIM calculation in equation 3.1, $\mu_{\hat{y}}$ and $\mu_y$ correspond to the means of an image window and its reference window, respectively. $\sigma_{\hat{y}}$ and $\sigma_y$ correspond to the variances of these windows, while $\sigma_{\hat{y}y}$ is the covariance between the two windows. The KR terms are used to stabilize divisions when the

denominator is small; R is the image's dynamic range, and $K_1$ and $K_2$ correspond to small (<<1) constants. In this evaluation, R was set to the displayed dynamic range of 50dB. Following the methods used in (Wang *et al.*, 2004), 0.01 and 0.03 were used for $K_1$ and $K_2$ and a window size of 11x11 was employed.

### 3.7.3.2 Resolution

To evaluate the resolution of beamformed ultrasound images, the lateral full width at half maximum (FWHM) was calculated for beamformed point targets. FWHM is expressed as the width that a point spans between half of its maximum value on each side. Since the beamformed images are logarithmically scaled, this was calculated as the distance between a 6dB drop on each side of a point target. An example of the lateral FWHM calculation for a point target is shown in Figure 3.8.

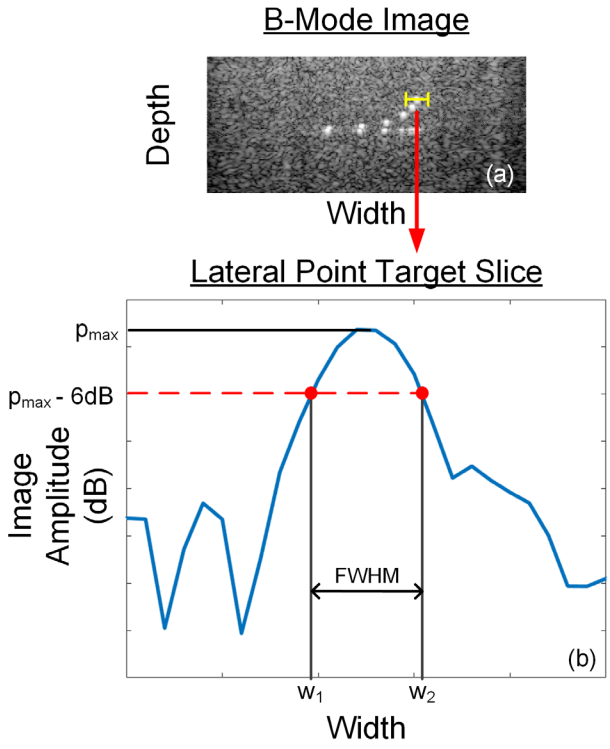


Figure 3.8. FWHM measurement of a point target. A lateral slice of the point target in (a) is displayed in (b). The width between a 6dB drop from the maximum on each side denotes the FWHM.

### 3.7.3.3 Contrast

*In Vitro Contrast Analysis*

The average image amplitude in spatially aliased regions was evaluated to determine the degree of *in vitro* artifact reduction provided by CNN-inferred RF data. Spatially aliased regions were chosen on images formed with a subset of receiving channels; the maximum value of a region that did not contain point targets was used as the center of the spatially aliased selection. A 4mm×4mm spatially aliased region was then centered around the maximum value. An example selection for the 2X downsampled -0.25° point target transmission is shown in Figure 3.9. The overall region examined for spatial aliasing artifacts is shown in the yellow box, with the green box denoting the selected spatially aliased region.

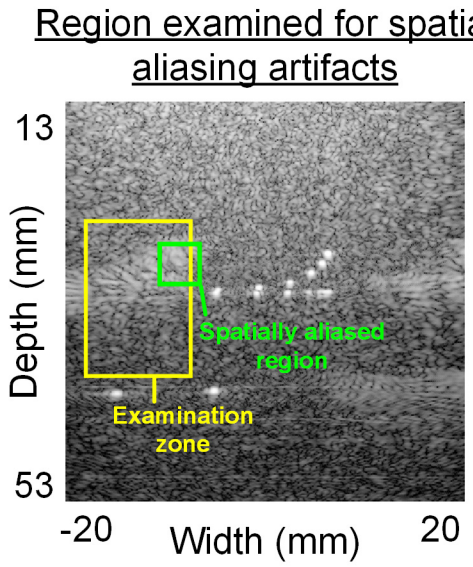


Figure 3.9. Spatially aliased region selection for a -0.25° point target transmission that was beamformed with half of the receiving channels.

### *In Vivo Contrast Analysis*

The contrast ratio (CR) of the carotid artery lumen to surrounding tissue regions was taken to evaluate the reduction of spatial aliasing artifacts *in vivo*. CR is measured as:

$$CR = \mu_{tissue} - \mu_{lumen} \tag{3.2}$$

where $\mu_{tissue}$ and $\mu_{lumen}$ correspond to the mean of a tissue and a lumen patch in the beamformed and logarithmically scaled ultrasound image. Contrast assessments were taken using both the thyroid and the carotid wall as tissue references, where the reference regions were manually segmented on fully

compounded images. Circular regions were selected in the center of the carotid lumen (L) and thyroid (T) for their reference regions, and the most reflective part of the bottom of the carotid wall (W) was selected for its reference region. An example segmentation is shown in Figure 3.10. CR changes were examined at both the individual level and over the larger sample of carotid scans from 9 volunteers. The segmentations for each volunteer can be found in Appendix A.

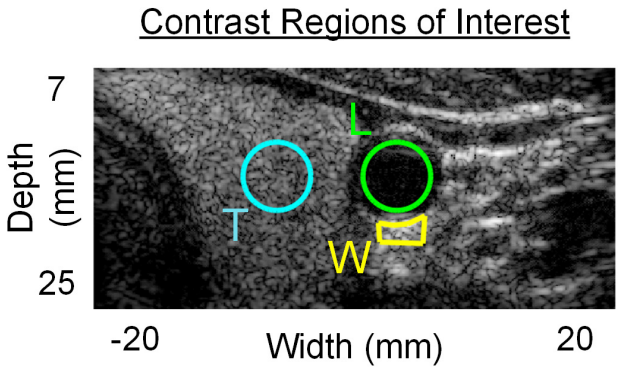


Figure 3.10. ROI Selection for contrast ratio measurement. L, T, and W denote the lumen, thyroid, and carotid wall, respectively.

#### 3.7.3.4 Statistical Substantiation of *In Vivo* Artifact Reduction

To determine if the inclusion of CNN-inferred RF data during beamforming provided a statistically significant change in spatial aliasing artifact amplitude, parametric testing was performed over the 0° acquisitions for each volunteer's carotid. Statistical testing was performed on the isolated set of 0° transmissions to ensure independence between samples, and to prevent making a large degree of comparisons, which would increase the chance of error being made during multiple hypothesis testing.

##### *Subject Grouping*

CR measurements from the 0° carotid artery scans of the 9 volunteers were grouped based on reference tissue type (T-L and W-L), image formation method (beamformed with a subset of receiving channels, beamformed with a subset of receiving channels + CNN-inferred channels), and downsampling degree (2X, 3X, 4X). This resulted in a total of 12 groups each with 9 subjects. Groups were paired based on downsampling level, transmission angle, and tissue type; for example, each subject's T-L CR after 2X downsampling and no CNN recovery was paired with their corresponding T-L CR after 2X downsampling with CNN RF recovery.

*Statistical Methods*

For the parametric test, a paired t-test (Student, 1908) was employed to determine if the inclusion of CNN-inferred RF data statistically changed the CR of beamformed images. Since multiple comparisons were being made between multiple tissue types, levels of downsampling, and transmission angles, Bonferonni correction (Armstrong, 2014) was employed to adjust the significance level α of the test. To describe the results of the paired t-test, p-values, confidence intervals (Bonferonni-corrected), and Cohen's d effect size (Lee, 2016) were all reported for each pairwise comparison. To ensure that samples conformed with the parametric assumptions of normality, a Shapiro-Wilk test (Shapiro & Wilk, 1965) was used on each pair's CR difference measurements.

### 3.7.4 RF Recovery Metrics

#### 3.7.4.1 Root Mean Squared

To characterize the RF data received from different tissue regions, the root mean squared (RMS) measure was used:

$$RMS = \sqrt{\frac{1}{n}\sum_{i=1}^{n} y_i^2} \tag{3.3}$$

where n refers to the total number of RF samples being investigated, and $y_i$ is a given sample of the original RF data. This metric provides an idea of the overall magnitude of reflections captured in a particular region of an RF image.

#### 3.7.4.2 Normalized Root Mean Squared

To compare the recovery error between different RF regions, the normalized root mean squared error (NRMSE) was used:

$$NRMSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}}{y_{max} - y_{min}} \tag{3.4}$$

where $y_i$ and $\hat{y}_i$ refer to the original and CNN-inferred RF samples, respectively. The denominator is used to normalize the error calculation, where $y_{max}$ and $y_{min}$ refer to the maximum and minimum of the original RF data values being examined. The normalization allows RF recovery comparison between multiple tissue types, despite potentially differing overall amplitudes due to each tissue's varying depth and echogenicity characteristics.

### 3.7.5 Compounding Evaluation

The compatibility of the recovery framework with the plane wave compounding process was evaluated by tracking image metric changes as beamformed images were compounded. Images were compounded by adding 1° transmissions sequentially, while omitting the 0° transmission due to a difference in the receiver delay time for these acquisitions. Therefore, the compounding pattern went as follows: 1 angle: -1°; 2-angle: -1°, 1°; 3-angle: -2, -1°, 1°; 4-angle: -2°, -1°, 1°, 2°; etc.

# Chapter 4
# Experimental Results of Receiver Channel Recovery

## 4.1 In Vitro Experimental Results

This subsection details the results from the *in vitro* RF inference experiments. First, an examination of RF inference success for hyperechogenic vs. hypoechogenic tissue is described. Second, B-mode images beamformed with and without CNN-inferred RF data are compared in terms of full reference quality, contrast, and resolution.

### 4.1.1 RF Analysis: More Successful Inference Along Hyperbolic Structure

For the *in vitro* point target RF data, relatively higher inference success was achieved on more hyperbolic RF structure. The B-mode image (full receiving channels) and the corresponding RF image for the point target phantom (-0.25° transmission) are shown in Figure 4.1 (a) and (b), where each point target in the B-mode image has a corresponding hyperbola in the RF image. Figure 4.1 (c)-(e) compare the single-channel inferred RF to the original received RF for each degree of downsampling. The channel under investigation is shown in yellow in (b) and the echo data is divided into pre-hyperbolic and hyperbolic reflections. The pre-hyperbolic reflections are from less reflective echoes from the point target "tissue", and the hyperbolic reflections are from the point targets themselves. Shown in (c)-(e), relatively higher inference success was achieved in the hyperbolic region of the channel data; the single channel inference follows the original RF much more closely once the hyperbola begins. This is reflected by the >2X lower NRMSE in the hyperbolic region for each level of downsampling.

