Noise Suppression in Ultrasound Beamforming Using Convolutional Neural Networks

By

Zhanwen Chen

A thesis submitted in partial satisfaction of the
requirements for the degree of
Master of Science

in

Computer Science

in the

Graduate Division

of the

Vanderbilt University

Committee in charge:

Professor Matthew Berger, Chair
Professor Maithilee Kunda
Associate Professor Brett Byram

Fall 2019

The thesis of Zhanwen Chen, titled Noise Suppression in Ultrasound Beamforming Using Convolutional Neural Networks, is approved:

Chair ____________________ Date __________

____________________ Date __________

____________________ Date __________

Vanderbilt University

Noise Suppression in Ultrasound Beamforming Using Convolutional Neural Networks

by

Zhanwen Chen

Abstract

Noise Suppression in Ultrasound Beamforming Using Convolutional Neural Networks

by

Zhanwen Chen

Master of Science in Computer Science

Vanderbilt University

Professor Matthew Berger, Chair

To my parents, Cheng Yuehong and Chen Feng. Thank you for years of unconditional support and encouragement.

Contents

List of Figures

List of Tables

Acknowledgments

Chapter 1

Introduction

1.1 Introduction to Ultrasound Beamforming

Diagnostic medical ultrasound has its roots in Sonar and ultrasonic metal flaw detectors. It is a noninvasive, affordable, portable, and real-time method to characterize the cross-sectional view of soft tissues compared with other imaging modalities such as Computed Tomography (CT) and magnetic resonance imaging (MRI). The underlying principle of ultrasound is the measurement of time elapsed between sending a signal and receiving its echo; given the sound speed *a priori*, we can thus calculate the distance to an object based on this duration.

Ultrasound imaging consists of three steps: emitting sound waves (transmit), receiving echoes (receive), and interpreting those responses to form an image. The transmit step is achieved with ultrasonic transducers - devices that convert electricity into ultrasound waves or vice versa. In ultrasound imaging, we use transceivers that both emit and receive echoes.

In practice, ultrasound scans are acquired with array transducers - a group of individual small transducers, each of whose pulse transmission is preciously timed by a computer. Because a single transducer element is only able to capture a single depth-dimension signal, a (horizontal) stack of such transducers can introduce the lateral (space) dimension.

The most basic case of ultrasound imaging is plane wave imaging, where all transducers in an array emit the same acoustic pulses at the same time. This creates a flat wavefront, which describes the shape of a group of waves by connecting their centers. After the transmit event, the waves propagate the field of view and are scattered back to the transducers as pulse-echo responses.

Each transducer element samples its measured acoustic pressure (in dB) at each timestep within a small time window. The distance to an object is then calculated as

$$distance = \frac{sound\,speed * timestep}{2}$$

, accounting for both directions of travel. With this relationship, the time dimension can be translated to the distance or the depth dimension. Thus, each transducer has a series of acoustic signals v for each depth, y. With a horizontal or lateral series of x transducers, we have a 2D matrix with each value $v = V[x, y]$ representing the sound signal/acoustic energy measured by the xth transducer at depth y. If we scale this image to grayscale, we have an image of width x and height y, where each pixel represents the relative intensity in the dynamic range of measured sound energy. A typical dynamic range is $[0, -60]dB$. This image modality is called brightness mode, or B-mode. The lateral dimension may be distinct from the number of transducer elements in cases of horizontal interpolatation to increase lateral resolution.

If the scanned object has no variation in sound impedance (for example, water, air), the flat wavefront sent out will return at the same timesteps, because all waves encounter the same level of resistence. Scanning pure water or air with plane waves produces a blank image because the intensity is the same (at all miniscule timesteps, each transducer receives the same amount of acoustic pressure). It is notable, however, that waves do not propagate straight, and even in the absence of scattering media, waves still disperse. In other words, the wave emitted by one transducer can be partially received by another.

Now consider the nontrivial example of plane wave imaging of a phantom, which is an artificial composite of materials of various shapes and sound impedance. As was the case previously, all elements emit the same pulse at the same timesteps. However, as each pulse wave travels through the composite, it encounters divergent impedance, and some of the wave energy gets scattered at various points in depth and at various degrees (refractions), depending on the location and the impedance of the component matrials. This difference in intensity generates signal in the resulting image, which should characterize the cross-sectional view of the phantom.

A fundamental limitation to this basic method of plane wave imaging is the lack of focus. The images are blurry because the received signals are not strong enough relative to noise. We want to focus our transducers because the responses from adjacent elements are more relevant than those far away. In practice, focus in ultrasound requires a subgroup of the total transducer array to form a single 1D depth-signal series, as opposed to one series for each element. A set of time-delayed (focused) transducer waves are called a beam. Focused imaging maximizes signal, minimizes noise, and results in a higher signal-to-noise ratio. Images generated this way are more helpful to clinical diagnosis.

To achieve focusing, we first select a subset of transducers (a *channel* or *aperture*) and slide the selection by one element for *num_total_elements* times, where *num_total_elements* is the number of total elements in the overall array. Only elements that transmit are set to receive. In other words, each aperture both transmit and receive before we slide the window. This results in a new channel/aperture dimension to our data, in addition to the depth and lateral ones. We call this new type of data matrix *channel data*.

Within each subset, we need to send out a focused wave of a curved wavefront by taking advantage of wave interference. The superimposition of waves can cause constructive or destructive interferences, depending on their relative phases and amplitudes upon contact. We preset a focus (typically controlled by a knob on an ultrasound machine), from which we then derive a desired wavefront. Working backwards from wave interference equations, we can determine how much time delay is needed for each transducer in each subarray. All subarrays use the same delay pattern. The ultrasound machine typically applies the delays automatically.

