Inverse Problem Formulation and Deep Learning Methods for Ultrasound Beamforming and Image Reconstruction

Sobhan Goudarzi

A Thesis
in
The Department
of
Electrical and Computer Engineering

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

November 2022

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the thesis prepared

By: **Mr. Sobhan Goudarzi**
Entitled: **Inverse Problem Formulation and Deep Learning Methods for Ultrasound Beamforming and Image Reconstruction**

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

Doctor of Philosophy (Electrical and Computer Engineering)

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.
Signed by the Final Examining Committee:

_____ Chair
Dr. Adam Krzyzak

_____ External Examiner
Dr. Jeremy Dahl

_____ External to Program
Dr. Thomas Fevens

_____ Examiner
Dr. Dongyu Qiu

_____ Examiner
Dr. Jun Cai

_____ Supervisor
Dr. Hassan Rivaz

Approved by _____
Dr. Jun Cai, Graduate Program Director

November 22, 2022
Date of Defense

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

Abstract

Inverse Problem Formulation and Deep Learning Methods for Ultrasound Beamforming and Image Reconstruction

Sobhan Goudarzi, Ph.D.
Concordia University, 2022

Ultrasound imaging is among the most common medical imaging modalities, which has the advantages of being real-time, non-invasive, cost-effective, and portable. Medical ultrasound images, however, have low values of signal-to-noise ratio due to many factors, and there has been a long-standing line of research on improving the quality of ultrasound images. Ultrasound transducers are made from piezoelectric elements, which are responsible for the insonification of the medium with non-invasive acoustic waves and also the reception of backscattered signals. Design optimizations span all steps of the image formation pipeline, including system architecture, hardware development, and software algorithms. Each step entails parameter optimizations and trade-offs in order to achieve a balance in competing effects such as cost, performance, and efficiency.

The current thesis is devoted to research on image reconstruction techniques in order to push forward the classical limitations. It is tried not to be restricted into a specific class of computational imaging or machine learning method. As such, classical approaches and recent methods based on deep learning are adapted according to the requirements and limitations of the image reconstruction problem. In other words, we aim to reconstruct a high-quality spatial map of the medium echogenicity from raw channel data received from piezoelectric elements. All other steps of the ultrasound image formation pipeline are considered fixed, and the goal is to extract the best possible image quality (in terms of resolution, contrast, speckle pattern, etc.) from echo traces acquired by transducer elements.

Two novel approaches are proposed on super-resolution ultrasound imaging by training deep models that create mapping functions from observations recorded from a single transmission to high-quality images. These models are mainly developed to resolve the necessity of several transmissions, which can potentially be used in applications that require both high framerate and image quality.

The remaining four contributions are on beamforming, which is an essential step in medical ultrasound image reconstruction. Different approaches, including independent component analysis, deep learning, and inverse problem formulations, are utilized to tackle the ill-posed inverse problem of receive beamforming. The primary goal of novel beamformers is to find a solution to the trade-off between image quality and framerate.

The final chapter consists of concluding remarks on each of our contributions, where the strengths and weaknesses of our proposed techniques based on classical computational imaging and deep learning methods are outlined. There is still a large room for improvement in all of our proposed techniques, and the thesis is concluded by providing avenues for future research to attain those improvements.

Acknowledgments

First, I would like to thank my dear supervisor, Dr. Hassan Rivaz, for his remarkable scientific guidance. His support has not been limited to technical content, as I found him dependable in all aspects. The way he supervises the students helped me comprehend how to work and live better. I truly appreciate your patience and all you have done during my studies at Concordia University.
Second, I had the pleasure of collaborating with insightful scientists Dr. Adrian Basarab and Dr. Amir Asif in some of my Ph.D. projects. I am really grateful for their help, guidance, and support.
I would also like to extend my deepest gratitude to the committee members, Dr. Jeremy Dahl, Dr. Thomas Fevens, Dr. Dongyu Qiu, and Dr. Jun Cai, for evaluating this dissertation and their thoughtful feedback.
My special thank goes to my wife and my family for their emotional support. The presence of my wife made the Ph.D. journey sweet and memorable. The encouragement received from my parents and siblings always pushes me forward.

Contents

List of Figures

List of Tables

List of Abbreviations

CUBDL	Challenge on Ultrasound Beamforming with Deep Learning
IUS	International Ultrasonics Symposium
ROI	Region Of Interest
T/R	Transmit/ Receive
ADC	Analog to Digital Converter
RF	Radio-Frequency
CNN	Convolutional Neural Networks
GAN	Generative Adversarial Network
MRI	Magnetic Resonance Imaging
MLP	MultiLayer Perceptron
ReLU	Rectified Linear Unit
NRMSE	Normalized Root Mean Square Error
SSIM	Structural SIMilarity
MG	Mean Gradient
CNR	Contrast to Noise Ratio
FWHM	Full Width at Half Maximum
MOS	Mean Opinion Score
REFoCUS	Retrospective Encoding For Conventional Ultrasound Sequences
MVB	Minimum Variance Beamforming

DL	Deep Learning
CT	Computed Tomography
IQ	In-phase/Quadrature
CPWC	Coherent Plane-Wave Compounding
USTB	UltraSound ToolBox
CR	Contrast Ratio
gCNR	generalized Contrast to Noise Ratio
PWI	Plane-Wave Imaging
PSF	Point Spread Function
TRF	Tissue Reflectivity Function
DNN	Deep Neural Network
BN	Batch Normalization
PICMUS	Plane-Wave Imaging Challenge in Medical Ultrasound
MSE	Mean Square Error
SSNR	Speckle Signal-to-Noise Ratio
ICA	Independent Component Analysis
SR	Simulation Resolution
SC	Simulation Contrast
ER	Experimental Resolution
EC	Experimental Contrast
CC	Carotid Cross
CL	Carotid Longitudinal
DAS	Delay-And-Sum

PnP	Plug-and-Play
ADMM	Alternating Direction Method of Multipliers
RED	REgularization by Denoising
CS	Compressive Sensing
EMV	Eigenspace-based Minimum Variance
KS	Kolmogorov–Smirnov
PCF	Phased Coherence Factor
GCF	Generalized Coherence Factor
F-DMAS	Filtered-Delay Multiply And Sum
SVD	Singular Value Decomposition
AWGN	Additive White Gaussian Noise
BCCB	Block Circulant with Circulant Block
TGC	Time Gain Compensation
BFGS	Broyden–Fletcher–Goldfarb–Shanno
UFSB	Ultrasound Fourier Slice Beamforming

Chapter 1

Introduction

A part of this chapter is based on our published paper [1].

This chapter starts with a short review of medical ultrasound imaging and its applications, followed by describing the problem of interest, its challenges, our motivations, and our goals. Afterward, a roadmap of the thesis, including an overview of each chapter, is presented. Finally, the publications culminated from the current Ph.D. dissertation are listed.

1.1 Medical ultrasound imaging

While it is hard to specify the beginning of medical ultrasound imaging, it may go back to World War II and has been developed for several years since 1942 [7]. Inspired by the successful applications of sonar and radar, the idea of using pulse-echo techniques for imaging the human body was established. Initial versions of the supersonic reflectoscope (an early instrument for inspecting inside solid parts) applied to the human body were commercialized in the late 1940s and early 1950s [8]. Since those days, this medical imaging modality has experienced several revolutions, among which the invention of ultrafast ultrasound imaging and portable scanners are examples of the most recent ones. Besides design optimization, medical ultrasound has also been applied to many new applications, such as shear-wave elastography, functional ultrasound, and transcranial ultrasound, to name a few.

1.1.1 Image formation pipeline

In short, the process of image formation in medical ultrasound imaging includes three main steps. First, a set of excitation electrical pulses are applied to the piezoelectric crystals in order to form an acoustic pressure field with desired characteristics. Second, the reflected pulse pressures from scatterers are converted to electrical signals through crystals. Finally, received signals are processed to obtain the desired image. Fig. 1.1 shows the critical components of a medical ultrasound system, and the role of each part in the imaging pipeline can be explained as following [9]:

- **System Processor:** the first step is to insonify the region of interest (ROI) with the desired wave. Several parameters need to be specified, such as the center frequency of the excitation pulse, bandwidth, amplitude, initial phase, etc. In addition to imaging settings, the sequence of transmission and synchrony of other hardware elements are all managed in

the system processor. Furthermore, display configuration, as well as the post-processing of radio-frequency (RF) image resulting from the receive beamforming, is also implemented in this module.

- **Transmit Beamforming:** depending on the imaging technique (e.g., line-per-line, plane-wave, or synthetic aperture imaging), the number of transmitting elements are selected, and their excitation pulses and firing times are set.
- **Pulser and Switch:** the selected settings are used to create the excitation pulses of the piezoelectric crystal elements of the transducer. Afterward, they are fed to the multiplexer using a set of transmit/ receive (T/R) switches.
- **Transducer:** the ultrasound probe is made of several piezoelectric elements in a specific shape depending on the probe type (e.g., linear, curvilinear, or phased array). These elements are used for the transmission of acoustic waves into the tissue and also the reception of the echo signals. Depending on the hardware configuration, a multiplexer/ demultiplexer (MUX/ DEMUX) might also be used to match the signal lines.
- **Analog Front-End:** once the ultrasound wave is transmitted, the elements record the backscattered signals, which will be transferred into an analog to digital converter (ADC).
- **Receive Beamforming:** in order to reconstruct a spatial map of the tissue echogenicity, it is necessary to trace back the backscattered echoes from each voxel of the medium.
- **Display:** once the raw RF image of the medium is created, it is subject to envelope detection and logarithmic compression for final display. The graphical user interface provided on the screen facilitates the selection of imaging settings and the investigation of their effect on the final image quality.

As mentioned before, ultrasound imaging systems have noticeably changed over time. Fig. 1.2 shows three different generations of Siemens ultrasound machines [10, 11, 12]. Thanks to advances in hardware technology, the probe sizes got smaller, and many processing steps can be digitally accomplished. In recent pocket-size versions shown in Fig. 1.2 (c), the raw data are collected on a battery-operated portable probe, and the image reconstruction is completed on cell phone or tablet devices.

1.1.2 Imaging techniques

The piezoelectric crystal elements can be used to create mechanical ultrasound waves, which commonly refer to sound waves above the human audible range (i.e., 20 Hz up to 20 kHz) [13].

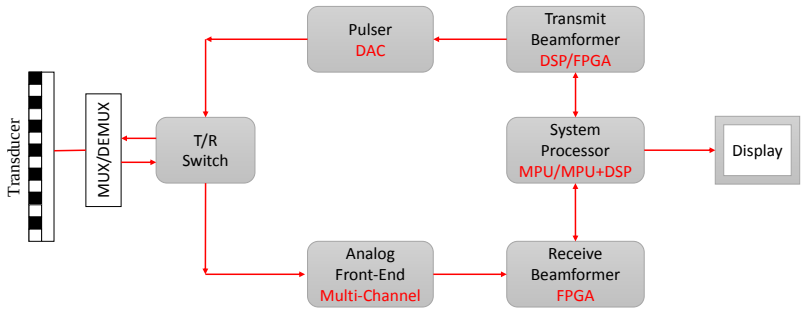

Figure 1.1: Key components of an ultrasound system.

The transducer is the most crucial part of an ultrasound imaging system, and there are different types of medical ultrasound probes, such as linear, curvilinear, phased array, etc. Each probe type results from arranging several small piezoelectric crystal elements in a specific architecture based on the application requirements. For example, linear array transducers generally have a higher transmission center frequency and a fixed field of view over depth. In comparison, convex array transducers work on the lower frequency to have a larger penetration depth. Generally, the key specifications of an ultrasound probe are the number of elements, elements' size, pitch (i.e., the distance between the centers of two neighbor elements), elevational focus, sensitivity, and frequency response.

Once the probe is fabricated, there are different ways to insonify the medium and form an image. In other words, the sequence of elements firing and recording the backscattered echoes specify the imaging technique. There are three common approaches as follows:

- **Classical line-per-line technique:** this approach, also known as focused imaging, is the most common technique in commercial scanners [7]. In short, this method entails firing a few elements to insonify a narrow ROI using a focused ultrasound beam, which results in a narrow image. This process is consecutively repeated to insonify the whole medium, and the final image is formed by concatenating the narrow images [7]. The main problem of this approach is the low framerate (i.e., temporal resolution) since it requires several transmissions. Moreover, the axial resolution is not constant over depth because the beam is focused on a single point. Multi-focus and multi-line acquisition are enhanced versions of this technique to solve the aforementioned problems, respectively [14].

- **Synthetic aperture imaging:** this approach, also known as element-by-element transmission, entails firing a single element in each step and recording the backscattered waves with all elements of the probe. The final image is formed by combining the resulting images of every single transmission. Synthetic aperture imaging suffers from a low signal-to-noise ratio (SNR) as well as a limited depth of penetration because of using a single element in emission [15]. Advanced versions of this method consider a virtual source behind the transducer, resulting in multi-element transmission and increased transmitted energy [16].

- **Plane-wave imaging (PWI):** this approach, also known as ultrafast imaging [17], entails the firing of all elements in a single shot to insonify the whole medium with a plane-wave.

(a) (b) (c)

Figure 1.2: Three different generations of Siemens ultrasound machines. (a) Vidoson 635 [5]. (b) ACUSON Redwood [6]. (c) ACUSON Freestyle [7]. All images obtained from Siemens company website with permission.

Thanks to hardware developments and parallel processing, the echo signals are recorded using all probe elements. Therefore, the framerate can reach several thousand per second since it only requires a single transmission. The main issue with ultrafast imaging is the poor quality of resulting images due to unfocused transmitted wave. Coherent plane-wave compounding (CPWC) [18] is an interesting way to improve the quality of plane-wave images, which consists of transmitting several tilted plane waves of different angles and combining the resulting RF images. This solution, however, poses a trade-off between the image quality and the framerate.

Once the acoustic wave is transmitted into the medium using any of the above techniques, the piezoelectric crystal elements work as receivers and record the backscattered signals. Consequently, the problem is reconstructing a spatial map of the tissue echogenicity by tracing back the echoes from each voxel of the medium.

1.1.3 Ultrasound beamforming

As seen in Fig. 1.1, beamforming is an essential step in the medical ultrasound image formation pipeline. In transmit beamforming, the shape of the excitation pulse and the firing time applied to each element of the probe are specified to insonify the medium with the desired acoustic wave [9]. Receive beamforming is to trace back the backscattered signals from each voxel of the medium [9] in order to reconstruct a high-quality ultrasound image, which is a spatial map of the target echogenicity. In other words, each pixel of the final image corresponds to an averaged tissue reflectivity function over the extent of the pixel. Receive beamforming can be accomplished in different ways, among which Delay-And-Sum (DAS) is the most common approach. Since the current thesis is mainly on ultrafast imaging and linear probe, a detailed explanation of image reconstruction using DAS for linear probes in CPWC is provided here. This can easily be extended to other imaging techniques and probe types. To keep this chapter concise, further explanations are not included, and interested readers can find more details in [7, 9].

Without loss of generality, let us assume a linear array of n elements, symmetrically lying on the x-axis, transmitting toward the positive direction of z-axis (illustrated in Fig. 1.3). Consider a plane wave with angle α which spans the domain with characteristic sound speed c. Backscattered signals received by element i located at x_i is denoted by $\mathbf{h}_i(t)$. d_t is defined as the transmission distance from the origin of the transmitted plane-wave to an arbitrary point (x, z) in the ROI and equals $zcos(\alpha) + xsin(\alpha)$ [18]. d_r is defined as the receiving distance from (x, z) to the location of element i and equals $\sqrt{(x - x_i)^2 + z^2}$ [18]. The RF data corresponds to (x, z), in $\mathbf{h}_i(t)$ can

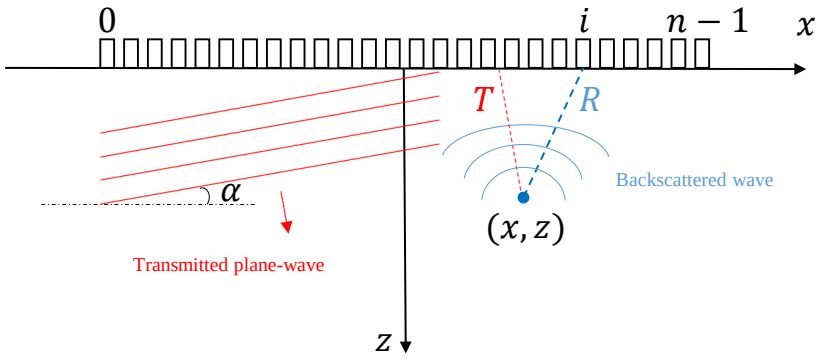

Figure 1.3: Geometrical illustration of the PWI. Notation α is the steering angle of the incidence wave (from [1]).

be found by applying propagation delay τ as follows (hereafter, capital and bold font variables represent matrices and vectors, respectively):

$$\tau(x,z) = \frac{d_t + d_r}{c} \Longrightarrow R_i(x,z) = \mathbf{h}_i(\tau(x,z)), \tag{1.1}$$

where R_i is the RF matrix of the ROI reconstructed from the output of channel i. S_α is defined as the resulting image of ROI when a plane-wave with angle α is transmitted. Each point (x,z) of S_α can be obtained through a weighted summation of RF data corresponding to receiving elements as follows:

$$S_\alpha(x,z) = \sum_{i=0}^{n-1} \mathbf{w}_r(i) R_i(x,z), \tag{1.2}$$

where $\mathbf{w}_r$ is the receive apodization window of length n. In practice, however, dynamic beamforming is implemented. More specifically, to fix the f-number ($f\#$) in the entire image, a specific number of elements contribute to each pixel value of the final image. Therefore, a part of probe elements with the width of l are considered for the reconstruction of each depth of the image as following:

$$f\# = \frac{z}{l}. \tag{1.3}$$

In CPWC, several plane-waves with different angles are transmitted to increase the quality of imaging. Therefore, each point (x,z) of the final image S is reconstructed as a linear combination of images of different angle S_α as follows:

$$S(x,z) = \sum_{j=1}^{m} \mathbf{w}_t(\alpha_j) S_{\alpha_j}(x,z), \tag{1.4}$$

where $\mathbf{w}_t$ is the vector of angular apodization. m is the maximum number of transmitted plane-waves in CPWC which is computed as follows [19]:

$$m = \frac{l_{max}}{\lambda \times f\#}, \tag{1.5}$$

where l_{max} is the aperture size, λ denotes the transmitted pulse wavelength. m steered plane waves linearly spaced between $[-\alpha_{max}, \alpha_{max}]$ wherein:

$$\alpha_{max} = \frac{1}{2 \times f\#}, \tag{1.6}$$

Once the final RF image of the medium is created, it is subject to envelope detection, which is commonly performed using the Hilbert transform. Afterward, *log* compression is applied in order to obtain the final B-Mode ultrasound image.

1.2 Problem statement

Generally, research in the field of ultrasound imaging can be classified into two main branches. First, design optimization deals with improving the quality of ultrasound imaging in terms of performance, efficiency, and cost. Second, studies on ultrasound imaging applications aim to introduce increasing utilization in diagnosis and treatment. The current thesis is among the first group of studies. There are three general categories of design optimization approaches focused on hardware development,

system architecture, and software algorithms. Our work is restricted to the third category based on our background knowledge and area of expertise.

As for the optimization of software algorithms, on the one hand, researchers study beam design and transmit sequencing. Design options include transmit beamforming, excitation sequence, transmit center frequency, and use of time-reversal approach. On the other hand, researchers investigate how to convert the raw RF echo traces acquired by transducer elements to a high-quality spatial map of target echogenicity. Design choices include receive beamforming, use of fundamental or harmonic frequency components, as well as post-processing algorithms such as temporal or spatial filtering, speckle reduction methods, phase aberration correction, and any other nonlinear transformations. Although we have investigated various problems during this Ph.D. study, the current dissertation can be considered among the second group, mainly focused on ultrasound beamforming and image reconstruction.

Conversion of the RF channel data to the final B-mode image is subject to three evaluation criteria, including the image quality, the speed of reconstruction (i.e., the computational cost), and the hardware requirements. Generally, finding a fast way which produces high-quality images using available hardware resources is challenging. Suppose we push aside the hardware limitations and assume that there is no limitation on practical implementation. In that case, there is still a trade-off between image quality and framerate (i.e., temporal resolution). For example, simple methods such as DAS do not provide high quality, while adaptive methods such as minimum variance beamforming (MVB) cannot be implemented in real-time. To be more precise, the image quality indexes are even divergent, meaning that improving one index may come at the expense of another. For example, MVB mainly improves the lateral resolution and extended versions, which simultaneously improves the contrast but disturbs the speckle information.

Considering the above challenges, the current thesis is motivated by the development of novel computing techniques to improve medical ultrasound imaging reconstruction. Our research covers different approaches of dealing with this problem. The classical approaches, as well as recent methods, which are mainly based on machine learning and deep learning, are adapted based on the requirements and limitations of the problems in order to acquire desired results. A big picture of each proposed method is provided in Section 1.3. Our primary goal is to reconstruct a high-quality image as fast as possible. By taking advantage of each class of methods, we propose novel ideas to push forward the classical limitations.

It has to be mentioned that since our work is not limited to a single problem or class of methods, instead of writing a general literature review in the introduction section, more specific literature reviews are provided at the beginning of each chapter. Moreover, a detailed review of the problem for each chapter is provided if it was necessary.