### 4.1.2 B-Mode Point Target Full Image Comparison

Beamforming *in vitro* point target images with CNN-inferred RF data resulted in recovery of the underlying image structure. Beamformed point target images for a -0.25° and a -14.75° transmission are shown in Figure 4.2. Progressive downsampling (2X to 4X) of received RF channels resulted in worsening spatial aliasing artifacts that obscure the beamformed images, as shown in (e)-(g) and (l)-(n). Even though the point targets are visible in these images, spatial aliasing incorrectly results in the impression that there other highly reflective regions in the imaging medium. This is contrasted with the images beamformed with CNN inferred data, as shown in (b)-(d) and (i)-(k). Spatial aliasing artifacts are reduced, and it becomes more apparent that the only objects in the medium are the point targets

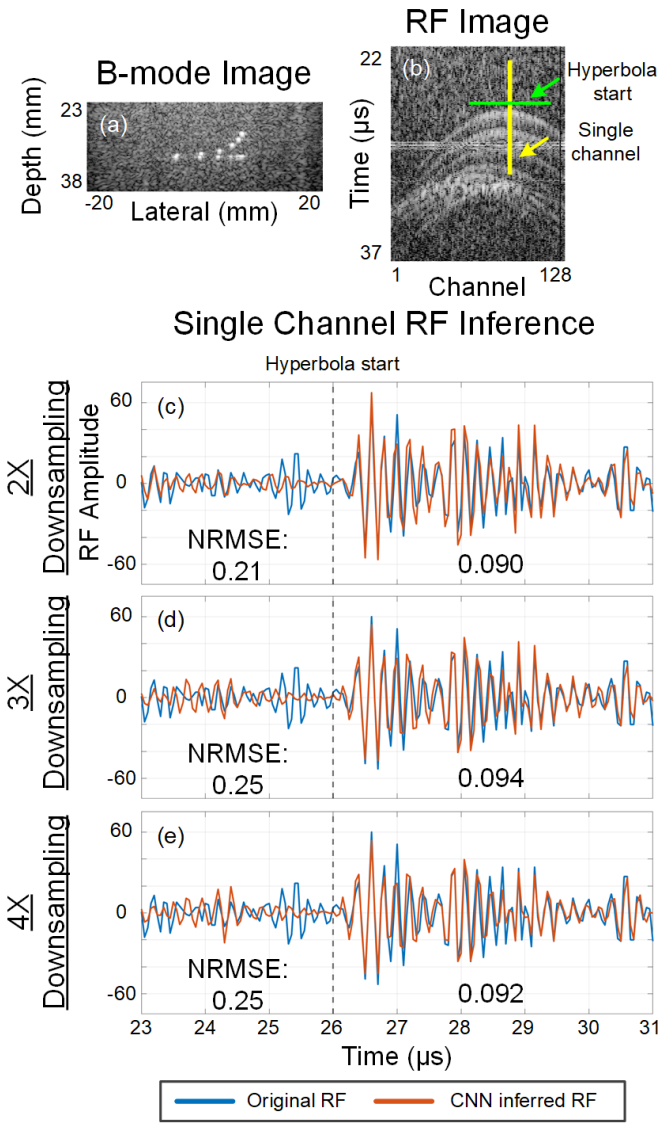


Figure 4.1. RF reconstruction evaluation for a 0° transmission of a point target phantom. (a) B-mode image of the point target image being examined, displayed with a 50dB dynamic range. (b) RF image of the point target with the channel examined highlighted in yellow, and the start of the hyperbolic point target reflections denoted in green. For visualization, the RF image is logarithmically scaled and displayed with a dynamic range of 40dB. (c)-(e) comparison of the original RF data to the inferred RF data for downsampling levels of 2X, 3X, and 4X, respectively.

themselves. This qualitative result is accompanied by a quantitative increase in SSIM; the SSIM measure increases for each beamformed image when CNN-inferred RF data is included during the beamforming process. While the spatial aliasing artifacts are alleviated, there is still an overall decrease in image quality with progressive downsampling, shown by the worsening SSIM values as the downsampling degree is increased.

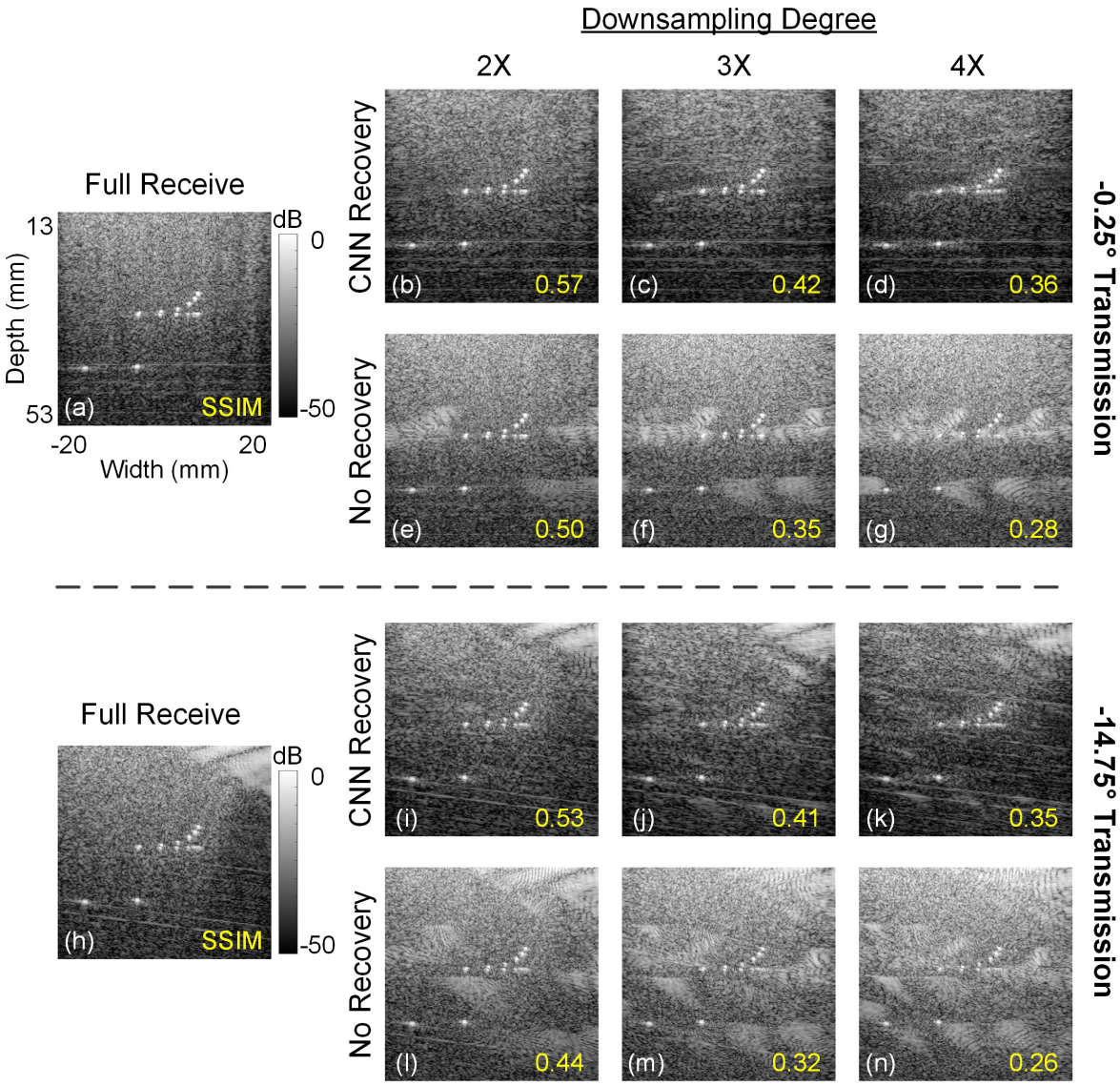


Figure 4.2. B-mode beamformed images of angled point target acquisitions. All images are displayed with a dynamic range of 50dB. (a) -0.25° transmission of the point target phantom, all receiving channels. (b)-(d) -0.25° point target images beamformed with 1/2 received + 1/2 inferred, 1/3 received + 2/3 inferred, and 1/4 received + 3/4 inferred channels. (e)-(g) -0.25° point target images beamformed with 1/2, 1/3, and 1/4 received channels. (h) -14.75° point target images, all receiving channels. (i)-(k) -14.75° point target images beamformed with 1/2 received + 1/2 inferred, 1/3 received + 2/3 inferred, and 1/4 received + 3/4 inferred channels. (l)-(n) -14.75° point target images beamformed with 1/2, 1/3, and 1/4 received channels.

### 4.1.3 Spatial Aliasing Artifact Reduction

Images beamformed with CNN-inferred RF data achieved suppressed spatial aliasing artifacts across multiple transmission angles. The average image amplitudes for spatially aliased regions are shown in Figure 4.3; average amplitude for 2X, 3X, and 4X downsampling are shown in (a), (b), and (c), with the selected spatially aliased image region for the -0.25° transmission case shown in (d), (e), and (f). As shown in (a)-(c), the inclusion of CNN-inferred RF data during beamforming suppresses the amplitude in the spatially aliased regions for all transmission angles. While the artifacts are removed when CNN-inferred RF data is used, the overall amplitude in these regions is slightly over-suppressed, as the image amplitude becomes lower than the case where the original received channel data is used. This over-suppression is substantially less than the heightened amplitude caused by the artifacts; the spatially aliased regions have average values that are 10.4dB, 11.8dB, and 12.9dB higher than the fully sampled case for downsampling levels of 2X, 3X, and 4X, respectively. This is compared to the regions beamformed with CNN-inferred data that have average values that are 1.8dB, 2.2dB, and 3.7dB lower than the fully sampled case for downsampling levels of 2X, 3X, and 4X, respectively.

### 4.1.4 Changes in Point Target Resolution

The resolution of the point targets experienced slight degradation as higher levels of CNN-inference were performed. The FWHMs for the 3 shallowest point targets from the -0.25° beamformed images (labelled in Figure 4.4 (a)) were evaluated and the results are shown in Table 4.1; compared to an average FWHM of 0.57 for the fully sampled case, the average FWHM was 0.58 (+1.8%), 0.60 (+5.2%), and 0.60 (+ 5.2%) for CNN recovery rates of 2X, 3X, and 4X. Qualitative examination of the bottom row of point targets (shown in Figure 4.4) shows that while resolution is mostly preserved, at higher levels of CNN recovery the point target's max value is reduced (the CNN-inferred image lines have slightly lower peaks than the fully sampled or just downsampled counterparts), which can cause a slight loss in FWHM/resolution.