The processing of channel data in order to form an image is called beamforming. The most basic method of beamforming is delay and sum (DAS). After receive, we undo the added time delays so that we receive a flat wavefront to reflect the true depth. After applying delays, we finally operate on the channel dimension. The dimension of our post-delayed channel data is $[depth, elements\_per\_channel, num\_total\_elements]$. To form each beam (vertical slice in the final image), we collapse its channel dimension by summing all 1D transducers responses in its aperture group, resulting in a new data matrix of size $[depth, num\_total\_elements]$, called radio frequency (RF) data. RF data is the basis for a beamformed image, which then undergoes filtering and upsampling, depending on the specific application requirements.

1.2 Challenges in Ultrasound Beamforming

Although widely accepted, DAS beamforming is not an ideal method for clinical application due to the presence of many noises or artifacts, of which we present three: off-axis scattering and reverberation.

Underlying these two artifacts is the basic mechanism of ultrasound - scattering. Scattering describes the reflection of an acoustic wave as it encounters the boundary between matters of differing impedance. There are many scatters (boundaries) in the field of view. We assume that a wave scatters once before returning to the transducer array and that a transducer subarray emits a straight wave aimed linearly down (the area directly under is called a region of interest - ROI). However, these assumptions are not true in reality. For example, speckles are formed by interference among scattered waves which stray from the ROI. Even though speckles are not considered an artifact, there are other properties of scattering that are.

1.2.1 Off-Axis Scattering

The first such artifact is off-axis scattering. We typically assume that pulse-waves propagate downward, but in reality, pulse waves exhibit diffraction when emitted by a transducer beam. The diffraction resembles the way sound travels. Using an analogy, when person A shouts a secret message facing person B, a bystander C can often hear the message as well (if less clearly). In ultrasound, this phenomenon can be illustrated by beam plots - the normalized far-field magnitude of the transmit pressure versus observation angle. The acoustic pressure we want to focus on is the mainlobe, and the sidelobes are the diffractions that dilute the energy and cause off-axis echos that degrade the image.

1.2.2 Reverberation

Another cause of image degradation is reverberation or multipath scattering. We assume that when a wave encounters a boundary, it is reflected back to and only to the transducers. However, in reality, the scattered wave can travel in all directions. In addition, the divergent scattered waves can be further scattered by boundaries outside the ROI of the emitting beam. As a result of bouncing around in the field of view, the wave takes longer than expected to return to the same transducer. The multipath-scattered responses interference with the desired vertical responses from the ROI and degrades the resulting image.

1.3 Noise Suppression Algorithms

Many algorithms have been developed to address ultrasound artifacts. Earlier methods include Tissue Harmonic Imaging, Time-Reversal Technique, Second-Order Ultrasound Field, Minimum-Variance Beamforming, PCA, SVD, and Short-Lag Spatial Coherence. These methods typically have a tradoff of a smaller field of view. Recently, statistical approaches have been proposed. One example is Aperture Domain Model Image REconstruction (ADMIRE) developed by Byram and Dei [1]. Machine learning approaches tend to work well, but they are too computationally demanding to be real-time, although progress has been made to improve its performance.

In addition, studies have shown that deep neural networks are effective in suppressing noises. Of particular relevance is the application of multilayer perceptrons (MLPs) in the aperture domain or channel dimension [4]. Lastly, convolutional neural networks (CNNs) have also been used in biomedical imaging in general [5] and various ultrasound imaging modalities [6]. In terms of ultrasound beamforming, work has been done using CNNs to learn the entire beamforming process [2].

1.4 Contribution

The aim of this thesis is to investigate the effectiveness of convolutional neural networks (CNNs) in suppressing off-axis scattering and reverberation clutter. Particularly, I apply CNNs to signals in the aperture short-time Fourier transform (STFT) domain to avoid having to train for for different pulse shapes, depth dependent attenuation, and other pulse parameters that may vary across patients and even across probes as they age. I will discuss the rationale in detail in Chapter 2.

1.5 Organization

Chapter 2 discusses background on deep neural networks and related work on noise suppression with deep learning and CNN-based beamforming. Chapter 3 describes the training data generation process and signal grouping. Chapter 4 explains the CNN architectures and the training pipeline, including a random hyperparameter search technique. Chapter 5 details the beamforming pipeline for evaluation. Chapter 6 addresses the limitations of this work, discusses the results, and concludes the thesis with potential future work.

Bibliography

[1] B. Byram. “Aperture Domain Model Image REconstruction (ADMIRE) for improved ultrasound imaging”. In: *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Mar. 2017, pp. 6250–6253. DOI: 10.1109/ICASSP.2017.7953358.

[2] D. Hyun et al. “Beamforming and Speckle Reduction Using Neural Networks”. In: *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 66.5 (May 2019), pp. 898–910. DOI: 10.1109/TUFFC.2019.2903795.

[3] J.A. Jensen. “Field: A Program for Simulating Ultrasound Systems”. In: *10th Nordic-Baltic Conference on Biomedical Imaging Published in Medical & Biological Engineering & Computing*. Vol. 34. 1996, pp. 351–353.

[4] A. C. Luchies and B. C. Byram. “Deep Neural Networks for Ultrasound Beamforming”. In: *IEEE Transactions on Medical Imaging* 37.9 (Sept. 2018), pp. 2010–2021. DOI: 10.1109/TMI.2018.2809641.

[5] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. “U-net: Convolutional networks for biomedical image segmentation”. In: *International Conference on Medical image computing and computer-assisted intervention*. Springer. 2015, pp. 234–241.

[6] R. J. G. van Sloun, R. Cohen, and Y. C. Eldar. “Deep Learning in Ultrasound Imaging”. In: *Proceedings of the IEEE* (2019), pp. 1–19. DOI: `10.1109/JPROC.2019.2932116`.