1.3 Roadmap of the thesis

As reviewed before, the raw RF data are the digitized version of backscattered waves recorded by the probe elements, which can be considered as our observation from the medium. Generally, the question partially answered in the current dissertation is how to extract a high-quality image of the medium using those observations. The remainder of the thesis is organized as follows:

Chapter 2 proposes a novel approach for multi-focus ultrasound imaging based on generative adversarial networks (GANs). The classical line-per-line technique involves the transmission of focused ultrasound beams into the medium. As a result, image quality is low out of the focus points. Therefore, several focused beams must be transmitted to have a uniform quality over depth. This solution, however, reduces the temporal resolution by a factor equal to the number of transmissions.

As a solution to this dilemma, GANs are adapted to construct a mapping function between the result of a single transmission and multiple transmissions. This method can be considered among super-resolution ultrasound methods that help not lose the framerate in multi-focus imaging.

Chapter 3 uses deep learning to propose a general ultrasound beamformer. MVB is among the best adaptive beamformers, providing high image quality regardless of the imaging settings. However, this method is computationally very expensive, such that a few minutes might be required to reconstruct a single image. Therefore, deep learning is used to mimic MVB and speed up the reconstruction process.

In Chapter 4, the potential of deep learning in ultrasound image reconstruction is further investigated. An artificial high-quality point spread function (PSF) is considered for the imaging system. We use this PSF for producing high-quality simulated images as training targets. Afterward, the generalization performance of the trained model is evaluated for unseen simulation and *in vivo* data.

While the first three chapters are about deep learning, the remaining chapters are not related to deep learning, and we investigate ultrasound beamforming in another framework. Chapter 5 introduces an adaptive ultrasound beamformer based on independent component analysis (ICA). More specifically, the tissue reflectivity function (TRF) of the medium is considered a single source observed by several transducer elements. In other words, each RF image of the medium reconstructed from elements' output is a single observation of the desired TRF. In this way, beamforming can be implemented using ICA through which the independent source is extracted from several noisy observations.

Chapter 6 introduces a solution for the inverse problem of ultrasound beamforming. While it is common in the literature to consider ℓ_1-norm regularizers to enforce the sparsity of solutions, the resulting image suppresses the speckle information of the medium. By taking advantage of recently proposed regularization functions, a novel framework is proposed in which any advanced denoising technique can be used as a regularization function.

Finally, in Chapter 7, we propose a framework for unifying the inverse problems of beamforming and deconvolution. The inverse problem of beamforming deals with the raw RF channel data to reconstruct the image of the medium, while the deconvolution problem uses beamformers' output (usually based on DAS) to recover the desired TRF. Our method poses a way to simultaneously solve both problems and find a better solution than the solutions of separated problems or even once they are applied sequentially.

1.4 List of publications

- **Journal papers:**

  1. Goudarzi, Sobhan, Amir Asif, and Hassan Rivaz. "Fast multi-focus ultrasound image recovery using generative adversarial networks." IEEE Transactions on Computational Imaging 6 (2020): 1272-1284.
  2. Goudarzi, Sobhan, Amir Asif, and Hassan Rivaz. "Plane-Wave Ultrasound Beamforming Through Independent Component Analysis." Computer Methods and Programs in Biomedicine 203 (2021): 106036.
  3. Goudarzi, Sobhan, and Hassan Rivaz. "Deep reconstruction of high-quality ultrasound images from raw plane-wave data: A simulation and in vivo study." Ultrasonics 125 (2022): 106778.

4. Goudarzi, Sobhan, Adrian Basarab, and Hassan Rivaz. "Inverse Problem of Ultrasound Beamforming with Denoising-Based Regularized Solutions." IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control (2022).
5. Goudarzi, Sobhan, Adrian Basarab, and Hassan Rivaz. "A Unifying Approach to Inverse Problems of Ultrasound Beamforming and Deconvolution." Submitted to IEEE Transactions on Computational Imaging.
6. Goudarzi, Sobhan, Jesse Whyte, Mathieu Boily, Anna Towers, Robert D. Kilgour, and Hassan Rivaz. "Segmentation of Arm Ultrasound Images in Breast Cancer-Related Lymphedema: A Database and Deep Learning Algorithm." Submitted to IEEE Transactions on Biomedical Engineering.

- **Conference papers:**

1. Goudarzi, Sobhan, Amir Asif, and Hassan Rivaz. "Multi-focus ultrasound imaging using generative adversarial networks." 2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019). IEEE, 2019.
2. Goudarzi, Sobhan, Amir Asif, and Hassan Rivaz. "High Frequency Ultrasound Image Recovery Using Tight Frame Generative Adversarial Networks." 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE, 2020.
3. Goudarzi, Sobhan, Amir Asif, and Hassan Rivaz. "Ultrasound beamforming using mobilenetv2." 2020 IEEE International Ultrasonics Symposium (IUS). IEEE, 2020.
4. Goudarzi, Sobhan, Amir Asif, and Hassan Rivaz. "Angular Apodization Estimation Using Independent Component Analysis in Coherent Plane-Wave Compounding." 2020 IEEE International Ultrasonics Symposium (IUS). IEEE, 2020.
5. Goudarzi, Sobhan, Amir Asif, and Hassan Rivaz. "Pruning MobileNetV2 for Efficient Implementation of Minimum Variance Beamforming." International Workshop on Advances in Simplifying Medical Ultrasound. Springer, Cham, 2021.
6. Goudarzi, Sobhan, and Hassan Rivaz. "Deep Ultrasound Denoising Without Clean Data." Accepted in SPIE Medical Imaging 2023.

Chapter 2

Multi-Focus Ultrasound Image Recovery Using Generative Adversarial Networks

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

In conventional line-per-line ultrasound imaging, transmitted beams are focused in order to have higher intensity and better lateral resolution at a specific depth. Indeed, focusing means aligning the pressure fields of all elements of the aperture to simultaneously arrive at a specific field point [21]. Focusing can be done through a physically curved aperture or electronic beamforming. Focused beams have a complex bowtie shape with side lobes and grating lobes [13]. In classical focused transmission, it is assumed that received echoes are brought about by scatterers from within the main transmitted ultrasound beam. However, if there is a strong reflector outside of the main beam, it may cause detectable echoes for transducer and will be falsely displayed. This problem is called beam width artifact [22]. Hence, the narrower the transmitted beam, the lower the beam width artifacts.

When the beam is focused, the quality of the image is optimal at the focal point and progressively degrades away from it. Therefore, in order to preserve optimal lateral resolution everywhere along the axial direction, several beams focused at different depths are often transmitted. Consequently, the multi-focus ultrasound image can be recovered. However, this approach drastically reduces the framerate which is inversely proportional to the number of transmissions. Therefore, there is a trade-off between the lateral resolution and framerate in classical focused transmission. It has to be mentioned that when the depth of imaging is limited, image degradation due to beam divergence is limited. Therefore, if there is no clearly discernible target such as a cyst or hyperechoic region, the difference between the quality of single and multi-focus images is difficult to observe. Another issue arising in this method is the assumption of having no relative motion between the tissue and the probe while transmitting several beams. This assumption is not practical in several applications such as in imaging regions close to the heart, a major artery, and in obstetric sonography. For example, in cardiac imaging, the motion blur is large even in between different lines, which has led to the advent of multi-line acquisition (MLA) methods [14]. Hand motion and tremor are additional sources of relative motions.

Inspired by the success of deep learning algorithms, we propose a data-driven method for multi-focus line-per-line ultrasound imaging with only a single focused transmission and without a loss in framerate. More specifically, we train a generative adversarial network (GAN) [23] to form a mapping function between non focused and focused ultrasound images.

Convolutional neural networks (CNNs) are able to efficiently extract necessary features from raw data, and there is no need to engineer hand-crafted features anymore [24]. CNNs have been successfully used in variety of applications such as classification, super resolution, denoising, etc. Defining a proper objective function to minimize during the training phase is a critical factor that influences the performance of the network [25, 26, 27]. GANs address this issue by using CNNs to automatically learn an objective function appropriate for satisfying the specific task. More specifically, GANs consist of generator and discriminator networks, which compete with each other. Generally, the generator does a mapping from input space to a real desired space, and discriminator specifies the quality of generated data. Hence, the discriminator is the objective function for generator, and the generator tries to fool the discriminator by generating more realistic data [23]. Both networks are interestingly trained during the training process, which entails solving a minimax game to find the Nash equilibrium of these two competing networks.

2.1 Related work

Training dynamics of GANs were theoretically investigated in [28], which has led to several contributions in improving the training process [29, 30], or finding a working architecture [31] tailored for specific applications. Subsequently, Arjovsky *et al.* [32] exploited the concept of integral probability metric [33] and introduced wasserstein GAN (WGAN). Although it resolved some issues, it has a limited success because of using weight clipping to enforce a Lipschitz constraint on the discriminator. This problem was solved by penalizing the norm of gradient of the discriminator over interpolation between generated and real data [34, 35]. Another notable contribution was proposed by Roth *et al.* [36] where a gradient norm penalty similar to [34] is introduced, except that there is no interpolation and f-divergences is instead used.

In spite of such important theoretical contributions, there is still no clear understanding on why the discriminator objective function is critical in stable training of GANs. Moreover, it has been shown that most of reviewed models can reach similar scores with non-saturating GAN introduced in [23], and there is no evidence that any of them consistently outperforms the non-saturating GAN. Using a different approach, another method for training GANs was proposed entitled boundary-seeking GANs (BSGANs) [37]. BSGAN is based on providing a policy gradient for training the generator that forces the generator to produce samples which are near the decision boundaries (i.e., the discriminator cannot distinguish real or generated data). In addition to better training behavior, BSGAN works for discrete as well as continuous data.

Application of GANs to different tasks such as classification and regression, image synthesis, image to image translation, and super-resolution is a growing interest. Herein, we confine our literature review on most important contributions in the filed of medical imaging. Yang *et al.* [38] used WGANs for denoising low-dose computed tomography (LDCT) images. They also took advantage of pretrained VGG-19 network [39] for feature extraction and defining a perceptual loss function instead of MSE loss function. However, VGG-19 was trained on color images, and they duplicated the gray-scale channels to be able to feed CT images to VGG-19 network [38]. Simultaneously, another work on LCDT denoising was published which utilizes a Conveying Pathbased convolutional encoder-decoder (CPCE) network as the generator in a WGAN structure [40]. In another application, conditional GANs were used for reconstruction of magnetic resonance

imaging (MRI) data recorded for a compressed sensing scenario [41]. The main idea in conditional GANs is conditioning both the generator and discriminator networks on some extra information [42]. In this work, frequency-domain information were used for conditioning the networks in order to have results that are similar in both time and frequency domains [41]. Nie *et al.* [43] used GANs for medical image synthesis. Their method was validated on reconstruction of MRI images from CT images and also generating 7T MRI from 3T MRI images. Recently, Mardani *et al.* [44] proposed a compressed sensing framework that uses GAN to remove the aliasing artifacts of undersampled MRI images.

In line-per-line ultrasound imaging, multilayer perceptron (MLP) was used for correction of phase aberrations [45] a long time ago. After many years, a deeper version of MLP was used for ultrasound beamforming [46], which trained several networks in frequency subbands to suppress off-axis scattering and remove clutter from channel data. This work used fully connected networks, which are prone to overfitting compared to CNNs. The reconstruction of B-Mode images from sub-sampled radio-frequency (RF) data using CNNs was investigated in [47]. Recently, CNNs were used for speckle reduction [48]. In ultrafast imaging, Gasse *et al.* [49] recovered high-quality plane-wave images from a limited number of transmitted angles using CNNs in a pilot study. Zhou *et al.* [50] improved the same idea and used multi-scale structure CNNs on different channels for recovery. To preserve the speckle information, wavelet postprocessing was added to the output of the network.As for the application of GANs in ultrasound imaging, in [51], a context-conditional GAN was used to acquire the quality of 128-channel B-Mode images from 32 channels. Speckle reduction was done using GAN in [52]. Recovery of high quality plane-wave images from a limited number of transmitted angles using GANs was performed in [53].

As for the purpose of multi-focal imaging, Bottenus [54] proposed a method based on formulating a new frequency domain transmit encoding matrix that incorporates both delay and apodization to recover synthetic transmit aperture dataset. This method allows for synthetic transmit focusing at all points in the field of view. However, it is originally designed for phased array sequences in which the radial scan lines increase in separation in the axial direction. Consequently, this method was demonstrated on a walking aperture curvilinear sequence [55]. Using the regularized inverse of encoding matrix, the possibility of recovering synthetic transmit aperture dataset at each frequency for walking sequences was demonstrated in [56]. Recently, Ilovitsh *et al.* [57] proposed an approach which relies on superposition of axial muti-foci waveforms in a single transmission. Despite substantially advancing the state-of-art, this method has two limitations. First, superposition can only be completed on a subset of probe crystals because of the piezoelectric maximal element response producing nonuniform quality in the axial direction. Second, it leads to an increase in thermal index due to transmissions of longer durations.

Herein, the central idea is generating several focal points by sending only one focused transmit beam. The nonlinear propagation pattern of the ultrasound beam is not stationary along the axial direction. Accordingly, in order to achieve a narrow beam everywhere, a mapping function between the single-focus ultrasound image and multi-focus version is estimated through different GANs. More specifically, the optimal focus depth of the transmit beam is found to be in the middle of imaging depth. The number of networks depends on the depth of imaging. In current study, we consider two networks to recover shallow and deep regions. A preliminary version of this work was presented in ISBI 2019 [58].

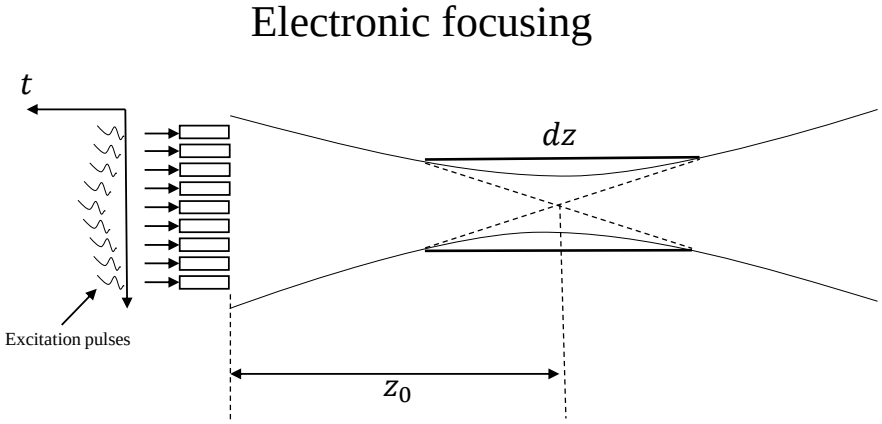

Figure 2.1: Electronic focusing of the transmit beam by applying the time delays shown in left.

2.2 Method

2.2.1 Focusing

In electronic beamforming, in order to focus at a specific axial depth (z_0), a set of excitation pulses with proper time delays are applied to the crystals. This method, as shown in Fig. 2.1, is always used in classical line-per-line imaging. The highest amplitude of acoustic potentials is achieved at focus point. Therefore, the distance between two points where the field on axis is 3dB less than at the focal point is defined as depth of focus (dz) [13]. The lateral resolution is optimum in this region. In order to preserve the lateral resolution (having optimal multi-focus image), the maximum distance between transmitted focal points has to be equal to the depth of focus. We formulate our problem as finding a nonlinear mapping function which transforms the bowtie-shaped focused beam (with one focal point) to a thin cylindrical beam. However, this nonlinear function is nonstationary along the axial direction. In other words, this function varies with depth and cannot be estimated through only one network. Therefore, different networks should be trained that correspond to different depths. Consequently, the proposed method is based on partial estimation of nonlinear function for multiple depth intervals. This is a common solution for addressing nonstationary problems such as spectrum estimation. Therefore, we break the image into limited number of intervals along the axial direction such that we get closer to the stationary assumption in training convolutional neural networks and have a lower amount of variation, and subsequently train a BSGAN for each interval.

2.2.2 Proposed recovery method

Let x be a sample of input space, $\{x^{(i)} \in \mathbb{R}^{r\times c}\}_{i=1}^{m}$, which is an ultrasound image with single focus point (m denotes the number of samples. symbols r and c, respectively, denote the number of rows and columns), and y be the corresponding sample of output space, $\{y^{(i)} \in \mathbb{R}^{r\times c}\}_{i=1}^{m}$, which is a multi-focus ultrasound image. We formulate the problem as:

$$y = \mathcal{F}(x), \tag{2.1}$$

where $\mathcal{F} : \mathbb{R}^{r\times c} \to \mathbb{R}^{r\times c}$ denotes the recovery function. Herein, a few main points have to be considered regarding the proposed recovery scheme. First, we assume that the recovery function $\mathcal{F}$ does really exist which means that it is possible to recover the multi-focus ultrasound image from a single-focus observation. Second, we assume that $\mathcal{F}$ can be estimated, with proper upper bound error, through GANs. In fact, the manifolds of input and output are in an unknown high

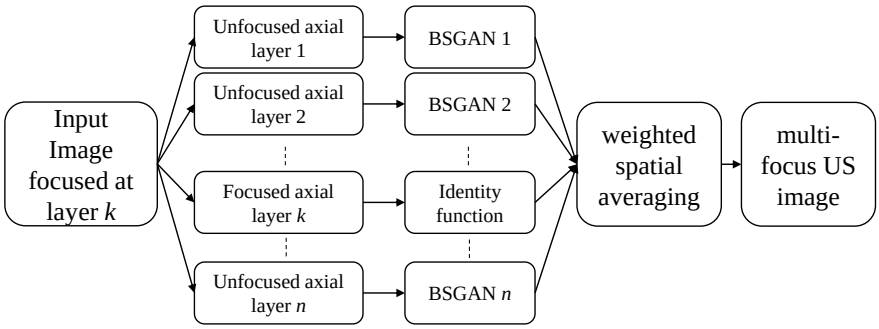

Figure 2.2: The proposed recovery scheme. The input image is focused at layer k on which no transformation is applied. All other unfocused axial layers are transformed through distinct BSGANs - one for each layer.

dimensional space ($r \times c$), and the problem is ill-posed. However, it has been shown that CNNs are able to efficiently represent the input data in middle layers and estimate any nonlinear function with a desired upper bound error [59, 60]. These are reasonable assumptions that need to be commonly made for deep learning, and their mathematical proof is beyond the scope of this thesis.

Our proposed recovery scheme is summarized in Fig. 2.2. First, the single-focus input B-Mode image is broken into a few axial layers. Then, for all layers where the transmitted beam is not focused, the mapping to the corresponding focused layer is achieved through a distinct BSGAN (i.e. a different network is trained for each axial layer). As the input image is focused at layer k, the output of this layer is the same as the input (i.e. an identity function is applied to this layer). Finally, all of the axial layers are merged together by minimal blending in small overlapping regions between layers in order to remove border effects as is the common practice, and multi-focus B-mode image is recovered.

2.2.3 Generative adversarial networks

Our aim is to estimate a nonlinear function that maps the input space to the target space. This aim can be fulfilled through CNNs. However, CNNs need an explicit differentiable objective function which scores the quality of results. Therefore, we need a distance measure $Dist$ between estimated output $\widehat{y}$ and desired output y. The problem can then be formulated as:

$$\widehat{\theta} = \arg\min_{\theta} \; Dist(\widehat{y}, y), \tag{2.2}$$

where θ is the parameters of the CNN. A long-running problem with CNNs is defining an appropriate distance measure. In other words, we still need to specify what we wish to minimize. As we will show in the Results Section, the commonly used MSE produces blurry results [61] because it averages across pixels. In the context of ultrasound imaging, this leads to incoherent averaging of the data which destroys the speckle pattern [18]. Fortunately, GANs give us the chance of reaching the desirable results only by specifying a high-level goal. What GANs learn is a loss function which classifies whether output is real or fake (the discriminator network) and a mapping function to minimize this loss (the generator network). Therefore, GANs consist of generator and discriminator networks, which compete with each other.

In classical form, GANs training is a min-max game between the generator and the discriminator [23]:

$$\min_{G} \max_{D} V(D,G) = \mathbb{E}_{y \sim p_{data}(y)} \left[log \; D(y) \right] + \mathbb{E}_{x \sim p_X(x)} \left[log(1 - D(G(x)) \right], \tag{2.3}$$

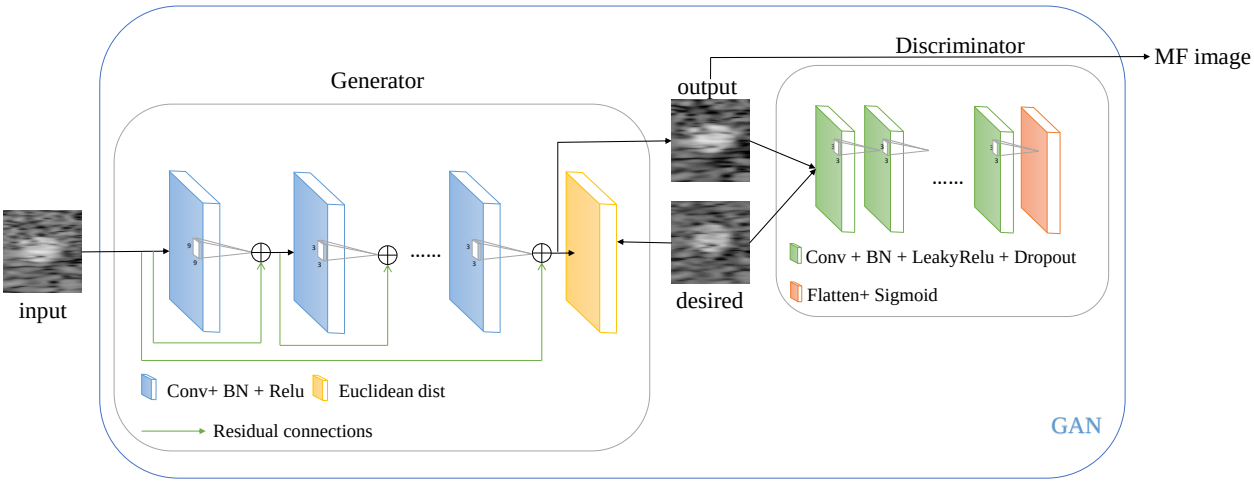

Figure 2.3: The structure of the proposed BSGANs.

where y and x are the desired and input respectively with $\widehat{y} = G(x)$ the estimated/generated output. E denotes the expected value, and D and G are the discriminator and generator, respectively, and $V(D, G)$ denotes the objective function for GAN training. $y \sim p_{data}(y)$ means y is a sample of data generating distribution while $x \sim p_X(x)$ means x is a sample of input distribution.