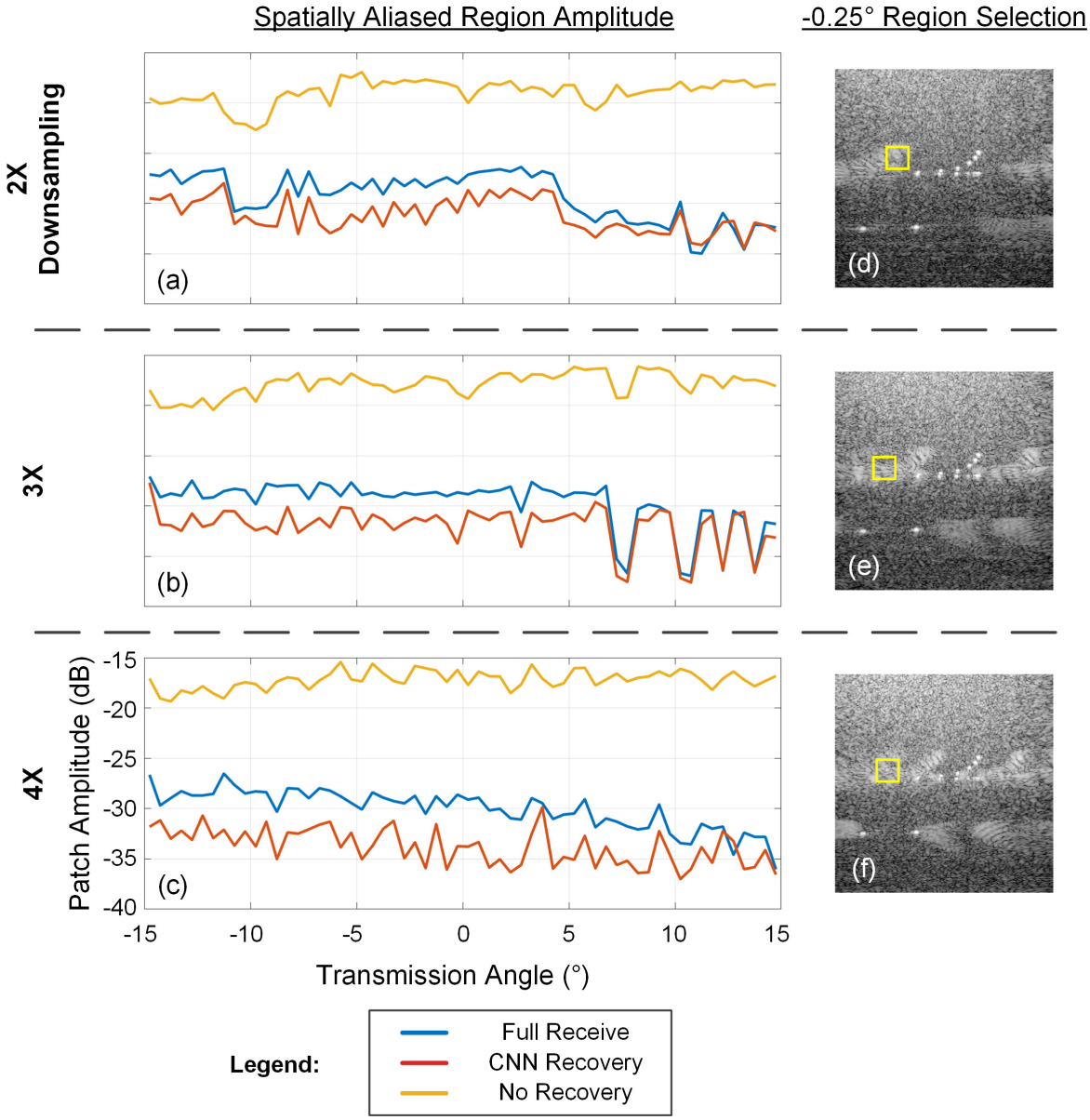


Figure 4.3. Spatial aliasing artifact reduction in point target images. Artifacts are evaluated over -14.75°:0.5°:14.75° transmission angles. (a)-(c) Artifact region mean amplitude comparisons for 2X, 3X, and 4X downsampling. (d)-(f) example region selection for a -0.25° transmission after 2X, 3X, and 4X receiver downsampling.

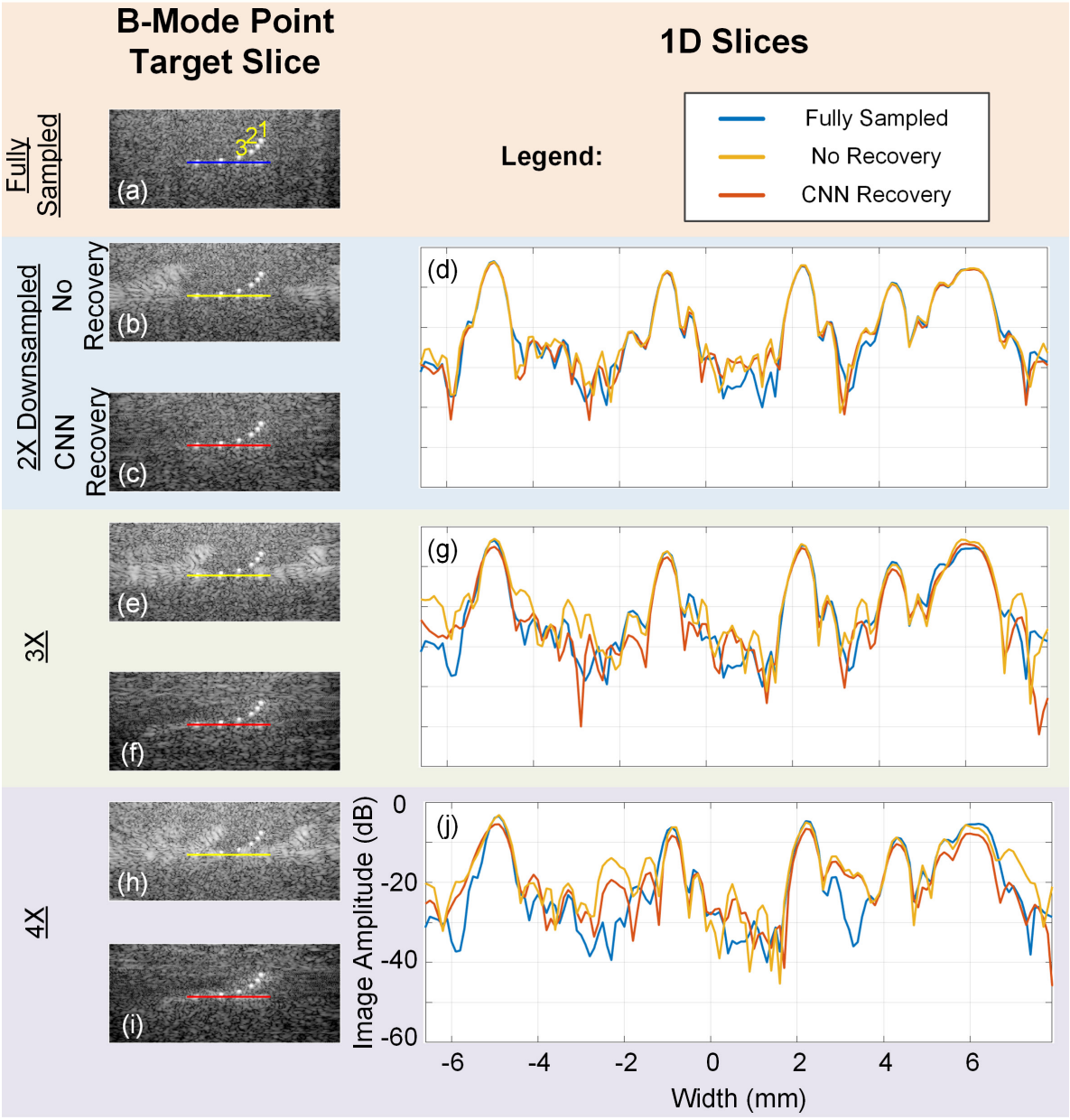


Figure 4.4. Lateral cross sections of beamformed point targets. (a) Reference slice of the -0.25° point target acquisition. Point targets evaluated for FWHM in Table 4.1 labelled. (b)-(c) Reference slices for point target images beamformed with 1/2 received and 1/2 received + 1/2 inferred RF data. (d) plots of the reference slices in (a)-(c). (e)-(f) Reference slices beamformed with 1/3 received and 1/3 received + 2/3 inferred RF data. (g) plots of the reference slices in (a), (e), and (f). (h) Reference slices beamformed with 1/4 received and 1/4 received + 3/4 inferred RF data. (j) plots of the reference slices in (a), (h), and (i).

# Table 4.1

**Single Transmission Point Target Resolution**

| Downsampling Degree | Recovery Type | Point 1 | Point 2 | Point 3 | Average |
|---|---|---|---|---|---|
| Fully Sampled | None | 0.57 | 0.57 | 0.56 | **0.57** |
| 2X | CNN | 0.59 | 0.59 | 0.57 | **0.58** |
| 2X | None | 0.56 | 0.57 | 0.57 | **0.57** |
| 3X | CNN | 0.6 | 0.59 | 0.61 | **0.60** |
| 3X | None | 0.56 | 0.57 | 0.59 | **0.57** |
| 4X | CNN | 0.59 | 0.65 | 0.56 | **0.60** |
| 4X | None | 0.55 | 0.63 | 0.56 | **0.58** |

## 4.2 In Vivo Experimental Results

This subsection details the results from the *in vivo* RF inference experiments. First, an examination of RF inference success for reflections from different tissue types is described. Second, B-mode images beamformed with and without CNN-inferred RF data are compared in terms of full reference quality and contrast. Third, image quality is examined as scans of a carotid artery are coherently compounded.

### 4.2.1 Raw RF Analysis

When performing *in vivo* imaging, more echogenic regions of the imaging medium produced RF data with hyperbolic structure. The B-mode image for a carotid/thyroid is shown in Figure 4.5 (a), with its corresponding RF image shown in (b). Different regions in the imaging medium are matched to their corresponding reflections in the RF image; 1 indicates the reflections from the hyperechogenic carotid wall, 2 indicates an echogenic region with lower amplitude reflections, and 3 indicates the less echogenic thyroid. The RMS values for each region's RF data are given in (b) to reflect the overall amplitude of reflections from these regions. As shown in (b), reflections from the echogenic carotid wall produce a more hyperbolic shape in the RF image, while reflections from the homogenous, less echogenic thyroid do not have any apparent structure in the RF image. Reflections from region 2 are lower amplitude (indicated by the lower RMS), but they are manifested in a clear hyperbolic shape.

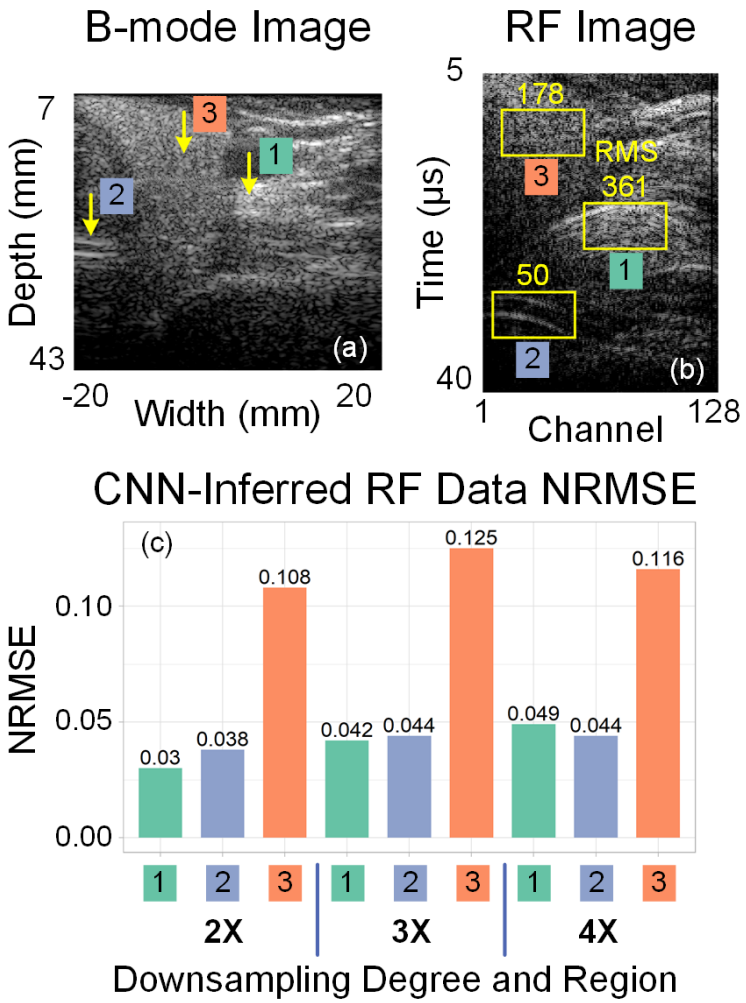


Figure 4.5. *In vivo* RF reconstruction evaluation for different regions of a carotid artery. (a) Highlighted tissue regions on a B-mode image, displayed with a dynamic range of 50dB. 1) carotid wall, 2) lower amplitude hyperechogenic region, and 3) thyroid. (b) RF reflections from each highlighted tissue region. Region # is displayed below its yellow bounding box, and the region's RMS is displayed above the box. For visualization, the RF image is logarithmically scaled and displayed with a dynamic range of 40dB. (c) NRMSE for the CNN-inferred RF data from each region.

Higher RF recovery success was achieved when inferring RF data from hyperbolic regions of the RF image. Figure 4.5 (c) displays the NRMSE values of inferred RF data from each of the regions highlighted in (a) and (b). Comparatively higher inference success was achieved in regions with hyperbolic RF data. Comparing the carotid wall (region 1) to the thyroid (region 3), the NRMSE was >2X lower for each degree of downsampling. Additionally, the NRMSE for region 2 was >2X lower than the NRMSE for the thyroid (region 3) for each degree of downsampling, despite the lower amplitude reflections in region 2.

### 4.2.2 B-Mode Full Image Comparison

Beamforming *in vivo* carotid and quadriceps images with additional CNN-inferred RF data revealed the underlying structure of the imaging medium. Beamformed images of carotid and quadriceps 0° transmissions are shown in Figure 4.6. Progressive downsampling (2X to 4X) of received RF channels resulted in worsening spatial aliasing artifacts that obscure the beamformed images, as shown in (e)-(g) and (l)-(n). These artifacts substantially reduce the visibility of the carotid lumen and the quadriceps muscle fibers. When CNN-inferred RF data is included in the beamforming process (shown in (b)-(d) and (i)-(k)), visibility of the underlying image structure is improved; the carotid lumen becomes visible, and the fibrous structure of the quadriceps is revealed. This qualitative result is accompanied by a quantitative increase in SSIM; the SSIM measure increases for each beamformed image when CNN-inferred RF data is included during the beamforming process. While the spatial aliasing artifacts are alleviated, there is still an overall decrease in image quality with progressive downsampling, shown by the worsening SSIM values as the downsampling degree is increased.

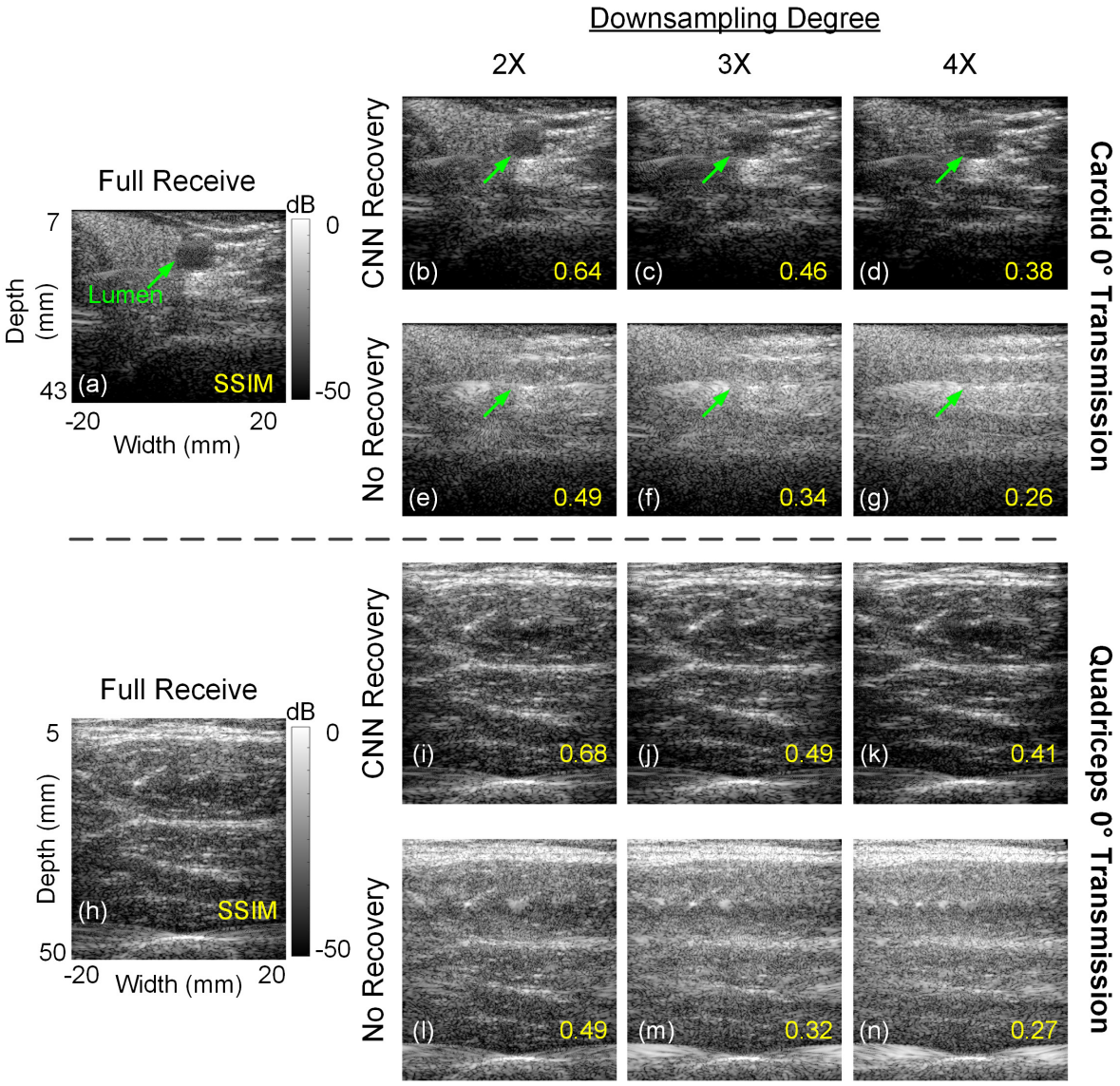


Figure 4.6. B-mode beamformed images from a carotid and quadriceps muscle. All images are displayed with a dynamic range of 50dB. (a) 0° transmission of a carotid artery with the lumen region highlighted, all receiving channels. (b)-(d) Carotid images beamformed with 1/2 received + 1/2 inferred, 1/3 received + 2/3 inferred, and 1/4 received + 3/4 inferred channels. (e)-(g) Carotid images beamformed with 1/2, 1/3, and 1/4 received channels. (h) 0° transmission of a quadriceps muscle, all receiving channels. (i)-(k) Quadriceps images beamformed with 1/2 received + 1/2 inferred, 1/3 received + 2/3 inferred, and 1/4 received + 3/4 inferred channels. (l)-(n) Quadriceps images beamformed with 1/2, 1/3, and 1/4 received channels.

### 4.2.3 Multi-Angle Contrast Analysis

#### 4.2.3.1 Single Acquisition

Suppression of spatial aliasing artifacts resulted in an increased contrast in the carotid artery over multiple transmission angles. The CR measurements for the carotid shown in Figure 4.7 (a) are displayed in (b) and (c). When the beamformed images are formed with downsampled sets of RF data, the wall-lumen (W-L) CR and the thyroid-lumen (T-L) CR are reduced across all transmission angles. When CNN-inferred RF data are included during beamforming, the W-L CR and T-L CR are improved across all transmission angles. Relatively higher recovery in contrast was achieved when the carotid wall was used as a reference, and this is reflected in the images of Figure 4.6 (b)-(d), as the thyroid loses some of its details at higher degrees of downsampling and CNN recovery.

#### 4.2.3.2 Multiple Acquisitions

Suppression of spatial aliasing artifacts resulted in an increased contrast of the carotid artery over multiple subjects. The CR box plot for all 31 transmission angles and all 9 subjects is shown in Figure 4.8 (a) and (b). Similar trends to the measurements in Figure 4.7 can be observed: downsampling of RF channels resulted in lower CR compared to the fully sampled case; inclusion of CNN-inferred RF data improved both the W-L CR and the T-L CR; and W-L CR had comparatively higher improvement compared to the T-L CR, which sees progressive losses as downsampling degree is increased.

The CR improvements achieved with the inclusion of CNN-inferred RF data were found to be statistically significant when evaluated on the 0° transmission angles. The box plots for all subjects' 0° transmission angles are shown in Figure 4.8 (c)-(d), where similar trends are present compared to the box plots that include all transmission angles. T-L and W-L CR improvements through the inclusion of CNN-inferred RF data were deemed significant ($p < \alpha_{adjusted} = 0.05/6 = 8.3\times10^{-3}$) for all degrees of downsampling. The Shapiro-Wilk test confirmed that all tested pairs conformed to the normality assumption used in the paired t-test ($p > \alpha = 0.05$). The detailed statistics for the paired 0° comparisons can be found in Table 4.2.