2.2.4 Boundary seeking generative adversarial networks

It can be shown from Eq. (2.3) that the optimal discriminator $D^*(y)$ is given by [23]:

$$D^*(y) = \frac{p_{data}(y)}{p_{data}(y) + p_g(y)}. \tag{2.4}$$

Hence, if the optimal discriminator with respect to generator is known, the global minimum of generator training will be $p_g = p_{data}$, wherein the desired distribution of output data is perfectly estimated by the generator, and the generator produces samples that are indistinguishable for discriminator. In practice, however, we are far from optimal case and the true data distribution, $p_{data}(y)$, could be achieved by weighting with the ratio of optimal discriminator as follows [37]:

$$p_{data}(y) = p_g(y)\frac{D^*(y)}{1 - D^*(y)}. \tag{2.5}$$

As the optimal discriminator is also unknown and hard to estimate, we always work with an approximation of $D^*(y)$. The intuition in training of GANs is that as we train the discriminator, it gets closer to $D^*(y)$, and consequently, the results improve. Eq. (2.5) means that the optimal generator is what makes the discriminator 0.5 everywhere, or a coin toss. In fact, $D(y) = 0.5$ is the decision boundary for a discriminator. So, BSGANs are a specific form of GANs in which generated data are close to the decision boundary of the discriminator [37].

The discriminator parameters ω are trained through the following optimization problem:

$$\widehat{\omega} = \arg\min_{\omega} L_D(\widehat{y}, y) = \arg\min_{\omega} L_{BCE}(D(y), 1) + L_{BCE}(D(\widehat{y}), 0), \tag{2.6}$$

where $L_D(\widehat{y}, y)$ is the loss function for the discriminator. Herein, we used binary cross entropy (BCE) which is defined as follows:

$$L_{BCE}(D(y), l) = -\sum_i [l_i log(D(y_i)) + (1 - l_i) log(1 - D(y_i))], \tag{2.7}$$

where l is the output label with values of $\{0, 1\}$. As can be seen from Eq. (2.6), the discriminator is trained to complete a two-class classification problem wherein the generated (i.e., not real) data is assigned to 0, and real data is assigned to 1.

The generator parameters θ are trained through the following optimization problem:

$$\widehat{\theta} = \arg\min_{\theta} L_G(\widehat{y}, y) = \arg\min_{\theta} \lambda_1 \parallel \widehat{y} - y \parallel^2 + \lambda_2 L_{BS}(\widehat{y}), \tag{2.8}$$

where $\parallel . \parallel$ is the second order norm, and λ_1 and λ_2 are regularization coefficients. The first term is the classical MSE loss function, and the second one $L_{BS}(\widehat{y})$ is the boundary seeking loss function which is defined as following:

$$L_{BS}(\widehat{y}) = \frac{1}{2m} \sum_{i=1}^{m} [log(D(\widehat{y}_i)) - log(1 - D(\widehat{y}_i))]^2. \tag{2.9}$$

In other words, we take advantage of both MSE and adversarial objective function to reach desirable results.

2.2.5 Proposed network

Our proposed network is shown in Fig. 2.3. The generator in Fig. 2.3 is a fully convolutional network with residual connections [62] consisting of 9 layers, where the first 8 layers contain 32 filters, and the last layer contains 1 filter. The first layer contains kernels of size 9×9, and other layers contain kernels of size 3×3. Each layer also contains batchnorm layer and a ReLU (rectified linear unit) activation function except for the last layer which uses tanh activation function in order to map the output values between $[-1, 1]$. As shown in Fig. 2.3, we used both overall and local residual connections. The discriminator in Fig. 2.3 consists of 4 layers containing 32, 64, 128, and 256 convolution filters with the same kernel size of 3. The first 3 layers have stride of 2, and the fourth layer has a stride of one. Each layer also contains LeakyReLU, batchnorm, and dropout (rate = 0.25) layers. The last layer is flattening with sigmoid activation for getting the output label. The number of filters and layers was chosen to maintain a minimum number of parameters for preserving the generalization performance and a more stable training. Kernel sizes were chosen empirically. We did not encounter checkerboard artifacts because the input and output patches have the same size. In summary, total number of trainable parameters for generator and discriminator networks are 68,000 and 400,000, respectively.

2.3 Experiments

2.3.1 Datasets

2.3.1.1 Simulated phantom

This dataset contains phantom simulations using the Field II program [63, 64]. The transducer configuration is described in Table 2.1. The sampling frequency is reduced to 10 MHz after envelope extraction to reduce the size. The phantoms typically consist of 100,000 scatterers (more than 30 scatterers per wavelength to ensure fully developed speckles) and a collection of three point targets, three cyst regions, and three highly scattering regions in three different axial depths. Four different shapes of cysts and highly scattering regions are simulated. For each shape, five different sizes are considered. Finally, for each size 40 independent realizations of scatterers are simulated. For each realization (i.e., each phantom), three different images were simulated by changing the location of

Figure 2.4: Real phantom experiment setup.

the focal point. Therefore, we have 4*5*40*3=2,400 different simulated images in total. The size of images is fixed as 40 mm lateral * 60 mm axial. We use line-per-line imaging with delay and sum beamforming.

2.3.1.2 Real phantom

Multi-Purpose Multi-Tissue Ultrasound Phantom (CIRS model 040GSE, Norfolk, VA) was used as real phantom. Ultrasound images were collected using an E-CUBE 12 Alpinion machine with L3-12H high density linear array and a centre frequency of 8.5 MHz. The sampling frequency of the RF data was 40MHz, and 384 RF lines were collected for each image. 20 independent images were collected at different locations of the phantom. At each location, three images with different

Parameter	Value	Unit
Array geometry	Linear	-
Number of elements	192	elements
Center frequency	3.5	MHz
Element width	0.44	mm
Element height	5	mm
Kerf	0.05	mm
Sampling frequency	100	MHz
Number of scan lines	50	lines
Speed of sound	1540	m/s

Table 2.1: Field II simulation setting

Algorithm 2.1 Minibatch stochastic gradient descent training of BSGANs. The number of steps to apply to the discriminator $N_D = 3$. All experiments in this chapter used Adam parameters, $\alpha = 10^{-4}$, $\beta_1 = 0.9$, $\beta_2 = 0.99$.

Require: set $\lambda_1 = 0.4$, $\lambda_2 = 90$.
Require: set the number of total epochs, $N_{epoch} = 100$, the batch size $m = 64$.
  calculate the number of iteration in each epoch
  $N_{iter} \leftarrow total\ number\ of\ training\ samples/m$
Require: ω_0, initial discriminator parameters. θ_0, initial generator parameters.
  for N_{epoch} **do**
    for N_{iter} **do**
      for N_D **do**
        sample a batch of input patches $\{x^i\}_{i=1}^m$
        sample a batch of ground truth patches $\{y^i\}_{i=1}^m$
        update the discriminator by descending its stochastic gradient:
        $\nabla_\omega[-\frac{1}{m}\sum_{i=1}^m log(D(y_i)) + log(1 - D(G(x_i)))]$
      end for
      sample a batch of input patches $\{x^i\}_{i=1}^m$
      sample a batch of ground truth patches $\{y^i\}_{i=1}^m$
      update the generator by descending its stochastic gradient:
      $\nabla_\theta[\frac{\lambda_1}{m}\sum_{i=1}^m (y_i - G(x_i))^2 \ + \ \frac{\lambda_2}{2m}\sum_{i=1}^m [log(D(G(x_i))) - log(1 - D(G(x_i)))]^2]$
    end for
    calculate average SSIM index over the validation set.
  end for
  select the model with highest SSIM index for test.

focal points were collected, while the probe was held with a mechanical arm to prevent any probe movement during changing the transmit focus point. This ensured that images with different focal depths were collected at the same location. Our setup is shown in Fig. 2.4. Although more images can be collected from a phantom, only independent images are of significance in training process, and repeated similar images from the same location do not help the generalization ability of the network.

2.3.1.3 *Ex vivo* data

These images were collected from a fresh lamb liver. Imaging parameters are the same as phantom experiments. Instead of placing the liver in a gel phantom to minimize its motion during data collection, which may lead to some loss of blood and other tissue changes, we placed the liver on a plate and wrote a script in Python (which is the Alpinion interface for using the machine in research mode) to collect three consecutive images at different focal points without altering other imaging settings. This allowed us to collect consecutive images at a very high framerate (i.e., more than 50 frame per second) to minimize the chance of misalignment between images. In addition, we attemped to hold the probe steady during data collection. These steps lead to a collection of images at different focusing depths with minimal relative motion between the probe and tissue. To have independent data points, we repeated the experiment five times by collecting images from different locations of the lamb liver.

2.3.2 Evaluation setting

For evaluation, we placed three real equispaced focal points in the axial direction of the ultrasound images, and blended the resulting three images by weighted spatial linear averaging as in commercial ultrasound machines. As such, the multi-focus image (desired) has 3 layers with 2 blended regions (Fig. 2.5 (b)). One of the images (Fig. 2.5 (a)) with the middle focal point is the input of our model. Therefore, the middle layer of the output (Fig. 2.5 (c-f)) is equal to the input, and two other layers are estimated from related layers of input through two BSGANs. Each layer was broken into 52×52 patches and fed to the network. During the test phase, we did not break the image, and each layer was fed to the generator to prevent the blocking artifact. For quantitative analysis, we tried to compare the results of the proposed method in terms of all image quality metrics. General metrics including peak signal to noise ratio (PSNR), normalized root mean square error (NRMSE), and structural similarity (SSIM) index were calculated between ground truth and both of the output of proposed network and input. Additionally, MOS test was performed to show which form of network and which type of training is more successful in recovery of perceptually convincing images. Monte Carlo simulation was performed on *ex vivo* data to investigate the ability of the proposed method on recovering the sharpness of images in terms of mean gradient (MG) index. Afterward, the proposed method, using specialized ultrasound assessment indexes including contrast to noise ratio (CNR) and full width at half maximum (FWHM), is compared with other approaches which are not based on NNs. The next subsection describes details of the MOS comparison.

2.3.3 Mean opinion score (MOS) testing

As common indices for image quality assessment have a limited potential to indicate how much an image is perceptually convincing, we performed an MOS test to improve the validation step. More specifically, 20 graduate students who work in the field of ultrasound imaging, as raters, were asked to assign a score from 1 (bad quality) to 5 (excellent quality) to images. 6 versions of simulated phantom image (Fig. 2.5 (a-f)) were rated. Images were presented in a randomized fashion to raters. Raters very consistently rated ground truth image as excellent quality and the original input image (with only single focal point) as bad quality. Moreover, we put two identical images in questionnaire to make sure that answers are reliable. The summary of all results is reported in Table 2.2.

2.3.4 Network training

The entire database was broken into three sets of training, validation, and test groups with sizes of 70, 15, and 15 percent of the total size of images, respectively. We first normalized the intensity input ultrasound images to [-1,1]. As it is common in training GANS [23], in each iteration, the discriminator is trained 3 times (N_D), and the generator one time. In all experiments, the Adam algorithm with learning rate ($\alpha = 10^{-4}$) was used for optimization [65]. The training procedure of the proposed BSGAN is shown in Algorithm 2.1. The code is implemented using Keras library using TensorFlow back-end, and training was done with an Nvidia Titan Xp GPU.
The solution to training a BSGAN network (which is a game between two players) is a Nash equilibrium. In fact, by having the optimal discriminator, the global minimum of generator's loss function is achieved if and only if $p_g = p_{data}$, which means that the discriminator gives the same probability of 0.5 to both generated and real data. Although the two players may suddenly reach an equilibrium, the training process oscillates between two modes and players repeatedly undo each other. In fact, as we never reach the perfect case (in which $p_g = p_{data}$), after finishing training process for a specific number of epochs, the model which has the best structural similarity to desired

on validation dataset is chosen as final model. The final model of training is saved and applied to the test set.

For real data (i.e., real phantom, *ex vivo* experiments), we used transfer learning to fine tune the networks trained on simulated data. Transfer learning was done in the same adversarial manner as before and used for fine tuning the weights of whole of the layers in generator and discriminator. More specifically, weights of the best network on simulated phantom data was used as initial point of training on new data. As before, model selection was done based on SSIM evaluation on validation data. Finally, selected generator was used for test part.

2.4 Results

2.4.1 Experimental methods

The first network used in comparison is a well-known structure named super-resolution CNN (SRCNN), a relatively shallow network with 3 layers without any residual connections, about which details can be found in [66]. The second, entitled residually connected fully CNN (RFCNN), is the generator in our proposed structure shown in Fig. 2.3, which is deeper and also has residual connections compared to SRCNN. Consequently, proposed RFCNN is used in a basic non-saturating GAN structure [23]. Finally, the basic GAN is extended to boundary seeking version. It has to be mentioned that non-GAN networks are only trained with MSE loss function.

2.4.2 Comparison on simulated phantom

In the first experiment, the performance of different networks is evaluated on the simulated phantom data. As can be seen in Fig. 2.5, both SRCNN and RFCNN do not perform very well and lead to over smooth images. The main reason for the loss of fine details is that the network is trained with only MSE as the loss function. In the GAN structure, however, the image quality is enforced indirectly by the discriminator in adversarial training, as the generator network tries to make images that look like real images. Between the basic GAN structure and the boundary seeking version, the latter works better because the training process of BSGAN is more stable and the discriminator is nearer to the optimal case. As can be seen in Fig. 2.5, the GAN result has some artifacts in the middle part of the cyst region. Furthermore, the GAN result has more contrast, but even more than the ground truth (b). So, as it has been shown in [37] for general images, GAN results are more artificial while results of BSGAN are more natural.

In order to provide better comparison among different methods, quantitative results are presented in Table 2.2 in which the input is a single focus image, and the desired output is a multi-focus image. As mentioned, common indices do not illustrate how much an image is perceptually convincing as these metrics are not developed for ultrasound images. Therefore, SRCNN and RFCNN have better values on some of those general indices because their results are very smooth.

Table 2.2: The results of PSNR, NRMSE, SSIM, and MOS between input-desired and output-desired pairs. The best values (highest mean and lowest std) are in bold font.

data	input				SRCNN				RFCNN				GAN				BSGAN			
index	PSNR	NRMSE	SSIM	MOS	PSNR	NRMSE	SSIM	MOS	PSNR	NRMSE	SSIM	MOS	PSNR	NRMSE	SSIM	MOS	PSNR	NRMSE	SSIM	MOS
mean	23.27	0.034	0.622	1	26.46	0.023	0.782	3.15	26.78	0.023	0.794	3.15	24.69	0.029	0.773	3.92	25.32	0.027	0.769	**4.07**
std	1	0.004	0.02	0	0.95	0.002	0.018	0.688	0.932	0.002	0.016	0.89	0.795	0.002	0.01	0.64	0.919	0.003	0.017	**0.49**
min	20.77	0.026	0.574	1	22.98	0.019	0.729	2	23.44	0.018	0.74	2	22.71	0.022	0.725	3	22.9	0.021	0.723	**3**
max	25.62	0.045	0.684	1	28.29	0.035	0.824	4	28.56	0.033	0.826	5	26.9	0.036	0.803	5	27.38	0.035	0.797	**5**
median	23.16	0.034	0.621	1	26.57	0.023	0.784	3	26.92	0.022	0.796	3	24.77	0.028	0.778	4	25.46	0.026	0.775	**4**

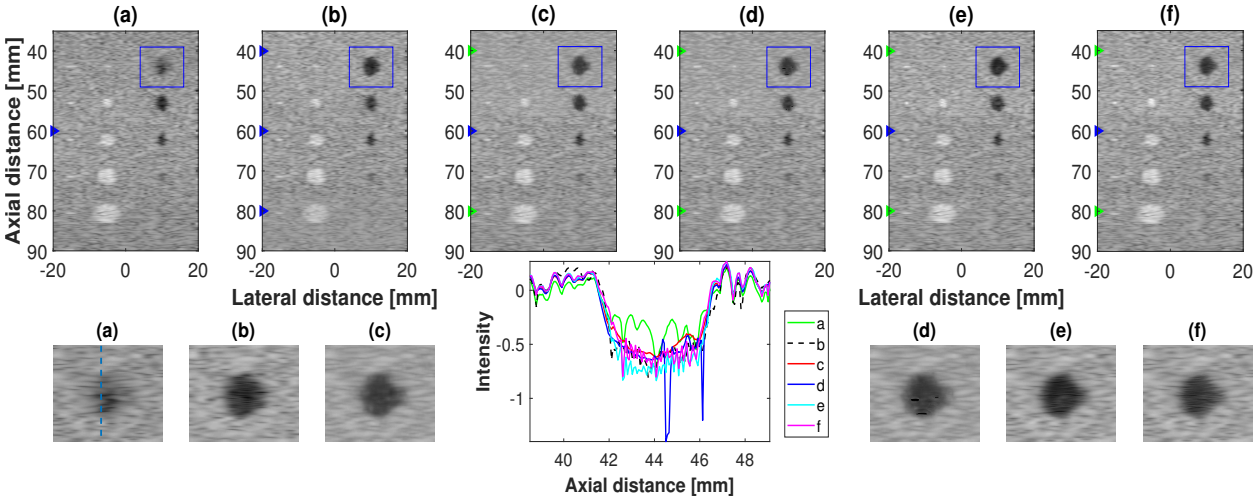

Figure 2.5: Results of the different methods on the simulated phantom data. Blue triangles indicate real transmit focal points, and green triangles indicate focal points added by the network. (a) Input image with a single focal point. (b) Desired image with 3 focal points. (c) Output of the SRCNN (d) RFCNN (e) GAN (f) the proposed BSGAN. The second row shows a zoomed in view of the blue rectangle in the first row, and edge spread function of different methods across the vertical line shown in zoomed view of (a) is in the middle.

However, their poor quality is uncovered by the expert raters, and GAN-based networks get much better scores. Comparing basic GAN and BSGAN, the second one has better results with lower standard deviation.

The second question that should be answered is about the selection of the input. In fact, we want to know whether the proposed method depends on the place of focus point in the input or not. To this end, we ran the algorithm for different scenarios. Results showed that the best selection is when the input image is focused at the middle position of axial direction, as expected. More specifically, we found that when the single focus point is in the middle part of the image, the similarity with multi-focus image is the highest value. So, this input is the most informative one and the mapping function from the input space to the output space is more straightforward.

Please note that the adversarial loss function does not reveal useful information in training GANs, and, as such, is not presented in this thesis. To check whether the training has converged or not, generating a few samples and looking at them during the training phase is instead commonly performed [29, 67].

2.4.3 Real phantom results

The proposed method was also validated on real phantom data. As for real phantom data, whenever the focus point was set on first or second axial layer, image of the last layer had a very low quality. Consequently, it is understood as noise by the network, and discriminator gives the probability of 0.5 to it which means that the discriminator is uncertain whether it is real or generated data. For real phantom data, therefore, the image focused on third axial layer was used as input and two other layers were estimated using BSGANs although this was not the best scenario as discussed in last subsection. Fig. 2.6 shows the result of different methods on test data, which depicts the sharp borders of cysts as well as the hyperechoic regions are preserved only in the output of the proposed method as the desired image. It can be easily understood that the proposed method outperforms

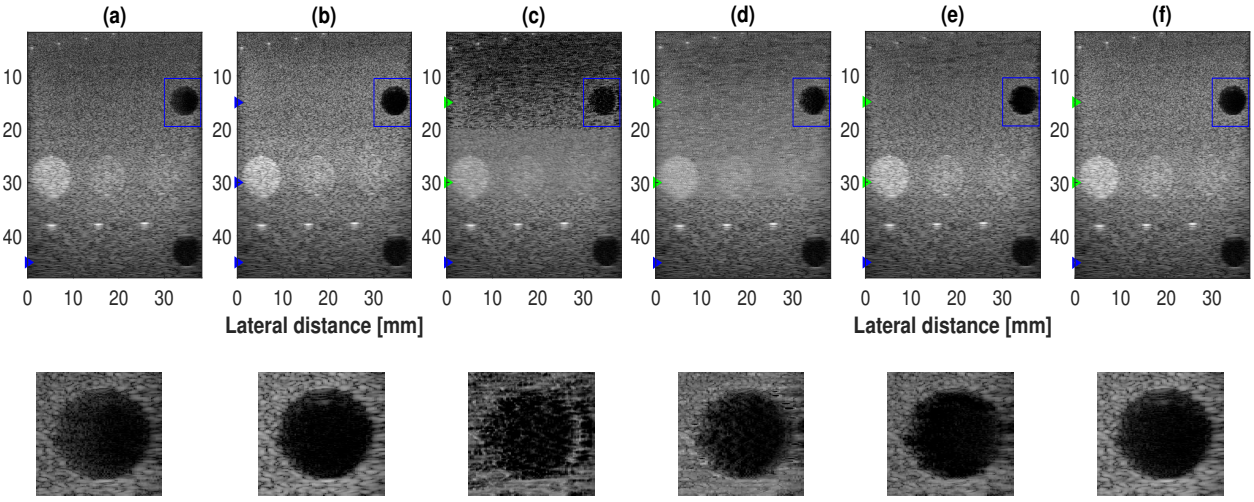

Figure 2.6: Results of the different methods on real phantom data. Blue triangles indicate real transmit focal points, and green triangles indicate focal points added by the network. (a) Input image with a single focal point. (b) Desired image with 3 focal points. (c) Output of the SRCNN (d) RFCNN (e) GAN (f) the proposed BSGAN. The second row shows a zoomed in view of the blue rectangle in the first row.

other approaches noticeably.