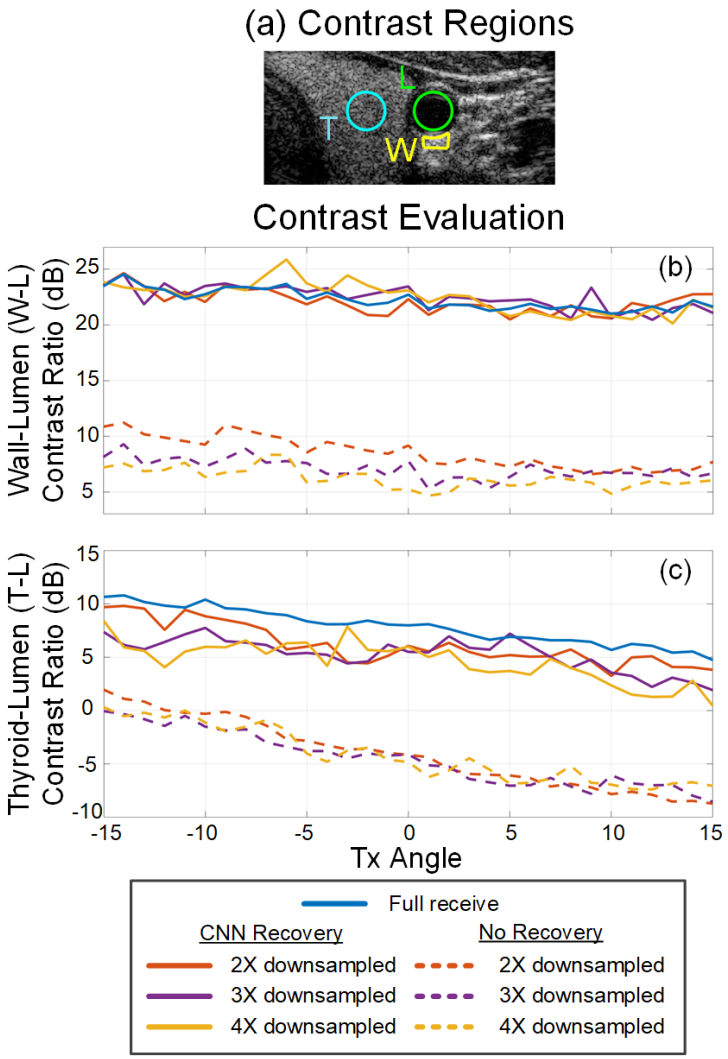


Figure 4.7. Contrast evaluation of a carotid lumen over multiple transmission angles. (a) Segmented regions for contrast assessment: Lumen (L), carotid wall (W), and thyroid (T). (b) W-L CR evaluations. (c) T-L CR evaluations.

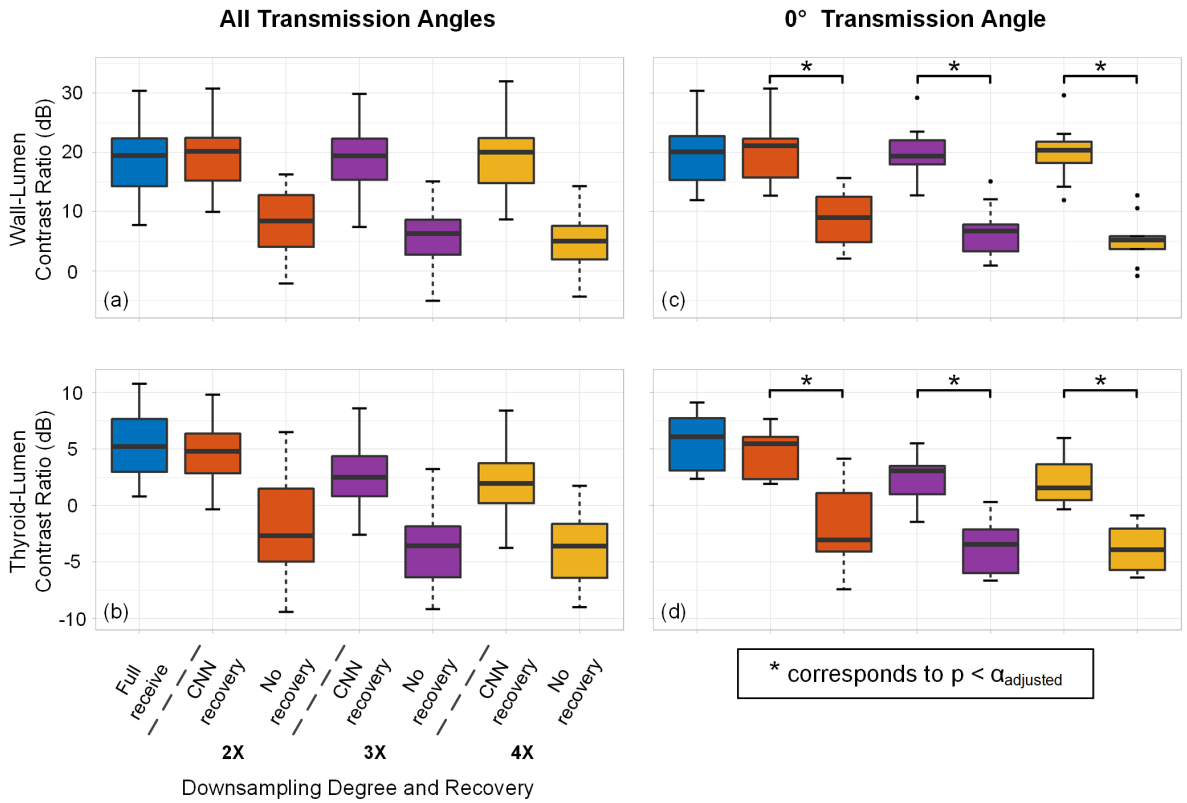


Figure 4.8. Contrast measurements over multiple subjects. (a)-(b) W-L CR and T-L CR box plots for all 31 transmission angles and 9 subjects (279 examples each group). (c)-(d) W-L and T-L CR box plots for 0° transmission angle and 9 subjects (9 examples each group).

# Table 4.2

**Statistical Testing of Tissue-Lumen Contrast Differences for 0° Transmissions**

| Tissue Point of Reference | Downsampling Level | P-value | Adjusted Confidence Interval (dB) | Mean Difference (dB) | Cohen's d effect size |
|---|---|---|---|---|---|
| Wall (W) | 2X | $4.12\times10^{-7}$ | [8.71, 14.0] | 11.4 | 2.25 |
| W | 3X | $2.90\times10^{-7}$ | [10.2, 16.0] | 13.1 | 2.72 |
| W | 4X | $8.70\times10^{-7}$ | [11.0, 18.6] | 14.8 | 3.12 |
| Thyroid (T) | 2X | $1.46\times10^{-3}$ | [1.7, 11.1] | 6.40 | 2.08 |
| T | 3X | $5.9\times10^{-4}$ | [2.2, 10.0] | 6.14 | 2.52 |
| T | 4X | $1.5\times10^{-4}$ | [3.0, 9.4] | 6.22 | 2.88 |

### 4.2.4 Coherent Compounding Analysis

The contrast and SSIM of images beamformed with CNN-inferred RF data were further improved with coherent plane wave compounding. The compounding process of the carotid artery in Fig. 4.6 (a) is shown in Fig. 4.9, where compounding resulted in a progressive improvement of contrast and SSIM for all beamforming scenarios. Similar to the single-angle cases, the inclusion of CNN-inferred data in the beamforming process enabled improved image quality compared to when only the subset of receiving channels were used. Firstly, the contrast between the lumen and carotid wall was enhanced beyond the case with all receiving channels when CNN-inferred RF data was used in beamforming. Secondly, the contrast between the thyroid and the lumen was consistently improved when CNN-inferred data was included during beamforming, despite the less hyperbolic RF data provided from thyroid reflections. Lastly, while the SSIM of all beamformed images improved at higher degrees of compounding, higher resultant SSIMs were achieved when CNN-inferred RF data was included during beamforming. The recovery attributed enhancement of compounded image quality can be observed in Fig. 4.10, where 7-angle-compounded images are displayed. While the compounding of images formed with downsampled RF data (Fig. 4.10 (e)-(g)) resulted in a higher quality image compared to their single transmit counterparts (Fig. 4.6 (e)-(g)), the images were still obstructed by spatial aliasing artifacts. The visibility of the carotid structure was enhanced when CNN-inferred data was also used in beamforming (Fig. 4.10 (b)-(d)).

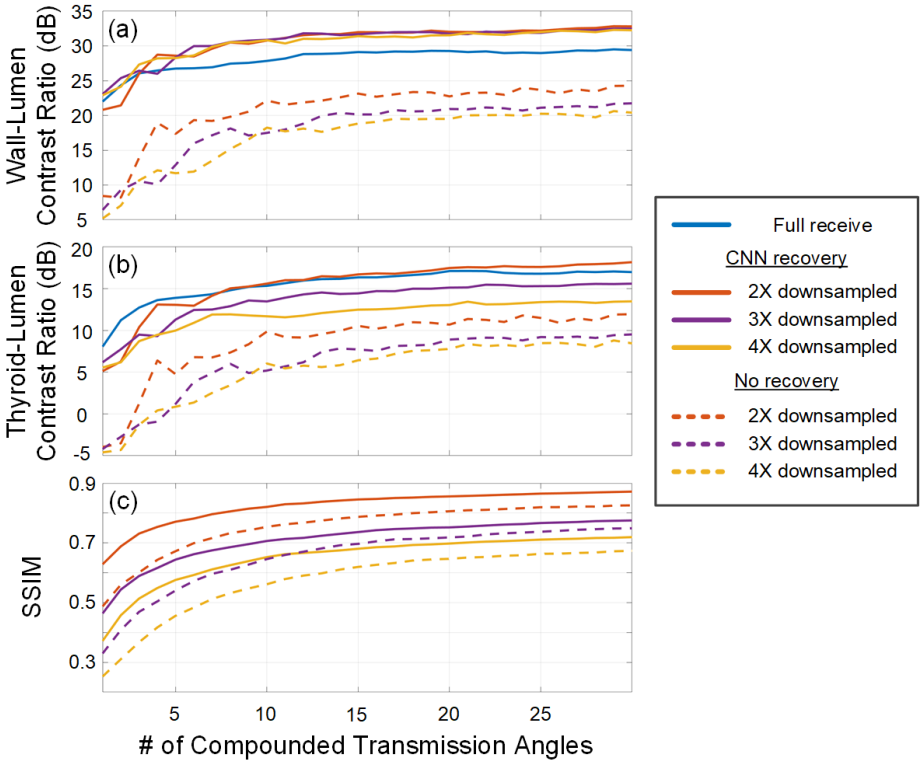


Figure 4.9. Progressive compounding analysis for the carotid artery in Figure 4.6. (a). (a) W-L CR, (b) T-L CR, and (c) SSIM metrics are tracked as transmission angles are compounded together.

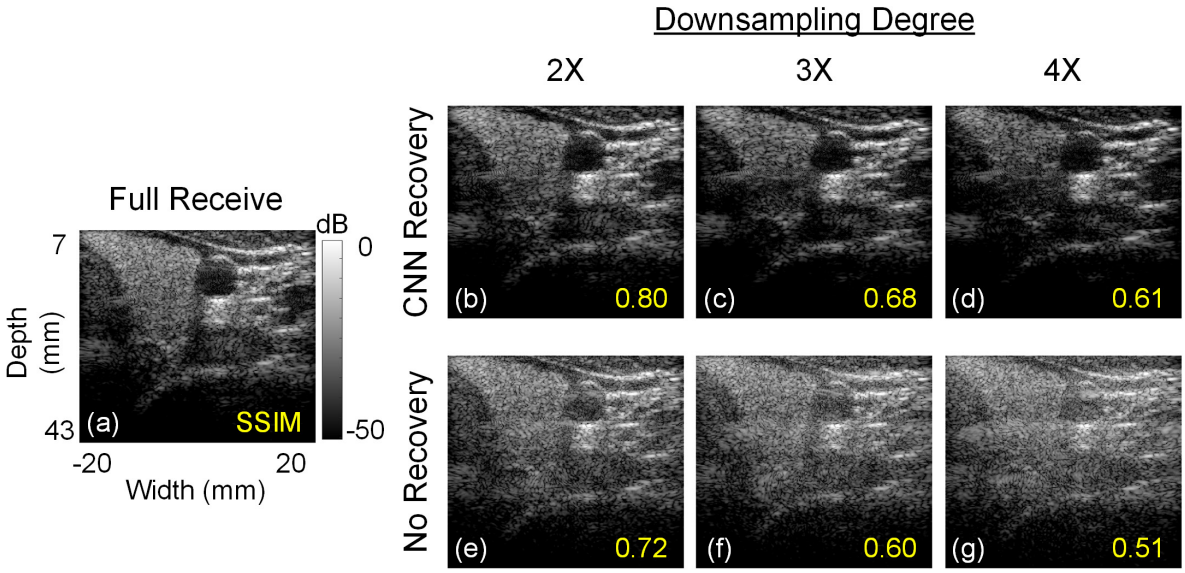


Figure 4.10. 7-angle compounded images for the carotid artery compounded in Figure 4.9. All images displayed with a dynamic range of 50dB. (a) Beamforming performed with all receiving channels. (b)-(d) beamforming performed with 1/2 received + 1/2 inferred, 1/3 received + 2/3 inferred, and 1/4 received + 3/4 inferred channels. (e)-(g) beamforming performed with 1/2, 1/3, and 1/4 received channels.