As for fine tuning using the real phantom experiment, the number of images is limited compared to simulation data. To reduce the risk of overfitting, two common approaches of training the weights of a specific layer or training for few epochs are commonly used [68, 69]. We chose the latter. In this way, we multiplied the learning rate with 0.1 and limited the number of epochs to 10. This ensures that weights only change slightly.

2.4.4 *Ex vivo* results

In real tissues, there are two main limitations preventing the multi-focus desired image to have a noticeable difference compared to single-focus input. First, there is no specific cyst or hyperechoic region in the tissue which makes the comparison more difficult to clearly visualize the improvement in the image quality. Second, the depth of imaging is limited which means the amount of degradation in image quality, because of beam divergence, may be difficult to notice.

Based on aforementioned reasons, Monte Carlo simulation is used to better investigate the performance of the proposed method on *ex vivo* data. More specifically, a PSF is convolved with the image to simulate large imaging PSF away from the focal point. As mentioned in Section 2.2.1, we assume that changes within each axial layer is negligible and for each axial layer one GAN is trained. The standard deviation (STD) of the Gaussian PSF is the parameter which specifies the level of blurriness and is composed of two deterministic and random parts as follows:

$$c = c_{det} + c_{rand}, \tag{2.10}$$

where c indicates STD of the Gaussian PSF. The deterministic part of STD (c_{det}) specifies a minimum level of blurriness which is set to 1. A positive random number taken from $\mathcal{N}(0, \sigma^2)$ is used as the random component of STD (c_{rand}). The random part is added to the deterministic component to specify the level of blurriness in each run. Consequently, Monte Carlo simulation is done for 10 different values of σ. For each value of σ, 100 runs are performed. Fig. 2.7 illustrates

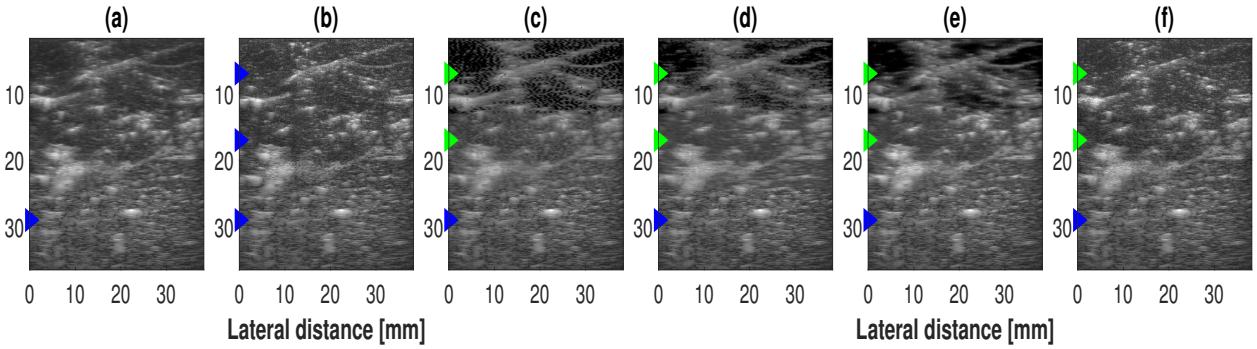

Figure 2.7: Results of the different methods on *ex vivo* data. Blue triangles indicate real transmit focal points, and green triangles indicate focal points added by the network.(a) Input image with a single focal point. (b) Desired image with 3 focal points. (c) Output of the SRCNN (d) RFCNN (e) GAN (f) the proposed BSGAN.

the results for *ex vivo* data. Fig. 2.7 is shown after convolving with a Gaussian PSF having a STD of 8. The blurring is not applied on the correct focused layer because there is no modification on that. As can be seen in Fig. 2.7, the proposed method successfully recovered the multi-focus image, very similar to the ground truth, while other methods failed to recover fine details from the blurry input. Fig. 2.8 summarizes the observed changes in image quality as the STD of the simulated Gaussian PSF is increased. More specifically, Fig. 2.8 illustrates the box plot of image quality indices obtained from a Monte Carlo simulation comprising of 100 runs for each value of σ. We want to make sure that the proposed method preserves its performance over a wide range of simulated blurriness. As shown in the first row of Fig. 2.8, the SSIM index between the blurred input image and desired multi-focus image rises as the amount of blurriness (c) increases. Therefore, other indices, such as the MG index, which reflects the sharpness and texture changes of the image should be used. As observed in the second row of Fig. 2.8, the output of the proposed method is substantially sharper than the input and much closer to the desired image for all levels of blurriness that we tested.

2.4.5 Comparison with other methods

In this subsection, the result of the proposed method is compared with other multi-focal methods which are not based on NNs. As reviewed in section 2.1, two multi-focal methods were proposed before us. Comparison with the method proposed by Ilovitsh *et al.* [57] was not possible for us because of two main reasons. More specifically, their method is based on the summation of electrical stimulation corresponding to different focused transmissions. So, one multi-focal beam which has a longer duration of time is transmitted instead of transmitting several single focus beams. However, the summation causes not only cross-talk, but also it can only be completed in a subset of probe crystals because of the piezoelectric maximal element response which causes nonuniform quality in the axial direction. This problem makes the comparison impossible. Moreover, we cannot implement the method on our research machine. However, the method proposed by Bottenus *et al.* named retrospective encoding for conventional ultrasound sequences (REFoCUS) could be applied using a conjugate transpose (REFoCUS adjoint) [54], or a regularized inverse (REFoCUS inverse) [56]], of the transmit encoding matrix at each frequency. Fig. 2.9 shows the results of our comparison based on a simulated phantom data with imaging details exactly the same as for the real experiments. Herein, we used fine-tuned networks on real phantom data. As observed

in Fig. 2.9, the result of proposed method is much more similar to the ground truth while other methods exhibit poor performance at the first row of cysts. In deeper regions, our method gives result similar to the ground truth while other methods are even better than the ground truth. Our method which is based on NNs does not require either channel data or any sort of matrix inversion, which is worth noting for practical applications since improved inversion matrix techniques are computationally expensive and time consuming while NNs in test case are on-line.

2.4.6 Ultrasound image quality metrics

The importance of using the adversarial loss function (GAN structure) as well as the boundary seeking method of training compared to other cases is demonstrated in subsection 2.4.1. The last subsection of the results is dedicated to assess proposed methods in terms of specialized ultrasound assessment indexes [70]. To this goal, the CNR [70] and the FWHM indices are calculated. As our method is proposed to preserve the lateral resolution, we only calculate the FWHM in the lateral direction. More specifically, for the simulation experiment, the point spread function of the imaging system is simulated by placing point targets on different focus points along the axial direction. Consequently, the FWHM is calculated for the input single-focus image, ground truth, and the results of the REFoCUS method. For real experiments, the calculation is performed using the point targets in the real phantom as shown in Fig. 2.4. It has to be mentioned that the comparison with the REFoCUS method only was possible for the simulation data. Moreover, we did not have the cyst region on the second axial layer of the real phantom, so, the CNR is only reported for the first layer. As it can be seen from Table 2.3, the REFoCUS method provides better resolution in terms of FWHM while the proposed method has a better performance in terms of contrast. However, the lower FWHM (narrower main lobe) value for the REFoCUS method comes

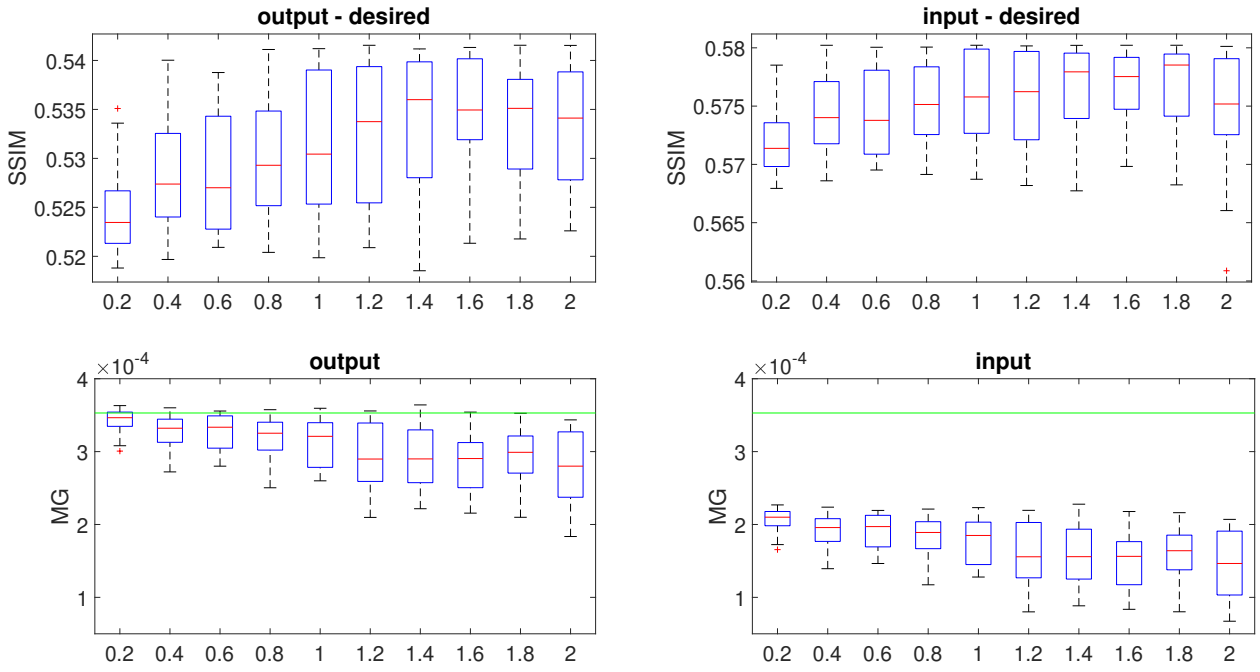

Figure 2.8: Results of Monte Carlo simulation. First row contains the box plot of SSIM versus σ, and second row illustrates the box plot of MG versus σ. Green line shows the MG of desired image.

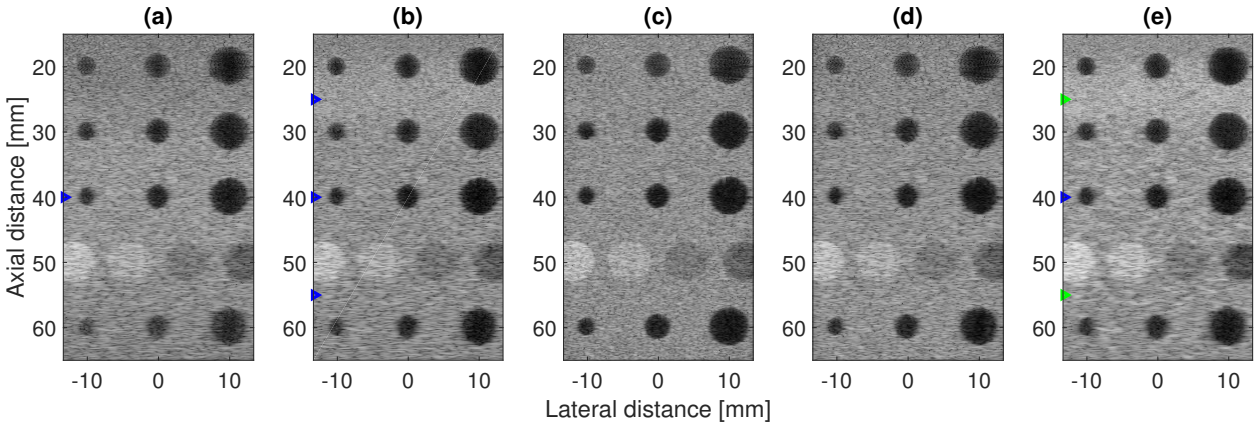

Figure 2.9: Comparison of results with other methods. Blue triangles indicate real transmit focal points, and green triangles indicate focal points added by the network.(a) Input image with a single focal point. (b) Desired image with 3 focal points. (c) Output of REFoCUS inverse (d) REFoCUS adjoint (e) the proposed BSGAN.

at the expense of worse side lobes. Table 2.3 also confirms the improvement of indexes for the real phantom experiment.

2.5 Discussion

The results have illustrated that the proposed method based on BSGAN noticeably outperforms our implementations of SRCNN, RFCNN and GAN learning structures. Having residual connections in the fully convolutional generator network provides better performance because it is easier to learn the difference between the input and output [62]. The necessity of using adversarial objective function in training besides basic MSE loss function for having a sharper image, which is more perceptually convincing, is rather significant. Moreover, using the boundary seeking method for training provides a policy gradient for training the generator, and generates samples near the decision boundary. This ultimately leads to improved stability in training.

The proposed method was also tested on real applications. Transfer learning was successfully performed from the simulated space to real space. In order to provide desired training data in real experiments, two different settings were used. First, we used a mechanical arm to prevent any probe movement during changing the transmit focus point. Second, we wrote a data collection script in Python to alleviate the problem of unavoidable movements during data collection. The network also works well in real experiments. Comprehensive qualitative and quantitative results with Monte Carlo simulations also verified the higher quality of recovered multi-focus images as compared to the single focus inputs.

Table 2.3: The results of CNR and FWHM indexes for simulation and real phantom experiments.

method		input		desired		BSGAN		ReFOCUS (adjoint)		ReFOCUS (inverse)	
index		FWHM	CNR	FWHM	CNR	FWHM	CNR	FWHM	CNR	FWHM	CNR
simulation	layer 1	1.3	7.2	1.01	8.32	1.09	8.02	1.15	7.2	1.04	7.56
	layer 3	2.13	6.29	1.88	7.3	1.95	6.95	1.37	6.9	.94	7.7
real phantom	layer 1	1.52	9.6	1.37	11.7	1.44	11.1	-	-	-	-
	layer 2	.88	-	.74	-	.83	-	-	-	-	-

The results of comparison with other non-NN based method (REFoCUS) show that the proposed method achieves similar results while it is faster and does not require matrix inversions. Moreover, a comparison on specialized ultrasound assessment indexes shows that the proposed method is able to simultaneously improve both resolution and contrast. Moreover, it is possible to combine the ideas in the proposed method and REFoCUS to further improve image quality, which is an interesting avenue for future work. For example, the output of REFoCUS method can be considered as the ground truth in the training step.

Currently, most of the commercial scanners are still using line-per-line acquisition, and plane-wave imaging is prohibitively expensive for affordable point-of-care ultrasound scanners. Therefore, most future scanner designs are likely to rely on line-per-line acquisition technique. For example, several next generation pocket-size ultrasound machines such as Extend R2 (GE Vscan), Sonon (Healcerion, USA) and Clarius all cost less than \$5K. In comparison, only the data acquisition board for plane-wave imaging usually costs approximately \$10K. In addition, plane-wave imaging also requires expensive computations using high-end GPUs. The proposed method in this manuscript requires a GPU for training and can be easily implemented on a CPU in the test phase making it a cost-effective choice for the next generation pocket-size ultrasound machines.

As the proposed method works on B-Mode images, its application to Doppler imaging and motion estimation algorithms, which are based on RF data, is limited. Moreover, an important issue in using machine learning methods for different medical image processing tasks, such as image synthesis, denoising and image reconstruction, is the reliability of generated results for the sake of diagnosis and surgical planning and guidance. In other words, are these results misleading or helpful for clinicians? In future, we plan to extend the proposed method to work on pre-beamformed RF data and test the performance of the proposed method in in-vivo applications and perform MOS tests with radiologists. In addition, we will explore training conditional GAN structures to be able to recover ultrasound images with a specific amount of reliability.

2.6 Conclusions

A reduction in the framerate and motion blurs are the main challenges associated with multi-focus line-per-line imaging technique. Herein, the proposed alternative works as a nonlinear mapping function from the input space (ultrasound image with one transmitted focused beam) to the optimum multi-focus output space. As shown above, GANs with boundary seeking method of training have been adapted to achieve the quality of multi-focus ultrasound images without any loss in framerate or appreciable drop in quality due to probe movement. The experiments confirm that the proposed approach provides perceptually convincing images with a higher resolution and contrast, while it is computationally efficient and does not require channel data. The proposed approach can potentially be used in applications that require both high framerate and lateral resolution.

Chapter 3

Ultrasound Beamforming using MobileNetV2

This chapter is based on our published papers [71] and [72]. It is noteworthy that we submitted our trained beamforming model to the Challenge on Ultrasound Beamforming with Deep Learning (CUBDL), held in conjunction with IEEE IUS 2020. In terms of image quality, our method ranked first and performed excellently on unseen *in vivo* test data as evaluated by the challenge organizers.

Minimum variance beamforming (MVB) is one of the best adaptive methods that performs well regardless of the imaging settings [73]. MVB, however, is computationally very expensive mainly because of the covariance matrix estimation step. Therefore, speeding up MVB is of crucial importance to make it applicable online [74, 75, 76].

In the past few years, deep learning (DL) has disrupted several low-level medical imaging tasks such as reconstruction of computed tomography (CT) and magnetic resonance (MR) images. DL has also been proposed for ultrasound image reconstruction [58, 77, 78, 79, 20]. There are a variety of approaches to accomplish this task. More specifically, DL can be designed to complete a single, few, or all of the reconstruction steps. Another advantage is that DL can simultaneously fulfill another objective such as speckle reduction or super-resolution in the reconstruction process. Nevertheless, the design of ultrasound image reconstruction using deep learning entails several challenges. The scarcity of training data as well as lack of ground truth are among the main limitations. Moreover, the changes in imaging settings cause a large domain shift in the high dimensional input space of DL, limiting its generalization.

Herein, we strive to address all of the aforementioned challenges. In essence, the proposed approach is designed to approximate MVB. Since MVB can be summarized in a set of closed-form mathematical steps, we can calculate the desired output for any input. As such, we do not have the problem of domain shift or lack of ground truth. Furthermore, the proposed method does not need many training images since MVB works on each pixel separately meaning that each pixel is a sample in the training process. We also consider the fact that all mathematical transformations, including DL, cannot generate new information that is not present in the input data. Therefore, necessary preprocessing steps are applied to raw RF channel data before feeding it to the network, and the network input contains all required information for estimating the result of MVB.

The current chapter is our submission to challenge on ultrasound beamforming with deep

learning (CUBDL) announced by 2020 IEEE international ultrasonics symposium (IUS) [80, 81, 82]. In terms of image quality, our method was ranked first. Overall, considering both image quality and network size, our method was jointly ranked first with another submission [83].

3.1 Minimum variance beamforming

Let us assume an ultrasound array that transmits a pulse into the medium with a sound speed of c. Without loss of generality, consider n elements record the backscattered signals denoted by $\mathbf{h}_i(t)$. The transmission distance between the origin of the transmitted pulse to an arbitrary pixel (x, z) in the region-of-interest (ROI) is denoted by d_t. Likewise, d_r is defined as the receiving distance from (x, z) to the location of element i. The radio-frequency (RF) data corresponding to (x, z) in $\mathbf{h}_i(t)$ can be determined using the propagation delay as follows (hereafter, capital and bold font variables represent matrices and vectors, respectively):

$$\tau(x,z) = \frac{d_t + d_r}{c} \Longrightarrow S_i(x,z) = \mathbf{h}_i(t) \mid_{t=\tau(x,z)}, \tag{3.1}$$

where matrix S_i contains the RF data recorded by crystal element i corresponding to ROI pixels. The final RF image is obtained using a weighted summation of all receiving elements as follows:

$$S(x,z) = \sum_{i=0}^{n-1} \mathbf{w}_i(x,z) S_i(x,z), \tag{3.2}$$

where $\mathbf{w}$ refers to the apodization window of length n. In practice, the number of active crystal elements considered for the reconstruction of each depth is determined using F-number. Moreover, S is subject to envelope detection and log compression for obtaining the final B-Mode ultrasound image.