# Chapter 5
# Discussion and Future Directions

## 5.1 Branching Encoder-Decoder CNNs as a New Framework for Channel-Wise RF Recovery

Each receiving channel in an ultrasound system imposes a tangible increase in the system's complexity, whether it is through an increased data transfer bandwidth, or through the requirement of additional receiver electronics such as analog to digital converters or low noise amplifiers. To enable HiFRUS systems with lower channel counts and lower system complexity, we have developed an RF recovery framework (Figure 3.2) that can be applied to uniformly channel-wise downsampled RF data. The proposed framework leverages novel branching encoder-decoder CNN architectures (Figure 3.3) to directly recover RF channels that were omitted during the receive process, restoring a full set of received RF data from a uniformly downsampled subset. By restoring access to a full set of RF channels, the proposed recovery framework can be used to improve performance of signal processing methods that are given raw RF data as inputs such as DAS beamforming.

Our experiments with DAS beamforming show that including CNN-inferred channels in the beamforming process recovered the underlying structure of beamformed images (Figures 4.2, 4.3, 4.6, 4.7, 4.8). This improvement was generalizable to multiple imaging scenarios, namely an *in vitro* point target, *in vivo* carotid arteries, and an *in vivo* quadriceps muscle. Furthermore, inferred RF channels provided similar improvements when the network's inputs were from varying transmission angles. This angle-independence of the framework's inputs enables its use with more advanced RF-processing techniques; when inferred RF data from steered transmissions were used for coherent plane wave compounding, progressive image quality improvement was also achieved (Figure 4.9 and 4.10). Lastly, the raw RF recovery analysis that was performed (Figure 4.1 and 4.5) can be used to explain any image quality degradation that may be apparent in beamformed ultrasound images (further detailed in section 5.2).

Overall, the performed experiments indicate the proposed framework's feasibility for RF recovery, and they characterize the framework's performance recovering RF data from different types of tissue (hyperechogenic vs. hypoechogenic). This feasibility was demonstrated for recovery from downsampling degrees beyond 2X, surpassing the *in vivo* recovery rates demonstrated by other

channel-recovery techniques (Ramkumar & Thittai, 2020; Anand & Thittai, 2021; Kumar *et al.*, 2020; Xiao *et al.*, 2022). Given the utility of inferred RF data in the DAS beamforming process, the presented RF recovery framework could enable use of additional HiFRUS techniques when operating with a reduced receiver channel count. This work could thus aid the adoption of HiFRUS principles into more compact systems, extending their reach into more remote and austere healthcare environments.

## 5.2 Insights on CNN-Based RF Channel Inference

### 5.2.1 Successful Spatial Aliasing Artifact Reduction Stems from Successful Hyperbolic RF Inference

The image quality improvements that CNN-inferred RF data provided to DAS beamformed images can be attributed to a reduction in spatial aliasing artifacts. The inhibiting features present in the downsampled images of Figures 4.2 and 4.6 are the aliasing artifacts that hide the imaging region's underlying structure. The strong reduction of these aliasing artifacts is expected due to 2 reasons: 1) the most prevalent artifacts will occur due to insufficient spatial sampling of the most echogenic reflections in the imaging medium (as detailed in section 2.3.2.3); and 2) reflections from more echogenic scatterers manifest themselves in RF image hyperbolas, enabling higher accuracy in RF inference (shown in Figures 4.1 and 4.5). Therefore, the RF samples provided by the inference framework are expected to be highly effective at suppressing the most prominent spatial aliasing artifacts in a medium, revealing the medium's underlying structure. This is visible in the B-mode images of Figures 4.2 and 4.6, as the spatial aliasing artifacts are minimal in the images beamformed with recovered RF data. This suppression was achieved among multiple transmission angles, indicated by the reduction in amplitudes of spatially aliased regions in Figure 4.3 and the carotid contrast recovery in Figure 4.7. Furthermore, this trend of improved carotid contrast was achieved over multiple subjects, with a statistically significant improvement observed at all levels of downsampling when tested on 0° transmissions (shown in Figure 4.8).

While reflections from hyperechogenic structures achieved relatively higher inference success, if there was a lack of hyperbolic RF image structure it resulted in a more difficult inference scenario. When comparing the NRMSEs of inferred RF data for *in vitro* (Figure 4.1) and *in vivo* (Figure 4.5) scenarios, the reflections from more echogenic sources had substantially lower NRMSEs. This effect was reflected in the DAS beamformed images: carotid W-L contrast had minimal losses when CNN-inferred RF data was used at all levels of downsampling (Figure 4.7 and 4.8). This is compared to the

T-L contrast which was progressively decreased with the downsampling + recovery rate. This loss in contrast can be observed in Figure 4.6 (b)-(d), as some thyroid detail is lost with higher degrees of downsampling + recovery. A similar effect is shown in the *in vitro* contrast analysis of Figure 4.3, where the spatially aliased regions successfully have their artifacts suppressed, but there is a decrease in region amplitude when CNN-inferred RF data is included in beamforming. The amplitude decrease is a result of the RF error, as beamformed samples will have a less coherent sum, subduing the resultant amplitude. This effect also resulted in a significant impact on SSIM measures, as any inconsistencies in less reflective speckle amplitudes caused a large degradation of the SSIM, even if the overall image structure was captured. Despite this difficulty, the inclusion of inferred RF data improved the SSIM of all beamformed images compared to their strictly channel-wise downsampled counterparts, due to the successful inference of spatial aliasing artifacts.

### 5.2.2 Inference with Multiple Transmission Angles Allows Enhanced Image Quality via Coherent Plane Wave Compounding

DAS image improvements were observed over the full span of examined plane wave transmission angles. Suppression of spatial aliasing artifacts was achieved for *in vitro* point target acquisitions from -14.75° to 14.75° (Figure 4.3) and *in vivo* carotid artery acquisitions from -15° to 15° (Figure 4.7). This confirms the angle-independence of the RF recovery framework's inputs, allowing it to be used alongside advanced signal processing techniques such as coherent plane wave compounding.

The progressive image quality improvement that was achieved throughout the compounding process indicates a general coherence between CNN-inferred data from angled transmissions. The SSIM, T-L CR, and W-L CR were all improved when a carotid artery acquisition was compounded (Figure 4.9). The compounding process also improved the quality of images beamformed only with the subset of RF channels, but image quality metrics remained higher when CNN-inferred RF data was included. The difference provided by the inclusion of CNN-inferred RF data can be seen in Figure 4.10, where the images beamformed with CNN-inferred data more closely resemble the image beamformed with all receiving channels.

### 5.2.3 Potential Ceiling for Channel Recovery Rate

Due to the different degrees of RF recovery success for different types of tissues, there is not a well-defined limit to the amount of CNN-based recovery that can be achieved, as it will depend on the

medium being imaged and the application at hand. Any errors present in inferred RF data become more prevalent as the degree of recovery is increased, imposing tradeoffs to the amount of acceptable recovery in a system. In this thesis, these tradeoffs were observed during the process of DAS beamforming; while the inclusion of CNN inferred data enabled recovery of underlying beamformed image structure due to strong suppression of spatial aliasing artifacts, point target resolution was slightly degraded with higher degrees of recovery (Figure 4.4) and some lower amplitude tissue data was suppressed at higher degrees of recovery (shown *in vitro* in Figure 4.3, and *in vivo* in Figures 4.7 and 4.8). The impact of these tradeoffs will differ depending on the application of the recovered RF data. For example, if compounding is to be performed to acquire higher quality beamformed images, the tradeoffs associated with RF recovery are partially mitigated (Figures 4.9 and 4.10). Additionally, if the RF data is to be used for a specialized beamforming method such as minimum-variance beamforming (Synnevåg *et al.*, 2009), these tradeoffs may be counteracted by the image quality improvements imposed by these methods.

## 5.3 Advantages and Limitations of Proposed Method

### 5.3.1 Implementing a Deep Learning Based Framework in a Compact System

By using a CNN-based solution for RF recovery, the proposed framework requires hardware that can facilitate a forward pass of a CNN. It is important that such a hardware solution would have a small form factor, and that it can be implemented in the back end of a system (after data transfer). GPUs are a potential candidate for implementation of the RF recovery framework. Convolutional operations are easily parallelized using GPUs, and recent innovations have seen significant downsizing of this technology in products such as the NVIDIA Jetson platform (Mittal, 2019). This coincides with the GPU's ability to parallelize several other tasks in an ultrasound processing pipeline (beamforming being a pertinent example), positioning GPUs as a strong candidate to be included in modern ultrasound scanners (So *et al.*, 2011). While GPUs are an attractive means of implementation, they are not the only solution, as recent innovations in portable deep learning hardware architectures (Zaman *et al.*, 2021) could be leveraged to implement a custom on-chip RF inference solution.

### 5.3.2 Performing HiFRUS Tracking in Different Mediums

A system that implements the proposed recovery framework should be well-suited to track dynamic events in the human body using HiFRUS. Given the higher accuracy when inferring reflections from

more echogenic scatterers, it is expected that the recovery framework would be of particular use when tracking movement of hyperechogenic tissues such as arterial walls (Couade *et al.,* 2011) or muscle fibers (Cortes *et al.*, 2015). While scenarios that track the movement of hypoechogenic scatterers may be more challenging, insights presented in this work could guide framework/acquisition modifications that improve inference success in these scenarios. For example, in flow imaging, contrast enhancement could be implemented to increase the echogenicity of blood reflections (Harvey *et al.*, 2001), with higher echogenicity expected to improve inference success.

## 5.4 Future Directions

### 5.4.1 Extension to Additional Imaging Schemes

With feasibility of the RF recovery framework demonstrated within the context of plane wave acquisitions on 1D linear arrays, additional research should be pursued for extending the framework to additional imaging schemes. The appearance of hyperbolic structure in RF images is not exclusive to plane wave acquisitions, therefore, it is expected that similar results would be seen when applying the proposed framework to other imaging schemes such as synthetic aperture imaging (Jensen *et al.*, 2006). Additionally, the demonstrated feasibility of >2X RF recovery raises the question of whether the framework could be extended towards 2D matrix arrays, since 2D sparse arrays typically require a larger reduction in channel count to adequately reduce system complexity (Roux *et al.*, 2018; Mattesini *et al.,* 2020; Yu *et al.*, 2020). This extension could be through a row/column-wise application of the framework to a downsampled matrix array, or through use of 3D convolutional kernels (Singh *et al.*, 2019) to infer reflections from 3D hyperbolic structure in RF matrices.