In Capon's MVB, $\mathbf{w}$ is adaptively estimated such that the output variance is minimized while the unity gain is preserved in the steering direction [73]. The final solution of MVB is as follows [73]:

$$\mathbf{w}_{MV} = \frac{R^{-1}\mathbf{a}}{\mathbf{a}^H R^{-1} \mathbf{a}}, \tag{3.3}$$

where R is the spatial covariance matrix. For delayed signals, the steering vector $\mathbf{a} = 1$. The robustness in estimating R is increased with temporal averaging over $2k + 1$ samples followed by another averaging over subarrays of length l as follows [73]:

$$\widetilde{R}(x,z) = \frac{\sum_{j=-k}^{k} \sum_{i=0}^{n-l} \overline{\mathbf{r}}_i(x, z-j) \overline{\mathbf{r}}_i^H(x, z-j)}{(2k+1)(n-l+1)}, \tag{3.4}$$

where:

$$\overline{\mathbf{r}}_i(x,z) = \Big[S_i(x,z), S_{i+1}(x,z), ..., S_{i+l-1}(x,z) \Big]^T. \tag{3.5}$$

A diagonal loading factor is added to the covariance matrix for numerical stability by $\widehat{R}(x,z) = \widetilde{R}(x,z) + \epsilon I$, where I is the identity matrix and:

$$\epsilon = \frac{\Delta}{l} trace(\widetilde{R}(x,z)). \tag{3.6}$$

The result of subarray averaging is a vector of length l. Finally, each pixel (x, z) of S, using MVB, can be computed as follows:

$$S_{MV}(x,z) = \frac{1}{n-l+1} \sum_{i=0}^{n-l} \mathbf{w}_{MV}^H \overline{\mathbf{r}}_i. \tag{3.7}$$

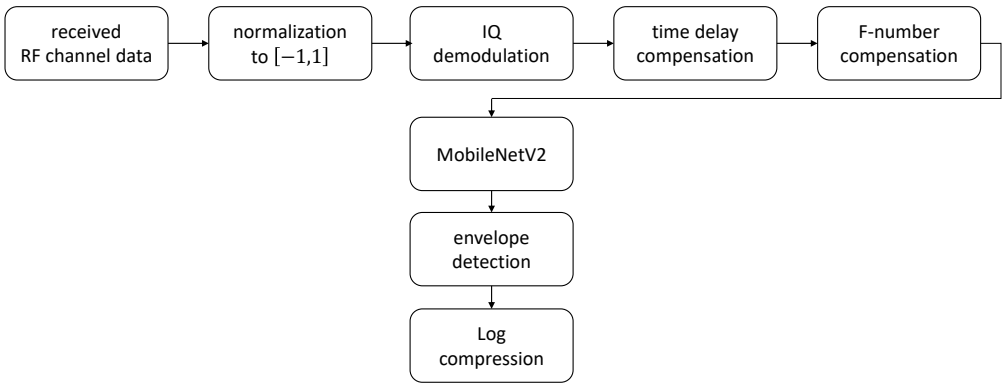

Figure 3.1: Diagram of the proposed method.

3.2 The proposed method

The proposed receive beamforming approach can be summarized in a few steps presented in Fig. 3.1. First, the network's input is supposed to be within the $\left[-1, 1\right]$ range otherwise RF channel data has to be scaled proportionally. Second, in-phase/quadrature (IQ) demodulation is applied to the RF channel data since MVB requires complex signals to compute complex weights allowing for beam patterns that are asymmetrical around the center of the beam. Third, time delays are compensated to reduce the load on the network. Finally, the F-number is fixed for all image depths in order to make the image quality uniform. Each pixel of the image is reconstructed separately as is the case for MVB. The network's input is a $2 \times m \times n$ matrix in which first the two channels are the real and imaginary parts of IQ data, n is the number of channels and m is the length of the window considered for temporal averaging to preserve the speckle statistics. The network output is a two-dimensional vector containing real and imaginary parts of the beamformed data. The network is designed to estimate the apodization window and apply Eq. (3.7) on the input IQ channel data. After reconstructing the whole output IQ data, it is subjected to envelope detection and log compression in order to obtain the final B-mode ultrasound image.

MobileNetV2 [84] is used as the network structure since it is a leading architecture for networks with low computational complexity and memory requirement. This is of critical importance for commercial success of deep learning beamforming given the very large ultrasound framerate and limited computational resources, especially in mobile ultrasound devices. The MobileNetV2 architecture is based on using depth-wise separable convolution building blocks. Moreover, it contains linear bottlenecks between the layers as well as shortcut connections between the bottlenecks. An overview of the MobileNetv2 architecture is shown in Fig. 3.2. More details regarding MobileNetV2 can be found in [84].

3.2.1 Training

As the trained network has to be able to generalize across the range of parameters provided in the CUBDL data guide, the network is trained with a variety of imaging settings such as the acquisition center frequency, sampling frequency, transducer shape, and number of transducer elements. More specifically, the training data contains one image collected with a phased array probe and 10 images collected with linear array probes. Among the second group, 2 images are from focused imaging dataset and 8 are from coherent plane-wave compounding (CPWC) dataset. 2 of CPWC data are collected with an Alpinion scanner (Seoul, South Korea) using a L3-8 probe and the other 6 are

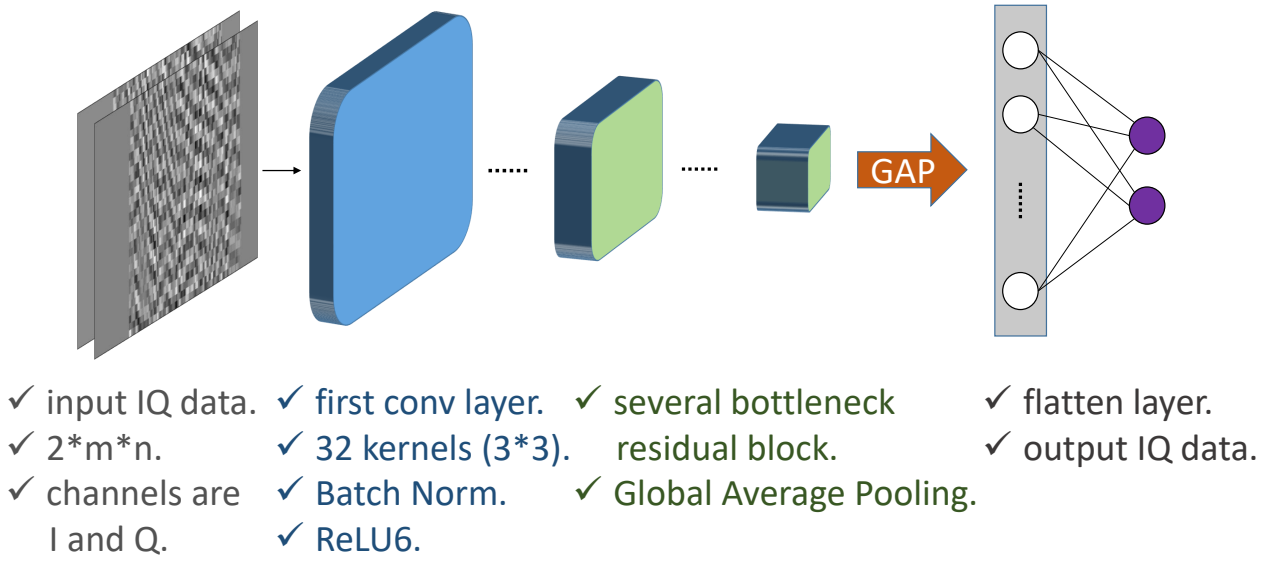

Figure 3.2: Overview of the MobileNetv2 architecture. GAP refers to Global Average Pooling.

collected with Verasonics Vantage 256 platform (Kirkland, WA, USA) and the linear L11-4v probe. All these datasets are publicly available through UltraSound ToolBox (USTB) [85]. The network's output for each image is reconstructed using the MVB code provided by USTB. The number of input channels (n) is different based on the probe specifications, while the length of window for temporal averaging (m) is set to 32 for all datasets.

The model is implemented using PyTorch library. The batch size is 50, and AdamW optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$ is used. The learning rate is linearly decayed from 10^{-3} to 5×10^{-5} during 50 epochs. Fig. 3.3 illustrates the training history of the network. As the network works on each pixel of the image separately, the total size of the input-output pairs is more than 1 million. 80 percent of the data is used for training and 20 percent for validation. Fig. 3.3 confirms that the weights are not overfitted to the training dataset.

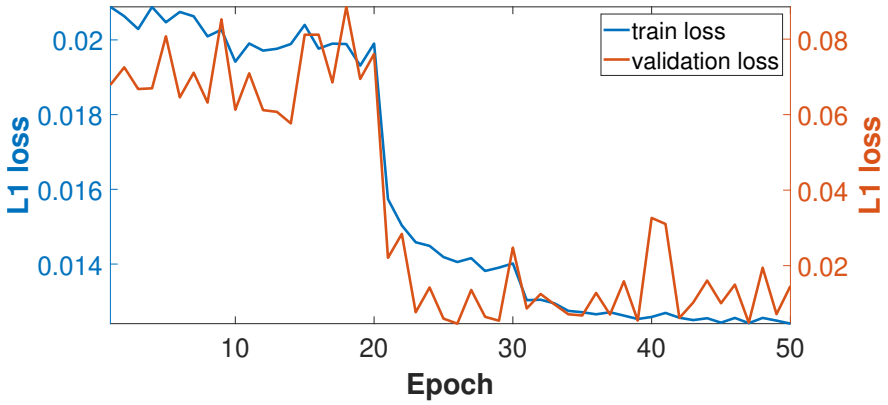

Figure 3.3: Plot of the training and validation losses during training.

3.2.2 Network pruning

Unfortunately, deep models require a large memory and are computationally expensive, which not only increase the hardware costs, but also precludes them from practical application on resource-constrained environments such as the point-of-care ultrasound machines. Neural network pruning is a common approach for solving these issues. Pruning involves the systematic removing of network parameters to produce a smaller model with similar performance [86]. Herein, an approach is used to prune the trained MobileNetV2 to reduce the number of parameters and computational complexity to speed-up beamforming.

A CNN architecture is defined as a function family f, and the trained model is defined as a particular parametrization of f, i.e., $f(x, \theta)$ for specific parameters θ. CNN pruning is the process of taking as input a model $f(x, \theta)$ and generating a smaller model $f(x, \hat{\theta})$ in which $\hat{\theta}$ is a subset of θ. There are numerous methods of pruning among which the method proposed by Han *et al.* [87] is the most popular one. In this general framework, after training a model, each network's parameter is issued a score, and the pruning is completed based on these scores. As pruning reduces the model performance, the resulting model might be subject to further fine-tuning. The explained pruning and fine-tuning process is often iterated to gradually reduce the network size.

In this section, the network parameters, which are kernel weights of convolutional layers, are scored based on their $L1$ norm. We believe that the impact of a kernel on the network's output is proportional to its norm. It is also observed, in our experiments, that reducing the size of convolutional layers closer to the network output causes relatively smaller changes as compared to making the same reduction in initial layers. Therefore, the pruning approach starts from the convolutional layer closest to the network output and works backwards towards the first layer. Our algorithm is implemented iteratively. More specifically, a specific fraction (α) of the kernels with the lowest $L1$ norm is pruned in each iteration. We start with a small value of α and gradually increase it until a reduction in the model's performance is observed. The value of α is upper bounded to a maximum value of 0.7. The same procedure is then successively applied to the earlier layers until the first layer is reached. To stop the iterative pruning process in each layer, specialized ultrasound assessment indexes including FWHM for resolution measurement and generalized contrast to noise ratio (gCNR) as well as contrast ratio (CR) for contrast are calculated.

3.3 Results and Discussion

In order to evaluate the performance of the proposed method, results on the simulation and experimental phantom datasets which are publicly available through USTB [85] are presented. More specifically, both datasets contain one image of point targets for measuring the spatial resolution, and one image of anechoic cysts for measuring the contrast. Fig. 3.4 provides a visual comparison of the original and pruned beamforming networks. As this figure illustrates, there is no discernible reduction in the network performance after pruning, and both pre- and post-pruned models are able to reconstruct images with the same perceivable quality to that of MVB. The quantitative comparison, provided in Table 3.1, also confirms that there are negligible differences in the values of the calculated indexes in terms of both resolution and contrast between two deep models. In terms of memory footprint, the pre- and post-pruned networks contain 2.3 and 0.3 million parameters, respectively (7.67 times smaller with pruning). For the reconstruction of a single image, the run-times of MVB, pre- and post-pruned models, are 4.05, 0.67 and 0.29 minutes, respectively. It should be noted that the deep reconstruction is tested on CPU. A small memory footprint is of critical importance for commercial success of deep learning beamforming given the very high framerate and limited computational resources, especially in mobile ultrasound devices. As mentioned before,

similar to MVB, our method reconstructs each pixel of the image separately which makes it quite slow and not close to real time. Therefore, one avenue for future work is to reconstruct the whole image simultaneously to further speed-up the method for real-time implementations. It has to emphasized that results presented in here correspond to the training/validation step. The trained network is blindly tested on the unseen datasets, and the proposed method was surprisingly generalizable as it improves the quality of images reconstructed with a single insonification angle. Detailed quantitative and qualitative test results are presented, compared, and discussed in the challenge paper [82]. Due to the copyright rules, test results cannot be reproduced in here, and interested readers can find them in [82].

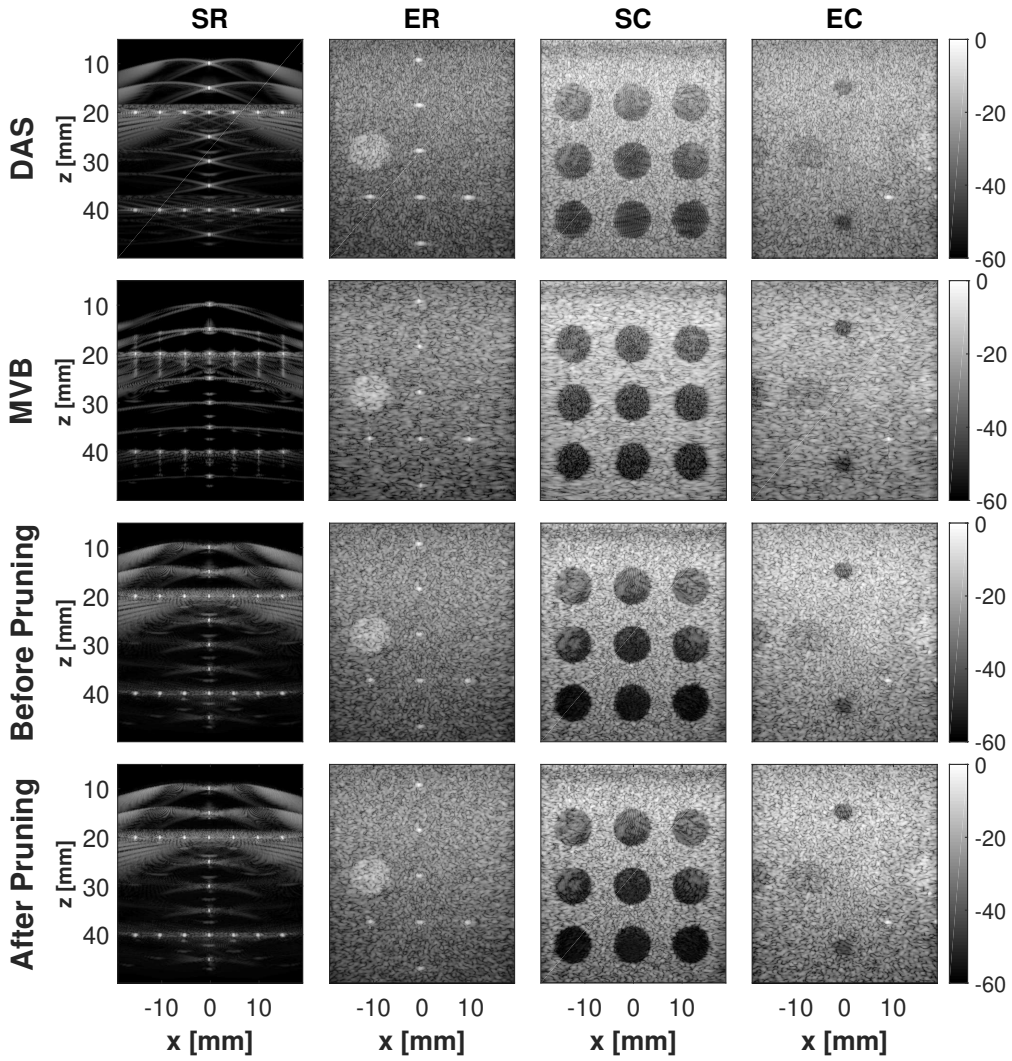

Figure 3.4: Beamforming results on the single 0^o plane-wave. Columns indicate different image datasets while rows correspond to the beamforming approaches. SR and SC refer to simulation resolution and contrast datasets, respectively. ER and EC refer to experimental phantom resolution and contrast datasets, respectively.

Table 3.1: Quantitative results on simulation and experimental phantom datasets in terms of resolution and contrast indexes. SR and SC refer to simulation resolution and contrast datasets, respectively. ER and EC refer to experimental phantom resolution and contrast datasets, respectively. Subscripts $._A$ and $._L$ refer to axial and lateral directions, respectively.

dataset	SR		SC		ER		EC	
index	$FWHM_A$	$FWHM_L$	CR	gCNR	$FWHM_A$	$FWHM_L$	CR	gCNR
DAS	0.4	0.82	-15.15	0.74	0.57	0.88	-13.79	0.57
CPWC	0.39	0.56	-31.44	0.97	0.56	0.56	-25.29	0.87
MV	0.41	0.1	-21.15	0.82	0.59	0.43	-16.74	0.69
MobileNetV2	0.42	0.273	-17.15	0.661	0533	0.773	-15.53	0.55
Pruned network	0.42	0.274	-16.86	0.659	0.531	0.767	-14.59	0.52

3.4 Conclusions

In this chapter, a deep learning framework for ultrasound beamforming has been presented. The proposed approach is based on MobileNetV2 structure and works on IQ channel data to reconstruct each pixel of the final image separately. The training results confirm that deep learning can be used as a general beamformer working on a wide variety of imaging settings. Moreover, the proposed pruning approach reduces the network size by a factor of 7.67 that makes the model 2.3 times faster.

Chapter 4

Deep Reconstruction of High-Quality Ultrasound Images from Raw Plane-Wave Data

This chapter is based on our published paper [88].

Plane-wave imaging (PWI) is of the highest framerate among different ultrasound imaging techniques. In this method, the whole medium is insonified in a single shot using a plane-wave produced by firing all transducer's elements simultaneously. Then, all piezoelectric elements record the backscattered signals from the medium. Therefore, the framerate can reach several thousand frames per second [17] since it is only limited by the depth of imaging and the speed of sound. Each element's output gives a low-quality spatial map of the target echogenicity, and the final image is the result of combining all elements' outputs.

The resulting image, however, is of low quality mainly due to the non-ideal point spread function (PSF) of the imaging system, where the PSF has a wide main-lobe and strong side-lobes as a result of the unfocused transmissions. Fig. 4.1 compares images created with PWI to an image created with a sharp Gaussian PSF as a high-quality ultrasound image. The common approach in PWI provides a low-quality version of the expected tissue reflectivity function (TRF), which is supposed to be a high-quality spatial-map of the target echogenicity (shown in Fig. 4.1(a)). There are several reasons that render the PWI image of low quality such as unfocused transmission, limited frequency response of the piezoelectric elements, and limitation of transducer design. Some of these reasons are intrinsic physical limitations which exist among all ultrasound imaging techniques. Given the very high framerate of PWI, improving its image quality is an active field of research.

Coherent plane-wave compounding (CPWC) is a skillful solution to the problem of reduction in image quality due to the unfocused transmission, which entails coherent compounding of images obtained by several insonifications of different angles [18]. This solution, however, comes at the expense of reduction in framerate, leading to a trade-off between image quality and framerate. In addition, this solution leads to motion blurs in applications such as shear-wave imaging and echocardiography. Therefore, a growing body of research has focused on transcending this limitation [70, 89, 90].

Deep learning has shown an immense potential for different applications in medical imaging such as segmentation, reconstruction and classification [91, 92, 93, 94]. It is an efficient tool for

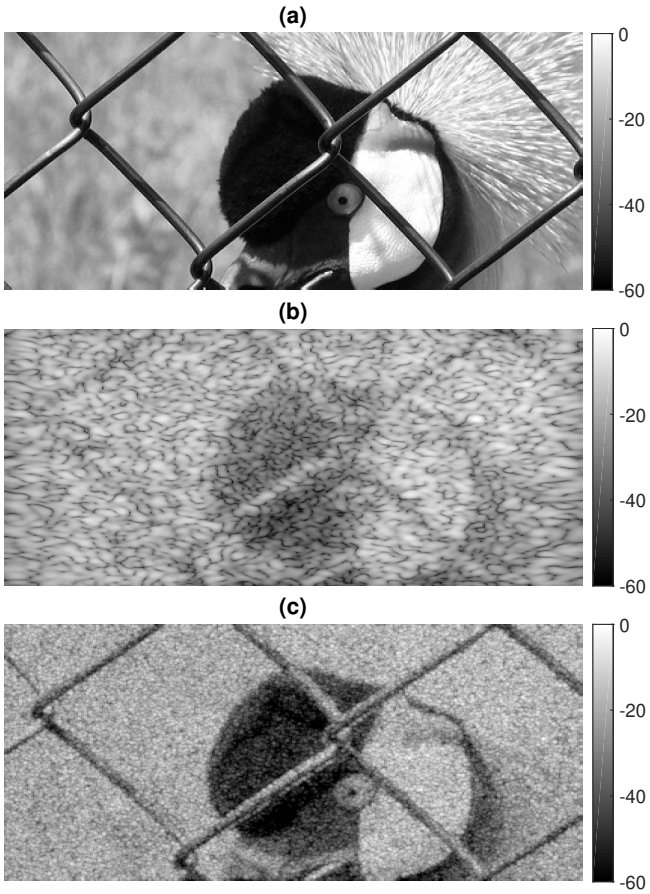

Figure 4.1: Demonstration of the PSF effect on the quality of ultrasound imaging. (a) Ground-truth echogenicity map extracted from a real photographic image from an ImageNet validation set. (b) Simulated ultrasound image from a single 0^o plane-wave transmission. (c) Desired ultrasound image reconstructed using the proposed PSF.

solving ill-posed problems with non-linear mapping functions between high dimensional input-output pairs. Deep learning, however, suffers from three main limitations in medical imaging applications including the scarcity of training data, lack of ground-truth labeled data, and domain shift between training and test data [95].