### 5.4.2 Feeding Recovered RF Channels into Advanced Signal Processing Algorithms

Given the framework's ability to directly recover missing RF channels, a logical progression of this work would be to provide inferred RF data as input to advanced signal processing algorithms that take raw RF inputs. As mentioned in section 2.4, some of these algorithms include methods of speed of sound mapping (Feigin *et al.*, 2020), image segmentation (Nair *et al.*, 2020), and beamforming (Synnevåg *et al.*, 2009; Matrone *et al.*, 2015; Cheng & Lu, 2006; Garcia *et al.,* 2013). The insights presented in this work can be used to guide investigation related to these other algorithms.

### 5.4.3 Preparation for Inference in a Clinical Setting

With initial feasibility demonstrated for the RF recovery framework, additional steps should subsequently be taken to strengthen the framework's suitability for clinical adoption. Implementation of a deep learning framework in a clinical imaging setting requires more robust preparation, since any errors made by the recovery framework can have negative impact on patient outcomes via misdiagnosis.

The framework needs to be proven to be generalizable, so that it can be applied to different imaging scenarios, on patients with differing backgrounds. Given that the framework was trained exclusively on *in vivo* carotid scans, the suppression of spatial aliasing artifacts on an *in vitro* point target and an *in vivo* quadriceps muscle provide evidence for the framework's applicability to different imaging regions. However, there are several additional regions of the body that are imaged with ultrasound, for example breast imaging (Szabo & Lewin, 2013). Prior to a clinical adoption of this framework, the training dataset should be expanded to cover a wider range of *in vivo* imaging scenarios. Additionally, data from a more diverse set of patients should be acquired, to prevent prediction bias for one particular type of patient (Esteva *et al.,* 2021). Acquiring a larger and wider range of training data should improve the generalizability of the framework and reduce the potential for inequalities in the framework's usage.

To meet the higher standard of trust that is required to implement a framework in a clinical setting, the machine learning model should have a degree of interpretability (Rueckert *et al.,* 2020; McCrindle *et al.*, 2021). This can pose a challenge in a deep learning system, as the complexity of a deep learning model often results in its treatment as a "black box" where its inner workings are unknown. However, there have been efforts recently to increase interpretability of these models. For example, model activity can be visualized by displaying input patterns that cause activations in intermediate layers (Zeiler & Fergus, 2014), and attribution maps can highlight regions of an input that are most relevant when producing a given output (Salahuddin *et al.*, 2021). Additional effort should be made to increase the interpretability of the RF recovery framework, to build the level of trust in its operation that is required for its use in a clinical imaging setting.

# References


Albawi, S., Mohammed, T. A., Al-Zawi, S. (2017). Understanding of a convolutional neural network. *In 2017 International Conference on Engineering and Technology (ICET),* 1-6.

Anand, R., Thittai, A. K. (2021). Towards practical implementation of the compressed sensing framework for multi-element synthetic transmit aperture imaging. *Ultrasonics,* 112, 106354.

Araujo, A., Norris, W., Sim, J. (2019). Computing Receptive Fields of Convolutional Neural Networks. *Distill,* 4(11), e21.

Armstrong, R. A. (2014). When to use the Bonferroni correction. *Ophthalmic & Physiological Optics,* 34, 502-508.

Baran, J. M., Webster, J. G. (2009). Design of low-cost portable ultrasound systems: Review. *In 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society,* 792-795.

Bercoff, J., Montaldo, G., Loupas, T., Savery, D., Mézière, F., Fink, M., Tanter, M. (2011). Ultrafast Compound Doppler Imaging: Providing Full Blood Flow Characterization. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control,* 58(1), 134-147.

Besson, A., Carrillo, R. E., Bernard, O., Wiaux, Y., Thiran, J. P. Compressed delay-and-sum beamforming for ultrafast ultrasound imaging. *In 2016 IEEE International Conference on Image Processing (ICIP)*, 2509-2513.

Boni, E., Yu, A. C. H., Freear, S., Jensen, J. A., & Tortoli, P. (2018). Ultrasound Open Platforms for Next-Generation Imaging Technique Development. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control,* 65(7), 1078-1092.

Bouvrie, J. (2006). Notes on Convolutional Neural Networks. MIT.

Carpenter, T. M., Rashid, W., Ghovanloo, M., Cowell, D. M. J., Freear, S., Degertekin, F. L. (2016). Direct Digital Demultiplexing of Analog TDM Signals for Cable Reduction in Ultrasound Imaging Catheters. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control,* 63(8), 1078-1085.

Cheng, J., Lu, J. (2006). Extended High-Frame Rate Imaging Method with Limited-Diffraction Beams. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control,* 53(5), 880-899.

Cikes, M., Tong, L., Sutherland, G. R., D'hooge, J. (2014). Ultrafast Cardiac Ultrasound Imaging: Technical Principles, Applications, and Clinical Benefits. *JACC: Cardiovascular Imaging,* 7(8), 812-823.

Cohen, R., Eldar, Y. C. (2018). Sparse Convolutional Beamforming for Ultrasound Imaging. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control,* 65(12), 2390-2406.

Couade, M., Pernot, M., Messas, E., Emmerich, J., Hagège, A., Fink, M., Tanter, M. (2011). Ultrafast imaging of the arterial pulse wave. *IRBM*, 32(2), 106-108

Cortes, D. H., Suydam, S. M., Silbernagel, K. G., Buchanan, T. S., Elliott, D. M. (2015). Continuous Shear Wave Elastography: A New Method to Measure Viscoelastic Properties of Tendons *In Vivo*, *Ultrasound in Medicine & Biology,* 41(6), 1518-1529.

Dargan, S., Kumar, M., Ayyagari, M. R., Kumar, G. (2019). A Survey of Deep Learning and Its Applications: A New Paradigm to Machine Learning. *Archives of Computational Methods in Engineering*, 27(4), 1071-1092.

David, G., Robert, J. L., Zhang, B. (2015). Time domain compressive beam forming of ultrasound signals. *The Journal of the Acoustical Society of America,* 137(5), 2773-2784.

Diarra, B., Robini, M., Tortoli, P., Cachard, C., Liebgott, H. (2013). Design of Optimal 2-D Nongrid Sparse Arrays for Medical Ultrasound. *IEEE Transactions on Biomedical Engineering,* 60(11), 3093-3102.

Donoho, D. L. (2006). Compressed sensing. *IEEE Transactions on Information Theory,* 52(4), 1289-1306.

Dumoulin, V., Visin, F. (2018). A guide to convolution arithmetic for deep learning. *arXiv:1603.07285.*

Estava, A., Chou, K., Yeung, S., Naik, N., Madani, A., Mottaghi, A., Liu, T., Topol, E., Dean, J., Socher, R. (2021). Deep learning-enabled medical computer vision. *NPJ digital medicine*, 4(1), 1-9.

Elharrouss, O., Almaadeed, N., Al-Maadeed, S., & Akbari, Y. (2020). Image Inpainting: A Review. *Neural Processing Letters,* 51(2), 2007-2028.

Feigin, M., Freedman, D., Anthony, B. W. (2020) A Deep Learning Framework for Single-Sided Sound Speed Inversion in Medical Ultrasound. *IEEE Transactions on Biomedical Engineering,* 67(4), 1142-1151.

Garcia, D., Tarnec, L. L., Muth, S., Montagnon, E., Porée, J., Cloutier, G. (2013). Stolt's f-k Migration for Plane Wave Ultrasound Imaging. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control,* 60(9), 1853-1867.

Goodfellow, I., Bengio, Y., Courville, A. (2016). Convolutional Networks. In *Deep Learning*, MIT Press, 326-366.

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., Chen, T. (2018). Recent advances in convolutional neural networks. *Pattern Recognition,* 77, 354-377.

Harvey, C. J., Blomley, M. J. K., Eckersley, R. J., Cosgrove, D. O. (2001). Developments in ultrasound contrast media. *European Radiology*, 11(4), 675-689.

He, K., Zhang, X., Ren, S., Sun, J. (2015). Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. *In Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 1026-1034.

Hecht-Nielsen, R. (1989). Theory of the backpropagation neural network. *Neural networks for perception.* 65-93.

Hujiben, I. A. M., Veeling, B. S., Janse, K., Mischi, M., van Sloun, R. J. G. (2020). Learning Sub-Sampling and Signal Recovery with Applications in Ultrasound Imaging. *IEEE Transactions on Medical Imaging,* 39(12), 3955-3966.

Humphrey, V. F. (2007). Ultrasound and matter – Physical interactions. *Progress in Biophysics and Molecular Biology.* 93(1-3), 197-211.

Jensen, J. A. (1996). Field: A Program for Simulating Ultrasound Systems. *In 10$^{th}$ Nordic Baltic Conference on Biomedical Imaging*. 351-353.

Jensen, J. A., Nikolov, S. I., Gammelmark, K. L., Pedersen, M. H. (2006). Synthetic aperture ultrasound imaging. *Ultrasonics,* 44(22), e5-e15.

Jensen, J. A., Svendsen, N. B. (1992). Calculation of Pressure Fields from Arbitrarily Shaped, Apodized, and Excited Ultrasound Transducers. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control,* 39(2), 262-267.

Kingma, D. P., Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv:1412.6980*.

Kumar, V., Lee, P. Y., Kim, B. H., Fatemi, M., Alizad, A. (2020). Gap-Filling Method for Suppressing Grating Lobes in Ultrasound Imaging: Experimental Study with Deep-Learning Approach. *IEEE Access,* 8, 76276-76286.

Lanza, G. (2020). Ultrasound Imaging, Something Old or Something New? *Investigative Radiology,* 55(9), 573-577.

Larson, J. D. III. (1993). 2-d phased array ultrasound imaging system with distributed phasing. *U.S. Patent Application No. 5,229,933.*

Lee, D. K. (2016). Alternatives to P value: confidence interval and effect size. *Korean Journal of Anesthesiology*, 69(6), 555-562.

Liu, H., Han, J., Hou, S., Shao, L., Ruan, Y. (2018). Single image super-resolution using a deep encoder-decoder symmetrical network with iterative back projection. *Neurocomputing,* 282, 52-59.

Lockwood, G. R., Li, P. C., O'Donnell, Foster, F. S. (1996). Optimizing the radiation pattern of sparse periodic linear arrays. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 43(1), 7-14.

Maas, A. L., Hannun, A. Y., Ng, A. Y. (2013). Rectifier Nonlinearities Improve Neural Network Acoustic Models. *In Proceedings of the 30$^{th}$ International Conference on Machine Learning,* 3.

Mao, X., Shen, C., Yang, Y. B. (2016). Image Restoration Using Very Deep Convolutional Encoder-Decoder Networks with Symmetric Skip Connections. *Advances in neural information processing systems*, 29.