Considering the aforementioned limitations, deep learning is exploited herein for PWI beamforming. Our goal is to extract a high image quality from the RF channel data. To do so, a high-quality PSF (sharp Gaussian function) is proposed for ultrasound PWI, and the result of convolution between the TRF and the proposed PSF is considered as ground-truth (Fig. 4.1(c)). Then, deep learning is used to construct a mapping function between the RF channel data and the proposed high-quality image.

4.1 Related work

The main focus of the current chapter is the application of deep learning in ultrasound imaging and specifically in image reconstruction and beamforming. Therefore, only related important researches, among a large number of recent literatures, are reviewed. Luchies *et al.* proposed an approach for ultrasound beamforming based on deep neural network (DNN) [46]. More specifically, a distinct

DNN is trained on in-phase/quadrature (IQ) data of each frequency sub-band to suppress off-axis scattering. They used simulated point target responses for training. Subsequently, the authors extended this approach by using simulated and phantom anechoic cysts for training [79]. The reconstruction of B-mode ultrasound images from sub-sampled RF data using deep learning was investigated in [47]. Convolutional neural networks (CNNs) have also been used to reconstruct the image quality of CPWC from a single transmission [50]. Hyun *et al.* made use of fully CNN (FCNN) for reconstructing despeckled ultrasound images [48]. An autoencoder network structure, with a single encoder and two decoders, was used to simultaneously obtain the segmentation map and B-mode ultrasound image from a single plane-wave transmission [96]. Mor *et al.* embedded the beamforming and speckle reduction steps of plane-wave imaging in a deep convolutional neural network [97]. Their beamforming subnet is based on spatial matched filters, and a U-Net architecture is used for speckle reduction [97]. Deep learning was employed as an adaptive ultrasound beamforming technique in [78, 77]. High-quality image reconstruction of diverging-wave ultrasound imaging from a small number of transmissions based on CNNs is proposed in [98]. Youn *et al.* adapted CNNs to detect the scatterers and estimate their positions from RF data [99]. More specifically, the network's output is a non-overlapping Gaussian confidence map of which the position of scatterers is considered as the local maxima. Although there are many differences, this work is similar to our approach on what the PSF is considered.

4.2 Methods

4.2.1 High-quality ultrasound PWI

One of the challenges of using deep learning for ultrasound image reconstruction is the lack of ground-truth. In other words, the best output quality in real *in vivo* images is unknown since TRF is unknown. In simulated ultrasound images, however, this problem can be resolved since a complete knowledge of scatterers' location and amplitude is available. And, researchers use a predefined desired model for the output. For example, in [48], a despeckled ultrasound image was considered as the output of the image reconstruction from raw data.

Herein, we assume that a high-quality ultrasound PWI system produces a sharp Gaussian mask as the resulting image of a point target. Indeed, we consider a high-quality impulse response (i.e., PSF) for the ultrasound imaging system which has a narrow main lobe without any side lobes in the axial and lateral directions. This assumption is held for the result of a single plane-wave transmission. And, it is a very high image quality that an imaging system can achieve.

As for the simulation data, using the proposed PSF as well as a complete knowledge of scatterers' location and amplitude (i.e., having the TRF), the corresponding high-quality image can be acquired using convolution. More specifically, it has been shown that using the first-order Born approximation and assuming weak scattering for soft tissues, the ultrasound images can be modeled as the convolution between a TRF and a PSF [100]. This linear model can be written as follows:

$$S = X \circledast H + N, \tag{4.1}$$

where S is the RF matrix, X is the TRF, H is the spatial impulse response (i.e., the PSF of the imaging system), and N denotes noise. Fortunately, there is no interfering noise in the simulation case. Therefore, our desired simulated image quality (shown in Fig. 4.1(c)) for training is created using this method (Eq. (4.1)).

In this chapter, an isotropic (i.e., circularly symmetric) 2D Gaussian kernel convolved with a 1D cosine function (along the z direction) is used as spatial impulse response H in Eq. (4.1). The

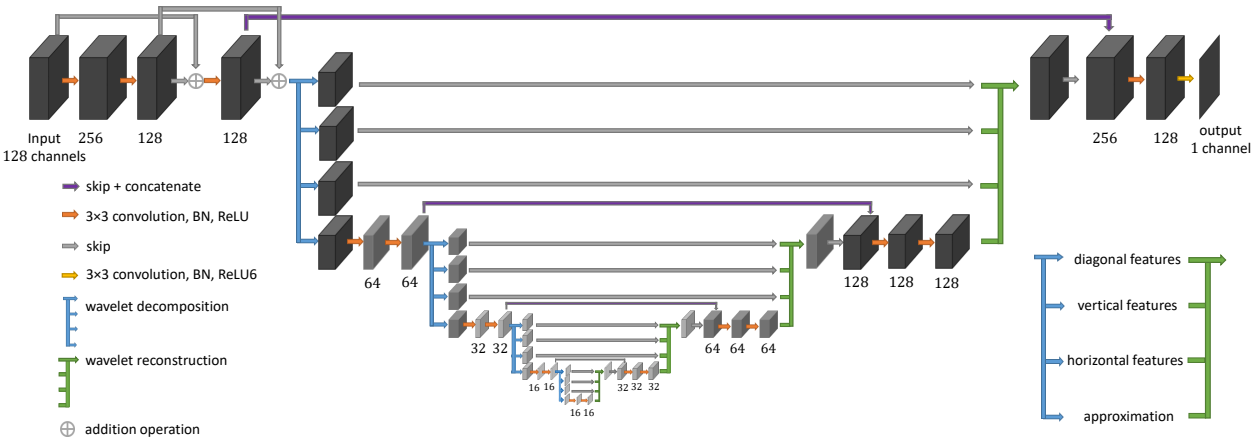

Figure 4.2: The structure of the proposed network.

frequency of cosine function is equal to the center frequency of the probe. It is assumed that H is locally invariant in the axial and lateral directions. The standard deviation of Gaussian kernel is set to 0.1 mm in order to achieve high axial and lateral resolutions, and the maximum projection is used for the overlapping Gaussian kernels.

4.2.2 Network structure

In order to efficiently learn the mapping function between input-output pairs, the tight frame U-Net [101] structure is modified based on the requirements of our beamforming problem. The proposed network is shown in Fig. 4.2. The motivation for the design choices in the proposed structure can be explained as follows. The input data contains all RF matrices of piezoelectric elements reconstructed using Eq. (3.1) without considering any apodization window. This is the rawest possible form of the data that can be used for the proposed beamforming approach in order to make sure that there was not any loss of information in preprocessing steps. As seen in Fig. 4.2, there are 3 convolutional layers with skip connections on top of the input to represent the raw data in the new hidden domain. Intuitively, initial layers extract the proper representation of input data for the following encoder-decoder layers while the skip connections provide an alternative path for the gradient. Next, the network contains encoder and decoder parts implemented using wavelet transform. It has been theoretically shown that pooling and unpooling layers in the original structure of U-Net [102] leads to an overemphasis of the low-frequency components of images due to the duplication of the low-frequency branch [103]. In contrast, using wavelet decomposition and reconstruction helps effective recovery of high-frequency content of the image [101]. Herein, Daubechies mother wavelet with 4 vanishing moments is used for the wavelet transform since it has the most similarity with the collected ultrasound RF signals (used as network's input) as compared to other mother wavelets. On one hand, the encoder part of the network successively represents the context of input data as it mainly focus on the approximation branch of wavelet transform. On the other hand, the decoder part successively reconstructs the output from different levels of input's abstraction acquired in encoder layers. The skip connections of encoder-decoder part is used for concatenating the detail branches of wavelet transform in order to keep the high-frequency information.

All convolutional layers contain kernels of size 3×3 (i.e., in the $x - z$ dimensions). Moreover, Batch Normalization (BN) as well as ReLU activation function are used after each convolutional layer. The depth of encoder-decoder part of the network, as well as the number of filters in each

layer are selected by trial and error to get proper result on the validation dataset using relatively lower number of trainable parameters.

As the input RF data is in the [-1,1] range and ReLU activation function only works on the positive portion of its input, the first convolutional layer contains the double number of filters to ensure that enough potential for successful mapping of data into the hidden feature space is provided. The target data is chosen to be ultrasound envelope data (not log compressed). Therefore, the last layer contains a scaled version of ReLU6 activation function in order to make sure that the network output is in the range [0,1]. In total, the network contains 1.5 million trainable parameters.

4.3 Experiments

4.3.1 Datasets

4.3.1.1 Simulation data

The Field II simulation package [64, 104] is used to simulate ultrasound images. Probe settings (Table 4.1) are set the same as 128 element linear array transducer L11-4v (Verasonics Inc., Redmond, 240 WA), and the plane-wave dataset is simulated using transmissions at 5.208 MHz with 67% bandwidth and the sampling frequency of 104.16 MHz. Scatterers' amplitudes are taken from standard normal distribution (i.e., Gaussian distribution with zero mean and unit standard deviation), and they are uniformly distributed in a $20 \times 45 \times 5$ mm^3 phantom centered at an elevational focus of 2 cm. Average scatterer density is set to 60 per resolution cell in order to simulate fully developed speckle. To have enough variety of ground-truth echogenicity maps in the training dataset, scatterers' amplitudes are weighted according to real photographic images. This approach for simulating phantoms with a wide variety of echogenicity maps was first proposed in [48]. Herein, 512 images are randomly selected from an ImageNet [105] validation set, and 512 images are randomly selected from the Places2 [106] test set. Each image is converted to grayscale and cropped into a 200 pixels $\times$ 450 pixels patch, and then mapped onto the axial and lateral extents of the phantom ($20mm \times 45mm$) to be used as the ground-truth echogenicity map. In other words, the scatterers' amplitudes are first taken from standard normal distribution and the resulting numbers are then weighted based on the pixel intensities of images according to their positions via bilinear interpolation. Each image is simulated using an independent realization of scatterers. Furthermore, to provide the network with images of sparsely distributed scatterers, we have considered a set of 15 predefined number of scatterers as $N = [2, 4, 6, 8, 10, 20, 40, 60, 80, 100, 200, 400, 600, 800, 1000]$. For each element $N(i)(i = 1, \ldots, 15)$, ten different images are simulated in which $N(i)$ scatterers are randomly distributed over an anechoic background. Therefore, in addition to 1024 fully developed

Table 4.1: The settings of linear array transducer L11-4v (Verasonics Inc., Redmond, 240 WA).

Parameter	Value	Unit
Array geometry	Linear	-
Number of elements	128	elements
Element width	0.27	mm
Element height	5	mm
Pitch	0.3	mm
Elevation focus	20	mm
Aperture	38.4	mm

speckle phantom images, we have also 150 ultrasound images of sparsely distributed scatterers. In total, the simulation dataset contains 1174 images.

In order to evaluate the performance of the proposed method, simulation ultrasound dataset available from the plane-wave imaging challenge in medical ultrasound (PICMUS) [70] is used. More specifically, the simulation dataset contains one image containing point targets distributed vertically and horizontally over an anechoic background designed to measure the spatial resolution, and one image containing anechoic cysts distributed vertically and horizontally over fully developed speckle to measure the contrast. It has to be mentioned that our simulation settings are the same as this benchmark dataset.

4.3.1.2 *In vivo* data

In vivo ultrasound dataset, publicly available through PICMUS, is also used for performance evaluation. This dataset contains cross-sectional and longitudinal views of the carotid artery of a volunteer. Plane-wave images were collected using a Verasonics Vantage 256 research scanner and a L11-4v probe (Verasonics Inc., Redmond, 240 WA).

4.3.2 Network training

Only our simulation dataset is used for training, leaving data provided through PICMUS for testing. More specifically, 90% of our images are randomly selected for training and the remaining 10% for validation. And, the same data split is used for trainings of different networks. The network's input consists of 128 channels corresponding to RF matrices of piezoelectric elements reconstructed using Eq. (3.1). Each element gives an RF matrix of size 2688×384, and is broken into two patches of size 1344×384 due to memory limitations during training. The network's output is the corresponding envelope image of the desired output created with the proposed approach outlined in Section 4.2.1. Therefore, we have 2113 input-output pairs for training and 235 for validation.

Let us consider a general form of the loss function defined between the desired output y and the estimated network's output $\widehat{y}$ as follows:

$$\ell_\alpha(y, \widehat{y}) = \frac{1}{m} \sum_{i=1}^{m} \mid y_i - \widehat{y}_i \mid^\alpha, \tag{4.2}$$

where m is the batch size of data in each iteration. The parameter α specifies the shape of loss function. For example, Eq. (4.2) becomes Mean Square Error (MSE) if α is set to 2. Herein, the network is trained in four different steps from coarse (ℓ_2) to fine ($\ell_{0.2}$) loss functions. In each step, the tuning parameter α is changes. More specifically, the training is started with ℓ_2 loss function with a learning rate of 10^{-3} for 25 epochs, which produces blurry results. This result can be considered as a rough approximation of the final output in which fine details are not present. In the second step, the ℓ_1 loss function with a learning rate of 5×10^{-4} is used for 50 epochs. Compared to the previous step, the network is trained to reconstruct more details of the output. Afterward, the network is trained with $\ell_{0.4}$ loss function, with a learning rate of 10^{-4} for 25 epochs. Finally, $\ell_{0.2}$ loss function with a learning rate of 10^{-5} is used for the last 25 epochs. During the last two steps of training, fine details of the output are reconstructed.

As mentioned above, the learning rate is successively decreased in each step to prevent a major distortion on trained parameters. Step-by-step training also helps training error reduces faster as opposed to starting with $\ell_{0.2}$ loss function from scratch. That is because in the initial stage of training, $\ell_{0.2}$ loss is very large. Therefore, the learning rate has to be selected as a very small number which renders training very slow.

In each step, the training procedure is stopped once the error curve becomes flat. The overfitting problem on the training set has not been noticed since the size of training and validation sets are very large. The proposed network is implemented using the PyTorch library and training was done with an Nvidia Titan Xp GPU. The batch size is 2, and AdamW [107] optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$ is used in all steps. As for the wavelet transform in the proposed network, the PyWavelets library is used, which provides the possibility of running wavelet transform on the GPU.

4.3.3 Evaluation metrics

The reconstructed images are evaluated in terms of resolution and contrast. As for resolution, the full-width at half-maximum (FWHM) index of point targets is calculated in both axial and lateral directions. As for the contrast, speckle signal-to-noise ratio (SSNR) [108] as well as contrast ratio (CR) [109] are calculated as follows:

$$SSNR = \frac{\mu_B}{\sigma_B}, \tag{4.3}$$

$$CR = 20log_{10}(\frac{\mu_{ROI}}{\mu_B}), \tag{4.4}$$

where μ_B and μ_{ROI} are the mean of envelope image (before log compression) over the background and the region of interest, respectively. σ_B is the standard deviation of the envelope image over the background region.

The other assessed contrast index is a generalization of contrast-to-noise ratio (CNR) definition, which was recently introduced [110]. It is a robust index of contrast and lesion detectability against dynamic range alterations and called generalized CNR (gCNR). gCNR is calculated as follows:

$$gCNR = 1 - \int_{-\infty}^{\infty} min\left\{p_{ROI}(x), p_B(x)\right\} dx, \tag{4.5}$$

where $p_B(x)$ and $p_{ROI}(x)$ are the histograms of pixels measured in the region of interest and background, respectively. gCNR determines how much the intensity distributions of two regions are overlapped regardless of grayscale intensity transformations. Lower distributions overlap leads to higher gCNR values. gCNR is equal to its maximum value of 1 when the two distributions are independent [110].

4.4 Results

Herein, all the presented results correspond to the test set, which have not been used during training or validation process. More specifically, the trained network, with the best results on the validation set, is evaluated on two test sets of simulation and *in vivo* data taken from PICMUS [70]. It has to be emphasized that the test experiments are completely blind and all results are produced without any further fine-tuning.

The proposed method is implemented using the proposed fully convolutional neural network (FCNN), which is trained based on the method explained in Section 4.3.2. The result of the proposed method is compared with other approaches including DAS beamforming on the single 0° plane-wave as well as 75 plane-waves (the result of CPWC), minimum variance (MV) beamforming [73], and our previously published method, referred to as MobileNetV2 method [71]. Hereafter, we use the boxcar apodization window with $f\# = 1.75$ for DAS and other methods on top of DAS.

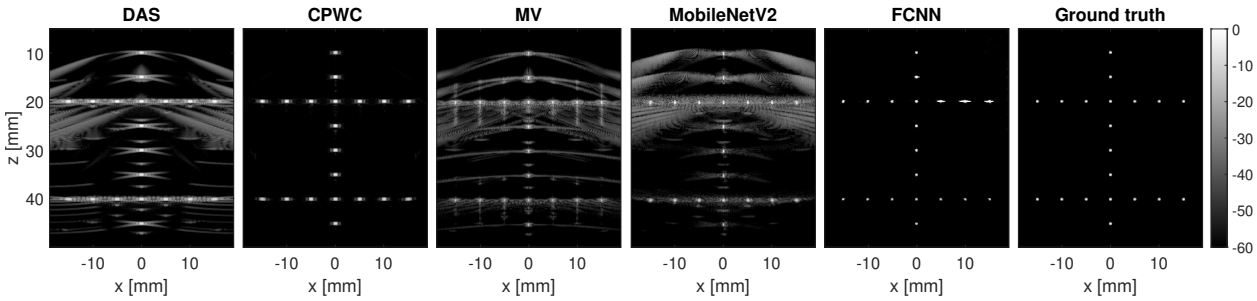

Figure 4.3: Results on simulation resolution data. Point targets are distributed vertically and horizontally over an anechoic background.

4.4.1 Simulation data

4.4.1.1 Resolution distortion

Fig. 4.3 illustrates the results of different methods on the simulation resolution test image. Although the MV beamformer, as well as MobileNetV2, produce high lateral resolutions, they mainly suffer from additional artifacts. In the CPWC method, these additional artifacts are completely removed by averaging among several insonifications at different angles. It is worth noting that the proposed ground-truth (produced with the method explained in Section 4.2.1) offers a high quality without any artifact with excellent lateral and axial resolutions, even better than that of optimal CPWC. The proposed beamforming approach using FCNN produces an image quality similar to the ground-truth. It is important to note that the gain in quality in CPWC comes at the expense of a considerable X 75 drop in the framerate.

Table 4.2 includes a quantitative comparison of different methods in terms of axial and lateral resolution. As mentioned, the proposed FCNN is of worse lateral resolution compared to MV and MobileNetV2. However, as seen in Fig. 4.3, it does not have any side-lobe artifacts. Furthermore, the proposed approach interestingly improves the axial resolution, while other beamforming methods do not have any effect on the axial resolution. It is important to note that while advanced beamforming techniques can be used for improving lateral resolution, options for improving the axial resolution are very limited, with the most common solution as increasing the center frequency. Increasing this frequency comes at the expense of a loss in penetration depth due to the rapidly increasing attenuation with frequency.

Table 4.2: Quantitative results in terms of resolution and contrast indexes for simulation test experiments.

dataset	Simulation Resolution		Simulation Contrast		
index	$FWHM_A$	$FWHM_L$	SSNR	CR	gCNR
DAS	0.4	0.82	1.34	-15.15	0.74
CPWC	0.39	0.56	1.21	-31.44	0.97
MV	0.41	0.1	1.16	-21.15	0.82
MobileNetV2	0.42	0.27	1.05	-17.15	0.66
FCNN	0.25	0.28	2.31	-39.44	0.99
Ground-truth	0.23	0.22	1.8	-47.64	1

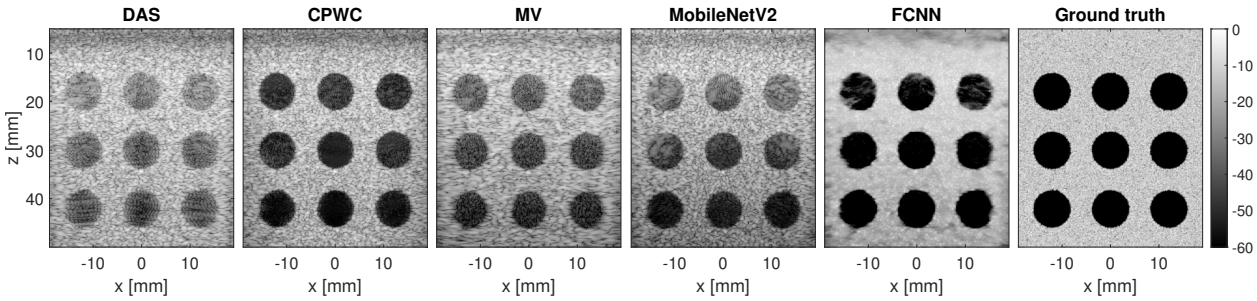

Figure 4.4: Results on simulation contrast speckle data. Anechoic cysts are distributed vertically and horizontally over fully developed speckle.

4.4.1.2 Contrast speckle

Fig. 4.4 shows the output of different methods for the simulation contrast test image. The quality of cyst regions is limited in DAS, MV, and MobileNetV2 methods mainly due to having unfocused transmissions degrading the cyst contrast. This problem is resolved in CPWC, and a noticeable improvement of contrast is visible in the cyst regions. The proposed ground-truth image, however, can be considered as the ideal ultrasound image quality that can be acquired since it gives perfect gCNR (Table 4.2). As shown in Fig. 4.4, the result of the proposed method using FCNN depicts the most similarity to the desired image. This point is also quantitatively confirmed in Table 4.2.

4.4.2 *In vivo* data

The results on the *in vivo* test datasets are shown in Fig. 4.5 and Fig. 4.6. Herein, we do not have the ground-truth anymore. However, we expect the artery to be fully dark, and the brightness within the artery in DAS, MV, and MobileNetV2 is a result of the clutter created by diffuse reverberation from shallow layers [111, 112, 113]. As shown in both cross-sectional and longitudinal views, the proposed method based on FCNN depicts the carotid artery without any clutter artifact. Other approaches including CPWC mainly suffer from the clutter artifact in the carotid artery regions. The *in vivo* results worth noting since there is always a domain shift between simulation and real experimental data due to several factors not modelled in simulation such as nonlinear acoustics, phase aberration, and multipath scattering. This domain shift adversely affects the network's performance. However, our proposed method maintains an acceptable performance in the new domain without any extra fine-tuning.

4.5 Discussion

We train a deep model to learn a mapping function between raw channel data and a desired high-quality image. Unlike traditional deep learning-based beamforming approaches, the high-quality image is not the output of a complex beamforming techniques such as MVB. Instead, it is the TRF convolved with a sharp Gaussian PSF, leading to a sharp image without any clutter caused by side-lobes. Although the proposed method is similar to other approaches for estimating tissues reflectivity functions, we have not used the Dirac delta function as the PSF in Eq. (4.1) since the mapping function between delayed channel data and TRF is not tractable. The standard deviation of Gaussian kernel used as desired PSF is selected experimentally by trial and error. On one hand, if a lower standard deviation was selected, the resulting image would be sharper. However, the mapping function between input-output becomes intractable and the training does not converge.

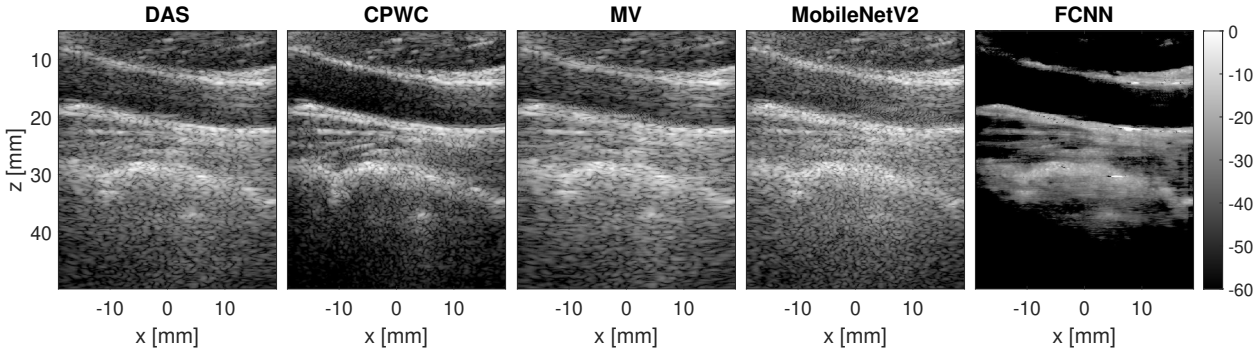

Figure 4.5: Results on *in vivo* data. This image shows a longitudinal view of the carotid artery of a volunteer.

On the other hand, using wider Gaussian kernel reduces the quality of training ground-truth. Therefore, a standard deviation of 0.1 mm, which translates to an axial and lateral FWHM of 0.22 mm and gCNR of 1 for the training targets.

The robustness of our proposed beamforming approach in the real test experiment mainly comes from the following reasons. First, using real photographic images as the ground-truth echogenicity map of simulation phantoms provides the network with enough variety of textures, contrasts, and objects with different geometries in the training phase. It also makes the simulation of a large number of training and validation data possible, which is important to prevent overfitting. Furthermore, using the simulation settings similar to the real experimental imaging settings of *in vivo* test data minimizes the unwanted domain shift between training and test datasets. Fig. 4.3, however, shows a large lateral spread in the top right scatterers, which is not expected and shows that the network did not learn the correct response to a spatially varying input. To understand the reason behind producing such a result, the simulation resolution test data is included in training and found that it is due to an unknown domain shift between training and test data because the nonblind intermediate result of the trained network shows a desired quality without any lateral spread in the top right scatterers. The performance of the proposed FCNN may be further improved by fine-tuning on wire phantoms wherein the exact locations of wires is known. Further insight into proper training of deep models [114] and shedding light on some of their counter intuitive properties [115] warrants a dedicated future work.

A major difference between the proposed method and speckle reduction [48, 97] is that our method generates sharper images whereas speckle reduction often leads to smooth and blurry images that are good for some high-level tasks such as image segmentation. The proposed method for simulating the training ground-truth provides images with a high quality in ultrasound imaging. As seen in the simulation resolution results (Fig. 4.3), the proposed ground-truth provides excellent lateral as well as axial resolution. It is also free from off-axis scattering. As for the simulation contrast results (Fig. 4.4), the ground-truth image is of perfect contrast.

There is a noticeable nonlinear attenuation with depth in the *in vivo* data, which causes the loss of contrast in the result of the proposed method for deep regions because the whole image is reconstructed at once. This behavior is not seen in the method based on MobileNetV2 [71] because each pixel of the image is reconstructed independently in that algorithm. Moreover, the method based on MobileNetV2 was trained using both simulation and real data which makes it more robust to the possible domain shifts of test data.

By considering RF matrices of all piezoelectric elements as the network's input, we ensure that all existing information of the backscattered signals are available in the input domain, and the

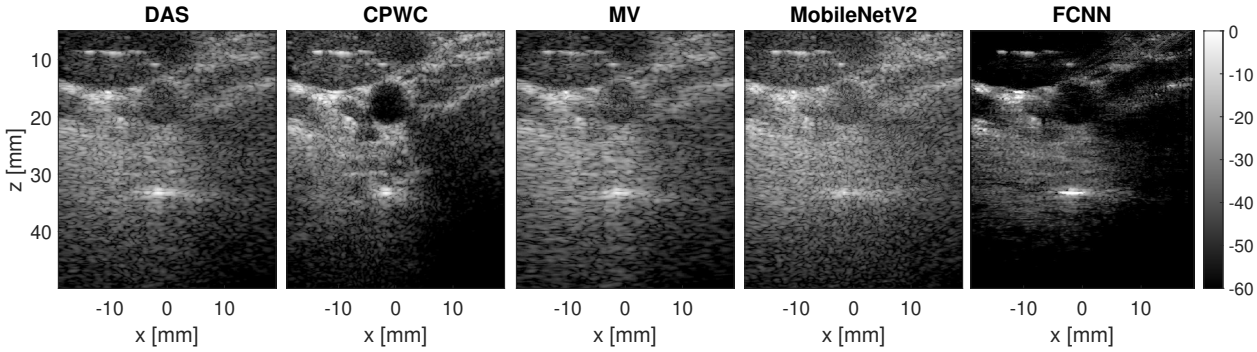

Figure 4.6: Results on *in vivo* data. This image shows a cross-sectional view of the carotid artery of a volunteer.

rawest possible format is provided for the network. It has to be mentioned that if we do not apply the propagation delays to the element's outputs, the task is not tractable for the network, and training does not converge. As for the network output, there are two possibilities including the definition of ground-truth in the format of RF or envelope data. Herein, the second one is used since our goal was to recover the envelope data. Since some applications need the RF data, we plan to extend the proposed method to define the desired PSF in the RF format in the future.

The achieved increase in quality of plane-wave ultrasound imaging is of crucial importance in practice since there is no longer a need to transmit several insonifications at different angles to achieve the optimal quality of CPWC. In other words, the proposed method goes beyond the classical limitations and can be considered as a solution to the intrinsic trade-off between framerate and image quality of plane-wave imaging. As shown in the results, the proposed method also provides improvements in the axial direction, which otherwise is usually achieved with higher center frequencies. As such, the proposed method can also be considered as a potential solution to the intrinsic trade-off between penetration depth and axial resolution. Further validation experiments are necessary to benchmark the efficacy of the proposed method in diagnosis and image-guided interventions.

How robust the method is to the changes of imaging settings is one of the pivotal questions regarding the usage of deep models in clinical applications. Herein, the proposed method is only trained and tested using a single imaging settings. We plan to extend the current work using adaptive instance normalization idea [116] which helps to have a single model for all possible changes in the imaging settings such as transducer specifications, transmission center frequency, etc. Future work can also investigate the effect of variations in the speed of sound and aberrations on the output of the network.

4.6 Conclusions

A reduction in the framerate is the main challenge associated with CPWC. Herein, the proposed beamforming approach works as a nonlinear mapping function from the input space (ultrasound RF channel data) to the high-quality output image. As shown in the results, an FCNN with the proposed step-by-step method of training has been adapted to achieve the high quality of ultrasound images without any loss in framerate. The experiments confirm that the proposed method reconstructs images with a high quality in terms of resolution and contrast, while it also preserves the performance on the *in vivo* datasets.

Chapter 5

Plane-Wave Ultrasound Beamforming Through Independent Component Analysis

This chapter is based on our published paper [117].

Ultrasound imaging experienced a revolution with the introduction of plane-wave imaging (PWI) in which framerate can reach several thousands per second. In contrast to other techniques, PWI fires all elements of the probe simultaneously to form a flat wavefront and span the whole region of interest in a single shot. This technique has been successfully applied to different applications such as imaging of shear waves, contrast imaging, and Doppler imaging of blood flow [17]. Having an unfocused transmitted beam, however, leads to poor resolution and low contrast in PWI. This drawback was addressed by coherent compounding of images obtained by several insonifications of different angles [18]. Consequently, there is always a trade-off between image quality and framerate. Hence, beamforming is witnessing a growing attention in order to enhance the quality of images without sacrificing the framerate.

Specifically for PWI, the minimum variance (MV) approach was applied in [118, 119, 120, 121]. Nguyen and Prager [122] proposed extensions to MV for coherent plane-wave compounding (CPWC). Beamforming based on compressive sensing for PWI was introduced in [123, 124, 4, 125]. Dei *et al.* [126, 127] investigated the performance of their beamforming method entitled aperture domain model image reconstruction (ADMIRE) on PWI. Beamforming in Fourier domain on PWI was first proposed by Lu [128], and then applied with two distinct strategies by Garcia *et al.* [129] and Bernard *et al.* [130]. Beamforming as a regularized inverse problem was introduced in [131] and applied at different depths separately. This point of view was extended in [132] to solve inverse problem for all image depths jointly. A statistical interpretation of beamforming entitled iterative maximum-a-posteriori (iMAP) was introduced in [133].

Herein, we propose a new framework for adaptive plane-wave beamforming wherein apodization weights are estimated through independent component analysis (ICA). In the field of ultrasound imaging, ICA has been mainly used for clutter filtering and noise suppression [134, 135, 136, 137, 138, 139]. ICA was used as a dimensionality reduction technique to speed up ADMIRE beamforming [140].

An overview of the proposed method is shown in Fig. 5.1. When an ultrasound wave is

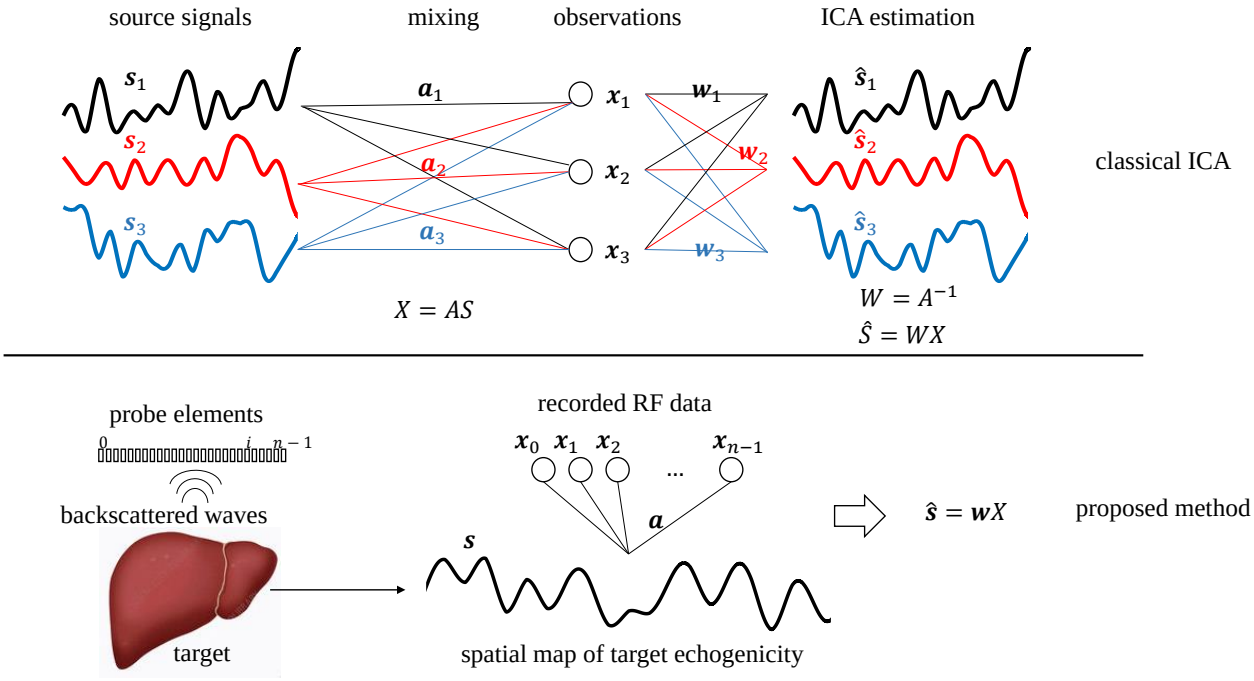

Figure 5.1: An overview of the proposed method and its correspondence with classical ICA. In the top, three sources are reconstructed using three observations. In the bottom, the source is 2D spatial echogenicity map, and the vectorized source s is reconstructed using n observations.

transmitted into the field, the most popular way to trace back the backscattered wave corresponding to each pixel of the target is by applying the associated propagation delay to each signal recorded by piezoelectric elements of the probe. Consequently, a discrete spatial map of the target echogenicity can be constructed as an ultrasound image. However, there is also the possibility of source ambiguity. More specifically, the backscattered waves from pixels at equal distances from the piezoelectric element lead to a single sensory data at the resulting RF signal. This physical limitation brought about by wave propagation is what motivates us to make use of ICA to suppress the pixel correlation in the imaging field after spatiotemporal mapping. Fortunately, the group of pixels which are indistinguishable from the output of each piezoelectric element are not the same. Therefore, our approach considers the signal recorded by each piezoelectric element as a non-independent observation of the target echogenicity and then uses ICA, as an adaptive beamforming method, to extract the independent spatial map of target echogenicity. Herein, the independency of spatial map means that each sample contains the trace of only one pixel. In the proposed method, the apodization window is first estimated using ICA algorithm, and then it is applied throughout the image based on a predefined f-number (f#).

5.1 Methods

Consider a linear array of n crystal elements, symmetrically distributed on the x-axis, transmitting along the positive z-axis (Fig. 5.2(a)). Let us assume that a plane-wave with angle α spans the domain with a sound speed of c. The backscattered signals received by crystal element i located at x_i is denoted by $\mathbf{h}_i(t)$ (Fig. 5.2(b)). Without any loss of generality, $zcos(\alpha) + xsin(\alpha)$ is the transmission distance d_t from the origin of the transmitted plane-wave to an arbitrary point (x, z)

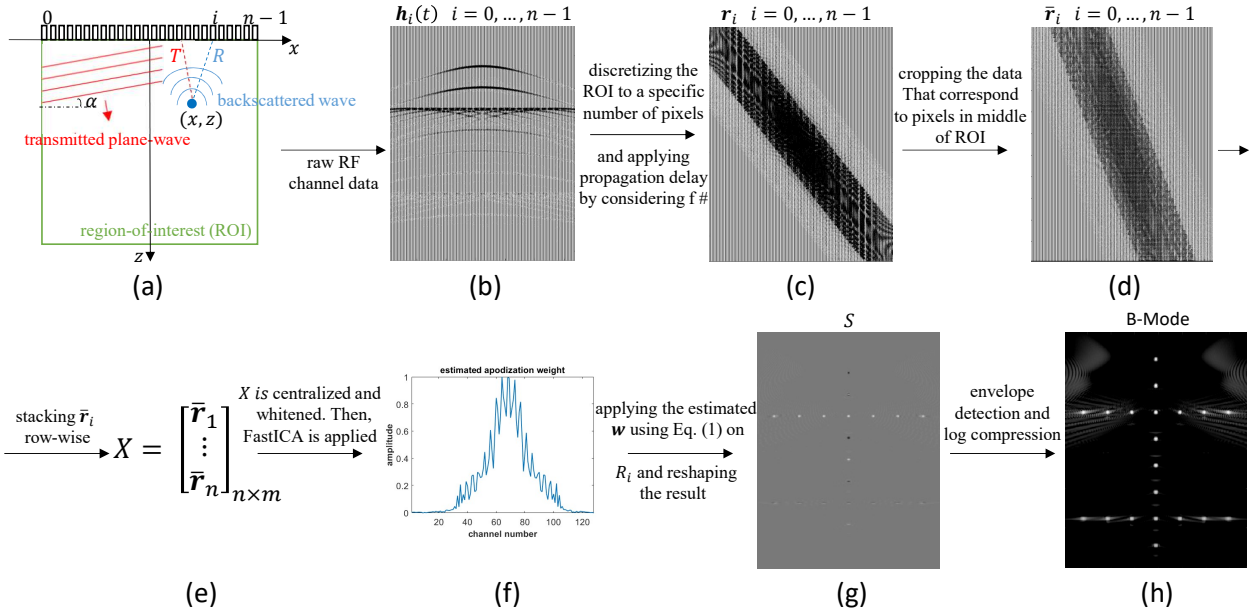

Figure 5.2: The proposed adaptive beamforming pipeline. (a) Geometrical illustration of the PWI. (b) The plot of backscattered signals recorded by piezoelectric elements. (c) The plot of vectorized RF matrices constructed by applying propagation delay to raw RF channel data by considering the f#. (d) The plot of vectorized cropped RF matrices corresponding to pixels in the middle of ROI. (e) The observation matrix of ICA constructed by stacking $\bar{\mathbf{r}}_i$ row-wise. (f) The estimated apodization window using ICA. (g) The beamformed RF matrix S. (h) The final B-Mode image.

in the region-of-interest (ROI) and $\sqrt{(x-x_i)^2+z^2}$ is the receiving distance d_r from (x,z) to the location of crystal element i (for more details see [18]). Let us define R_i as a matrix containing the RF data recorded by crystal element i corresponding to each point (x,z) in the ROI and its elements can be found by applying the associated propagation delay to $\mathbf{h}_i(t)$ as follows (hereafter, capital and bold font variables represent matrices and vectors, respectively):

$$\tau(x,z) = \frac{d_t + d_r}{c} \Longrightarrow R_i(x,z) = \mathbf{h}_i(\tau(x,z)). \tag{5.1}$$

As shown in Eq. (5.1), each piezoelectric element gives one RF matrix of the ROI (vectorized R_i are denoted by $\mathbf{r}_i$ and shown in Fig. 5.2(c)). Therefore, the beamformed RF matrix S (Fig. 5.2(f)) is the result of information fusion among different crystal elements, and each element of $S(x,z)$ can be obtained through the following weighted summation:

$$S(x,z) = \sum_{i=0}^{n-1} \mathbf{w}_i(x,z) R_i(x,z), \tag{5.2}$$

where $\mathbf{w}$ is the apodization window of length n (Fig. 5.2(e)). In practice, however, we utilize dynamic beamforming where the f# is fixed for the entire image. Therefore, l is defined as the number of crystal elements considered for the reconstruction of each depth (z) of the image and is calculated as follows [18]:

$$f\# = z/l. \tag{5.3}$$

After construction of S, it is subject to envelope detection and *log* compression in order to obtain the final B-Mode ultrasound image (Fig. 5.2(g)). Herein, our goal is to estimate the apodization window $\mathbf{w}$ using ICA.

5.1.1 Independent component analysis

A full mathematical description of ICA is provided in [141]. We include a short summary here to make this chapter self-contained. Our proposed method is explained in the next section. ICA is a framework used to separate signal components mixed in observations recorded at different transducer elements [141]. Assuming an n-dimensional signal space, i.e., an n-dimensional observed data $\mathbf{x}$, n-dimensional independent sources $\mathbf{s}$, and a square transformation matrix W of size $n \times n$, the mixing model can be written as follows [141]:

$$\mathbf{s} = W\mathbf{x}. \tag{5.4}$$

With the assumption of having independent and non-Gaussian sources (at the most one independent Gaussian source is allowed), both W and $\mathbf{s}$ can be estimated using the ICA algorithm. In practice, the objective function for ICA estimation can be formulated using different measures of non-Gaussianity such as kurtosis, negentropy, and mutual information. Moreover, it is very useful to center and whiten the observations before applying ICA. One of the most famous algorithms of ICA implementation is FastICA, where a unit vector $\mathbf{w}$ is computed such that the dot product $\mathbf{w}^T\mathbf{x}$ maximizes negentropy. The FastICA algorithm can be summerized in four steps as follows [141]:

1. Random initialization of vector $\mathbf{w}$.
2. $\mathbf{w}_{new} = E\{\mathbf{x}g(\mathbf{w}^T\mathbf{x})\} - E\{g'(\mathbf{w}^T\mathbf{x})\}\mathbf{w}$
3. $\mathbf{w} = \mathbf{w}_{new}/\|\mathbf{w}_{new}\|$
4. Return to step 2 until the direction of $\mathbf{w}$ does not change.

notation E refers to the expectation operation. Symbols g and g' are first and second derivatives of a non-quadratic nonlinear function f, respectively. It was shown that either of the two functions f is robust for negentropy estimation [142]:

$$f(u) = \frac{1}{a_1}\log\cosh(a_1 u), \text{ or } f(u) = -\exp(-u^2/2), \tag{5.5}$$

where $1 \leq a_1 \leq 2$. More details regarding the FastICA algorithm can be found in [141].