Matrone, G., Savoia, A. S., Caliano, G., Magenes, G. (2015). The Delay Multiply and Sum Beamforming Algorithm in Ultrasound B-Mode Medical Imaging. *IEEE Transactions on Medical Imaging,* 32(4), 940-949.

Mattesini, P., Ramalli, A., Petrusca, L., Basset, O., Liebgott, H., Tortoli, P. (2020). Spectral Doppler Measurements With 2-D Sparse Arrays. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control,* 67(2), 278-285.

McCrindle, B., Zukotynski, K., Doyle, T. M., Noseworthy, M. D. (2021). A Radiology-focused Review of Predictive Uncertainty for AI Interpretability in Computer-assisted Segmentation. *Radiology: Artificial Intelligence*, 3(6), e210031.

Mittal, S. (2019). A Survery on optimized implementation of deep learning models on the NVIDIA Jetson Platform. *Journal of Systems Architecture,* 97, 428-442.

Montaldo, G., Tanter, M., Bercoff, J., Benech, N., Fink, M. (2009). Coherent Plane-Wave Compounding for Very High Frame Rate Ultrasonography and Transient Elastography. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control,* 56(3), 489-506.

Nair, A. A., Washington, K. N., Tran, T. D., Reiter, A., Bell, Muyinatu A. L. (2020). Deep Learning to Obtain Simultaneous Image and Segmentation Outputs From a Single Input of Raw Ultrasound Channel Data. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control,* 67(12), 2493-2509.

Nelson, B. P., Sanghvi, A. (2016). Out of hospital point of care ultrasound: current use models and future directions. *European Journal of Trauma and Emergency Surgery,* 42(2), 139-150.

Ohri, K., Kumar, M. (2021). Review on self-supervised image recognition using deep neural networks. *Knowledge-Based Systems,* 224, 107090.

Perdios, D., Vonlanthen, M., Martinez, F., Arditi, M., Thiran, J. P. (2020). Single-Shot CNN-Based Ultrasound Imaging with Sparse Linear Arrays. *In 2020 IEEE International Ultrasonics Symposium (IUS),* 1-4.

Powles, A. E. J., Martin, D. J., Wells, I. T. P., Goodwin, C. R. (2011). Physics of ultrasound. *Anaesthesia & Intensive Care Medicine,* 19(4), 202-205.

Perrot, V., Polichetti, M., Varray, F., Garcia, D. (2021). So you think you can DAS? A viewpoint on delay-and-sum beamforming. *Ultrasonics*, 111, 106309.

Ramkumar, A., Thittai, A. K. (2020). Compressed Sensing Approach for Reducing the Number of Receive Elements in Synthetic Transmit Aperture Imaging. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control,* 67(10), 2012-2021.

Ronneberger, O., Fischer, P., Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *In International Conference on Medical image computing and computer-assisted intervention*, 234-241.

Roux, E., Varray, F., Petrusca, L., Cachard, C., Tortoli, P., Liebgott, H. (2018). Experimental 3-D Ultrasound Imaging with 2-D Sparse Arrays using Focused and Diverging Waves. *Scientific Reports*, 8(1), 1-12.

Rueckert, D., Schnabel, J. A. (2020). Model-Based and Data-Driven Strategies in Medical Image Computing. *Proceedings of the IEEE,* 108(1), 110-124.

Salahuddin, Z., Woodruff, H. C., Chatterjee, A., Lambin, P. (2021). Transparency of deep neural networks for medical image analysis: A review of interpretability methods. *Computers in Biology and Medicine*, 140, 105111.

Shaprio, S. S., Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4), 591-611.

Simonyan, K., Zisserman, A. (2015) Very Deep Convolutional Networks for Large-Scale Image Recognition. *In ICLR 2015*.

Singh, R. D., Mittal, A., Bhatia, R. K. (2019). 3D convolutional neural network for object recognition: a review. *Multimedia Tools and Applications,* 78(12), 15951-15995.

Sippel, S., Muruganandan, K., Levine, A., Shah, S. (2011). Review article: Use of ultrasound in the developing world. *International Journal of Emergency Medicine*, 4(1), 1-11.

So, H., Chen, J., Yiu, B. Y. S., Yu, A. C. H. (2011). Medical Ultrasound Imaging: To GPU or Not to GPU? *IEEE Micro,* 31(5), 54-65.

Springenberg, J. T., Dosovitskiy, A., Brox, T., Riedmiller, M. (2015). Striving for Simplicity: The All Convolutional Net. *In The International Conference on Learning Representations (ICLR)*.

Student. (1908). The Probable Error of a Mean. *Biometrika*. 6(1), 1-25.

Synnevåg, J. F., Austeng, A., Holm, S. (2009). Benefits of Minimum-Variance Beamforming in Medical Ultrasound Imaging. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control,* 56(9), 1868-1979.

Szabo, T. L., Lewin, P. A. (2013). Ultrasound Transducer Selection in Clinical Imaging Practice. *Journal of Ultrasound in Medicine,* 32(4), 573-582.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A. (2015). Going Deeper With Convolutions. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1-9.

Tanter, M., Fink, M. (2014). Ultrafast imaging in biomedical ultrasound. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 61(1), 102-119.

Van Sloun, R. J. G., Cohen, R., Eldar, Y. C. (2020). Deep Learning in Ultrasound Imaging. *Proceedings of the IEEE,* 108(1), 11-29.

Wang, S., Cao, J., Yu, P. S. (2020). Deep Learning for Spatio-Temporal Data Mining: A Survey. *IEEE Transactions on Knowledge and Data Engineering*, early access. doi: 10.1109/TKDE.2020.3025580

Wang, Z., Bovik, A. C., Sheikh, H. R., Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing,* 13(4), 600-612.

Xiao, D., Pitman, W. M. K., Yiu, B. Y. S., Chee, A. J. Y., Yu, A. C. H. (2022). Minimizing Image Quality Loss after Channel Count Reduction for Plane Wave Ultrasound via Deep Learning Inference. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control,* early access. doi: 10.1109/TUFFC.2022.3192854

Yiu, B. Y. S., Yu, A. C. H. (2016). Least-Squares Multi-Angle Doppler Estimators for Plane-Wave Vector Flow Imaging. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control,* 63(11), 1733-1744.

Yu, J., Yoon, H., Khalifa, Y. M., Emelianov, S. Y., Design of a Volumetric Imaging Sequence Using a Vantage-256 Ultrasound Research Platform Multiplexed With a 1024-Element Fully Sampled

Matrix Array. (2019). *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control,* 67(2), 248-257.

Zaman, K. S., Reaz, M. B. I., Ali, S. H. M., Bakar, A. A. A., Chowdhury, M. E. H. (2021). Custom Hardware Architectures for Deep Learning on Portable Devices: A Review. *IEEE Transactions on Neural Networks and Learning Systems,* Early Access, 1-21.

Zeiler, M. D., Fergus, R. (2014). Visualizing and Understanding Convolutional Networks. In *European conference on computer vision*, 818-833.

# Appendix: Hand Segmented Regions of Interest for Contrast Evaluation

This appendix contains the hand-segmented regions of interest for carotid contrast evaluation. The ROI selection is shown below in Figure A.1., where the lumen, wall, and thyroid regions were selected for 9 volunteers.

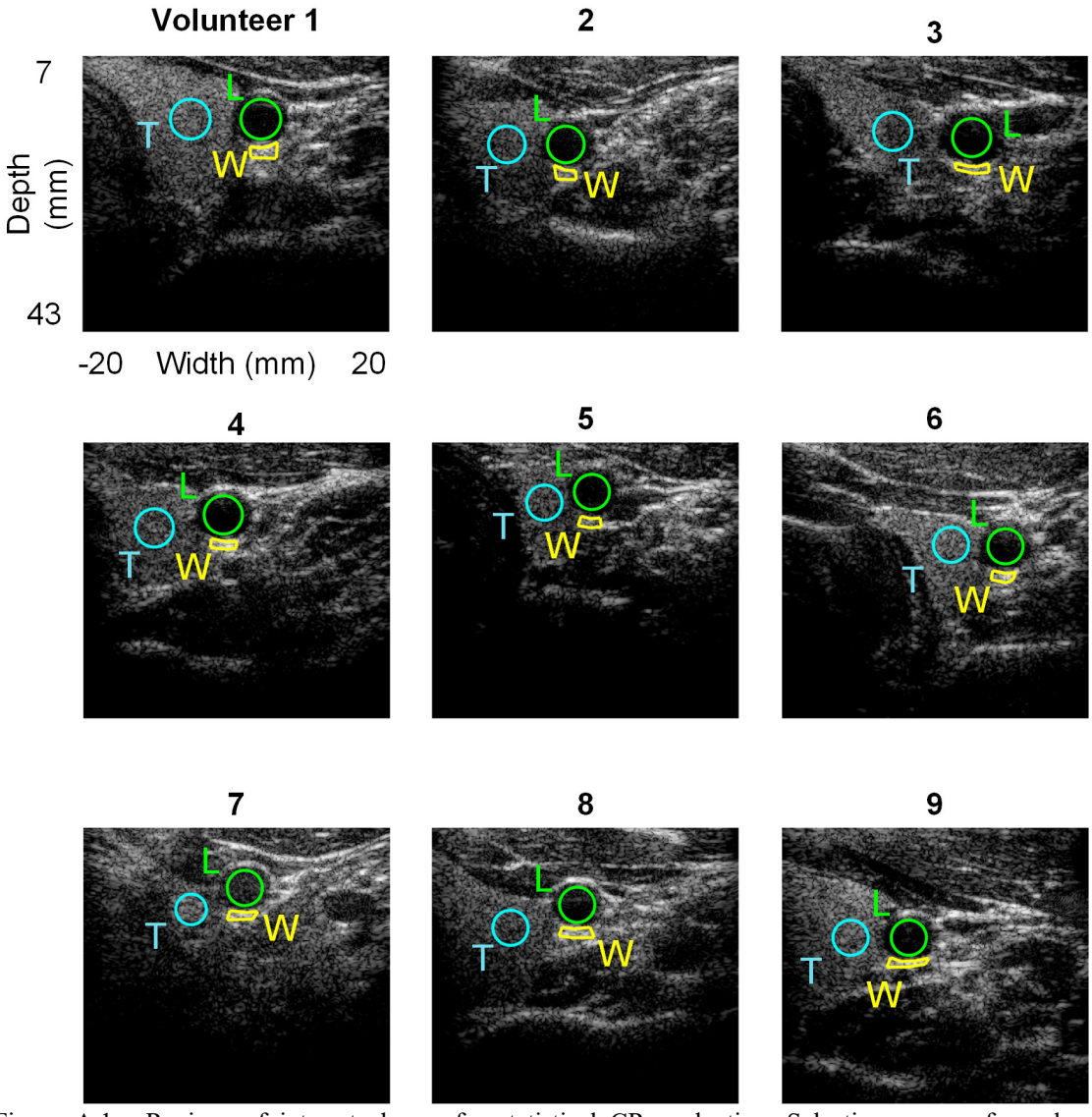


Figure A.1. Regions of interest chosen for statistical CR evaluation. Selection was performed on fully compounded images. All images are displayed with a dynamic range of 50dB.