5.1.2 Beamforming using ICA

In general, our goal is to reconstruct a high-quality ultrasound image which is a spatial map of the target echogenicity. More specifically, we discretize the map of scatterers that leads to pixels. Each pixel corresponds to an averaged tissue reflectivity function over the extent of the pixel. When the RF data corresponding to each pixel of the final image is extracted from the output of each crystal element (using Eq. (5.1)), there is also the possibility of source ambiguity. More specifically, the backscattered waves of at least two different pixels at equal distances from the piezoelectric element arrive simultaneously and lead to a single sensory data at the resulting RF signal. In this section, first, this problem is mathematically demonstrated and then our proposed solution is explained.

Without loss of generality, when $\alpha = 0$, the backscattered waves of two distinct pixels (with indices 1 and 2) arrive at the same time in crystal element i if and only if they have the same propagation delay τ. Form Eq. (5.1) and if the first pixel is in the lateral position x_i, it can be written that:

$$\tau_1 = \tau_2 \Longrightarrow 2z_1 = z_2 + \sqrt{(x_2 - x_i)^2 + z_2^2}. \tag{5.6}$$

Moving z_2 to the left side of the equality and squaring both sides, we get:

$$4z_1^2 + z_2^2 - 4z_1z_2 = (x_2 - x_i)^2 + z_2^2. \tag{5.7}$$

Cancelling z_2^2 from both sides, the resulting expression can be rearranged as:

$$4z_1z_2 = 4z_1^2 - (x_2 - x_i)^2. \tag{5.8}$$

Diving both sides by $4z_1$ (which is always nonzero) gives:

$$z_2 = z_1 - \frac{(x_2 - x_i)^2}{4z_1}. \tag{5.9}$$

So, for $z_2 < z_1$, there are a bunch of pixels located on an ellipse, whose reflections arrive at the same time as for pixel 1. In other words, all those pixels have the same value in the extracted RF matrix $R_i(x, z)$ from Eq. (5.1). In the continuous case, this problem is fully addressable. In the discrete case, however, there is the error due to quantization as well. Although this problem was shown for the specific case of $\alpha = 0$, the concept can be extended for different angles.

As seen from Eq. (5.9), the group of scatterers from whom reflections arrive simultaneously are not the same for each crystal element. In other words, the group of indistinguishable pixels in each $R_i(x, z)$ is distinct. This point provides the opportunity of source separation using ICA. If we consider each $\mathbf{r}_i$ as a non-independent observation of the discretized map of scatterers, our task is to extract the beamformed RF matrix S out of these non-independent observations (Eq. (5.2)). Therefore, we adapt ICA to estimate the apodization window $\mathbf{w}$. In the ideal case, the explained pixel correlation in the imaging field after spatiotemporal mapping is perfectly suppressed, and the desired S contains independent elements containing the trace of only one distinct pixel. It has to be mentioned that in practice, the axial and lateral resolutions are based on the sampling frequency of the system, center frequency of the transmitted wave, and transducer design. So, one pixel results in one voxel of the ROI.

In ultrasound beamforming, an issue is that the apodization window is not fixed throughout the image. More specifically, ICA works with a fixed transformation matrix W in Eq. (5.4). In ultrasound images, however, the apodization weight is not spatially invariant, rendering a different set of weights for different pixels. Two points make the apodization weights spatially variant. First, for pixels lying at the two lateral ends of the image, there are crystal elements predominately lying along one side. Second, as explained in Section 5.1, pixels at different depths of the image are reconstructed using a different number of elements to keep the f# fixed across the entire image. Hence, if we do not consider these points, ICA fails to estimate the source and apodization windows, leading to images that are even lower in quality than DAS.

To solve the aforementioned problem, first, we consider the f# while constructing the $R_i(x, z)$ (Fig. 5.2(c)). More specifically, $R_i(x, z)$ contains a nonzero value only if crystal element i is considered for reconstructing depth z of the image (Eq. (5.3)). Second, we consider only the central pixels of the image around which the crystals are symmetric as the input to the ICA algorithm (Fig. 5.2(d)). In this way, the cropped portion of $\mathbf{r}_i$ is used to construct the observation matrix X. In our ICA formulation, therefore, the observations are RF data corresponding to central pixels of final image that are recorded by crystal elements of the probe. It has to be mentioned that the random initialization of the transformation matrix W only affects the number of iterations that the algorithm takes to converge, and it does not lead to any variation in the results after beamforming.

As shown in Fig. 5.2, our proposed adaptive beamforming algorithm for PWI using ICA includes the following steps:

1. The region of interest is discretized into pixels. By considering a specific f#, the propagation delays are applied to the raw RF channel data to generate $\mathbf{r}_i, i = 0, ..., n-1$ (Fig. 5.2(c)).

2. Each $\mathbf{r}_i$ is considered as an observation of the field. First, it is cropped (denoted by $\bar{\mathbf{r}_i}$) and then the observation matrix X is constructed by stacking the row vectors ($\bar{\mathbf{r}_i}$). Finally, matrix X is centralized and whitened before running ICA (Fig. 5.2(d-e)).

3. The independent source and corresponding mixing vector are estimated using the FastICA algorithm [143] by maximizing Negentropy as the measure of non-Gaussianity (Fig. 5.2(f)).

4. The apodization window (the estimated transformation vector in the last step) is applied throughout the image based on a predefined f# (Fig. 5.2(g)).

In step 3, apodization window $\mathbf{w}$ is iteratively updated to maximize negentropy (estimated using $f(u) = -\exp(-u^2/2)$) with respect to $\mathbf{w}^T X$. More specifically, the fixed-point iteration scheme [143] uses g and g' which are first and second derivatives of the non-quadratic nonlinear function f to find the optimal value of $\mathbf{w}$. As discussed in [141], the number of sources, in the ICA algorithm, has to be equal or less than the number of observations. ICA can be considered as a variant of the projection pursuit algorithm [144], which enables one-by-one estimation of the independent components. Herein, the number of independent sources in the ICA algorithm is set to one since we only look for a single source, which is a collection of all scatterers. This is an important feature that substantially reduces the computational load.

5.2 Experiments

5.2.1 Dataset

In this section, we use a publicly available PWI dataset, entitled PICMUS, which was provided by the IEEE International Ultrasonics Symposiun (IUS 2016) in order to benchmark novel beamforming methods [70]. The PICMUS data utilized in this work include:

1. Simulation resolution (SR): A simulated ultrasound image containing point targets distributed vertically and horizontally over an anechoic background designed to assess the performance of beamforming methods in terms of spatial resolution.

2. Simulation contrast (SC): A simulated ultrasound image containing anechoic cysts distributed vertically and horizontally over fully developed speckle designed to assess the performance of beamforming methods in terms of contrast.

3. Experimental Resolution (ER): An experimental ultrasound image was recorded on a CIRS Multi-Purpose Ultrasound Phantom (Model 040GSE) in the regions containing several wires against speckle background to assess the performance of beamforming methods in terms of spatial resolution.

4. Experimental contrast (EC): An experimental ultrasound image was recorded on the same phantom as ER but in the regions containing two anechoic cysts against speckle background to assess the performance of beamforming methods in terms of contrast.

In addition, PICMUS dataset also contains two *in vivo* ultrasound images, showing cross-sectional (denoted by Carotid Cross (CC)) and longitudinal views (denoted by Carotid Longitudinal(CL)), recorded on the carotid artery of a volunteer. All of the phantom and *in vivo* data were collected

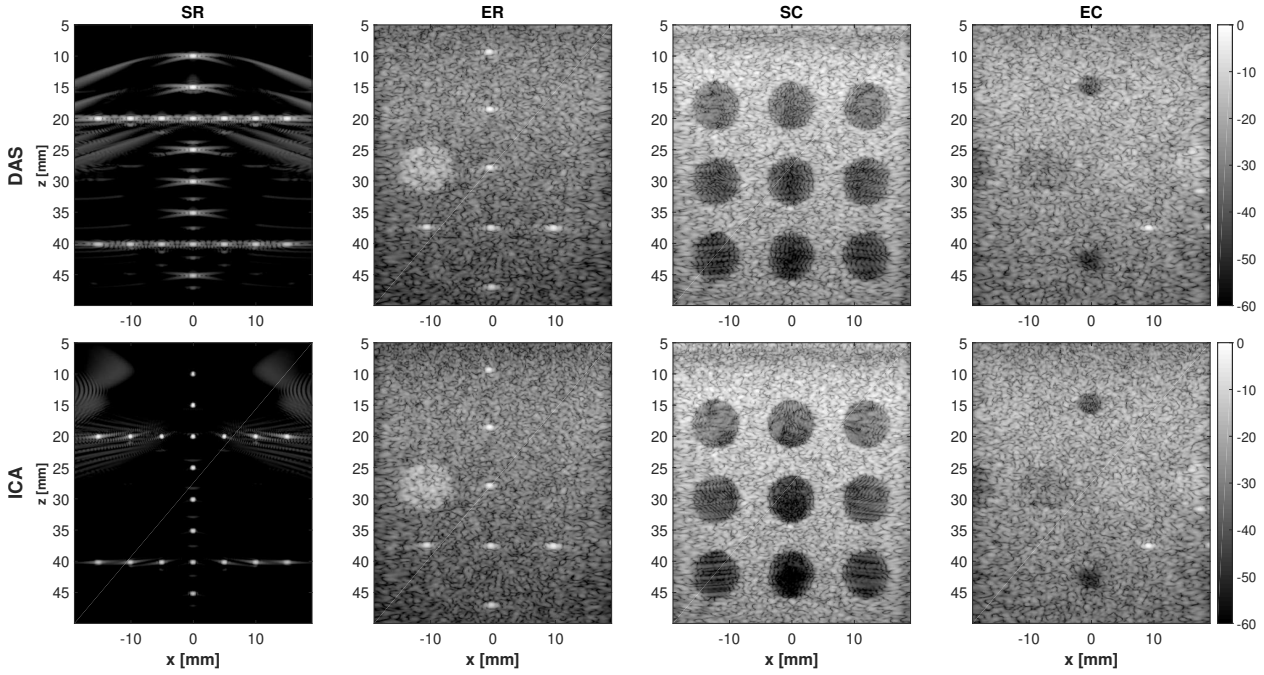

Figure 5.3: Beamforming results on the single 0° plane wave. Columns indicate different image data sets while rows correspond to beamforming methods.

using a Verasonics Vantage 256 research scanner and a L11 probe (Verasonics Inc., Redmond, WA). The simulation settings were selected to be as similar as possible to the experimental setup.

For each of mentioned groups, a collection of received prebeamformed data corresponding to 75 steered Plane-Waves covering the angle span from $-16°$ to $16°$ was provided. Both RF and IQ (phase quadrature) formats of data were provided. The proposed algorithm works on the RF version of data. More details regarding PICMUS dataset can be found in [70].

5.2.2 Implementation details

As explained in Section 5.1.2, the FastICA algorithm is used to estimate the apodization window. The maximum number of iterations is set to 100 and the stopping criterion is set to be $\epsilon = 10^{-6}$. The weights are initialized with random numbers extracted from standard distribution. The reduction of dimension through PCA is not used and the best results which are most reproducible are attained by considering all eigenvalues in the estimation procedure. We use the Matlab implementation of the Fixed point ICA, the main algorithm of FastICA, which is publicly available online `http://research.ics.aalto.fi/ica/fastica/code/dlcode.shtml`.

Throughout the results section, we consider $f\# = 1.75$ and use Tukey (tapered cosine) window with constant parameter set to 0.25 for DAS and other adaptive methods on top of DAS.

It is not possible to theoretically prove the convergence of FastICA algorithm with the mentioned parameters. In practice, however, we set the maximum number of iterations equal to 100 and observe that for all of data sets, the algorithm converges in a lower number of iterations.

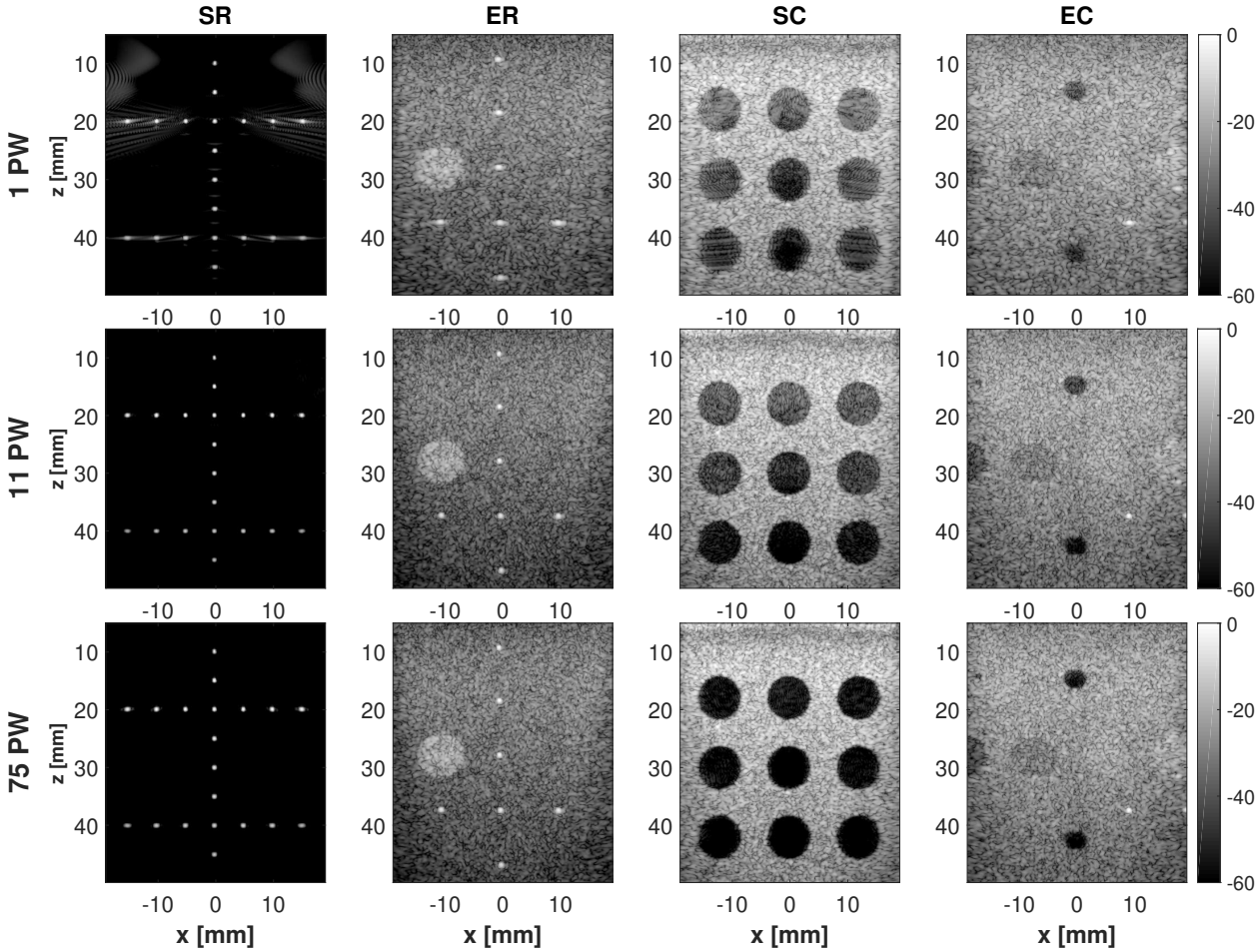

Figure 5.4: ICA beamforming using 1, 11, and 75 plane waves. Columns indicate different image data sets and rows correspond to the number of transmitted plane waves.

5.3 Results

5.3.1 Simulated and experimental data

The results of DAS beamforming versus our proposed method based on ICA on a single 0° plane wave of simulated and experimental data are illustrated in Fig. 5.3. As seen from this figure, the proposed beamforming method outperforms DAS and improves the resolution as well as contrast on both simulated and experimental phantom data. In order to better investigate the amount of improvement, the quantitative indices are reported in Table 5.1. What causes the improvement is the window used for apodization. So, as Table 5.1 confirms, improvement in resolution can only be acquired in the lateral direction. The highest improvement in lateral FWHM is 36.5% on simulated plane-wave data of only one single transmission. In terms of CNR, approximately 9% of improvement is achieved on the experimental cyst data of a single transmission while boarders of the cyst are also sharper. As mentioned before, for pixels lying at the two lateral ends of the image, symmetrical channel data is not available and data is predominately limited to one side. This point forced us to only consider the central pixels of the image. Therefore, there will be a reduction in image quality in those border regions because the weights are not optimized for those region. Furthermore, this reduction in quality is more visible in shallow regions of the image (the

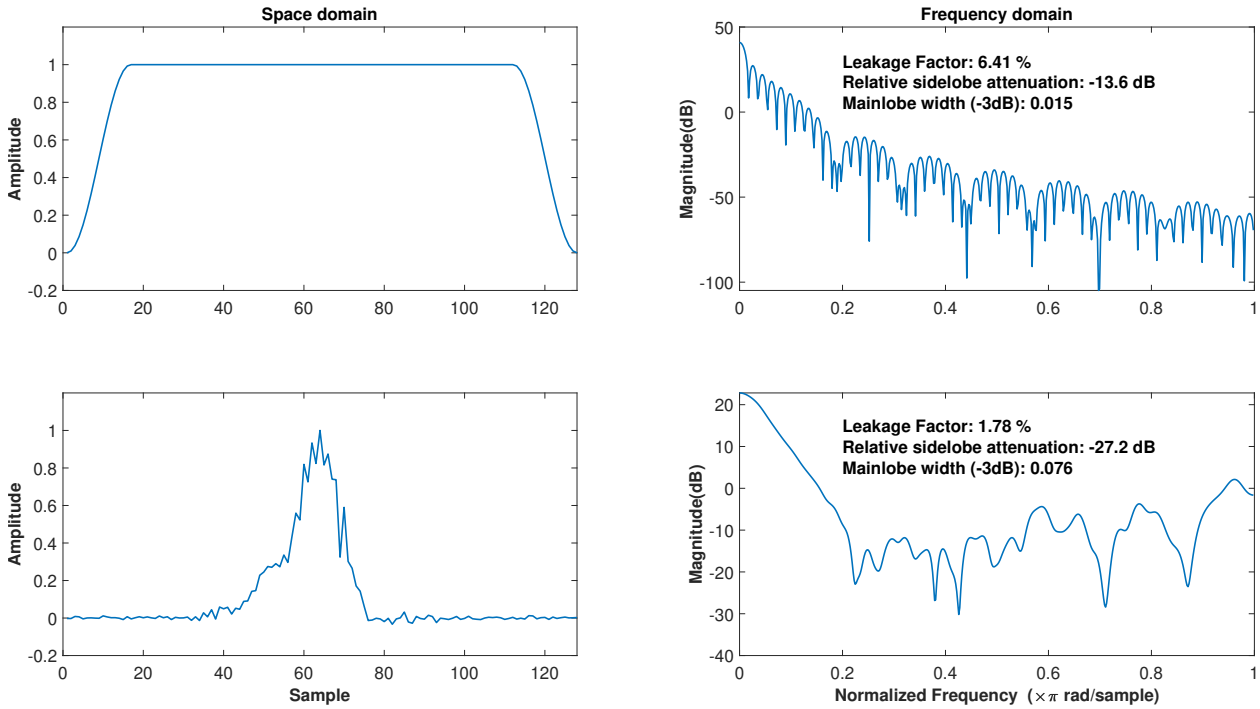

Figure 5.5: Comparison of apodization window estimated using ICA (second row) and Tukey25 (first row) used in DAS. Windows are shown in both space and frequency domains.

reconstructed SR image with ICA in Fig. 5.3) since a limited number of elements are considered for the reconstruction of pixel intensities. It has to be mentioned that this problem is one of limitations of ultrasound imaging regardless of the beamforming method.

In order to investigate the effect of CPWC, the results of the proposed method on higher number of plane waves are illustrated in Fig. 5.4. The indexes of Table 5.1 as well as Fig. 5.4 confirm that CPWC improves the image quality. As for CPWC, we do not repeat beamforming for each angle and use the apodization weights of the 0° plane wave for the remaining angles. Moreover, to limit the sources of achieved improvement, we do not apply any angular apodization. In fact, our main focus is on beamforming of the received signals.

To better understand the effect of proposed method, Fig. 5.5 shows a comparison between Tukey25 window used in DAS and the apodization weights estimated by ICA on ER dataset. The estimated window has a lower leakage, calculated as the ratio of power in the sidelobes to the total power, factor as well as a relative side lobe attenuation while its main lobe is wider. The estimated window is of a different shape which can not be found among predefined common windows. So, this point confirms the necessity of estimating the apodization window from the received data.

Fig. 5.6 demonstrates qualitative improvements with ICA and DAS methods as a function of the number of PWs. As can be seen in Fig. 5.6, the proposed approach achieves better lateral resolution with only 3 plane waves compared with DAS using 75 angles. In terms of contrast, however, the proposed approach achieves better CNR with 51 plane waves compared with DAS using 75 angles. Therefore, it is possible to reduce the number of plane wave transmits needed to achieve image quality similar to a fully sampled transmit.

Table 5.1: Quantitative results in terms of CNR and FWHM indexes for simulation and real phantom experiments.

dataset		SR		ER		SC	EC
index		$FWHM_A$	$FWHM_L$	$FWHM_A$	$FWHM_L$	CNR	CNR
1 PW	DAS	0.4	0.82	0.57	0.88	9.95	8.15
	ICA	0.39	0.52	0.57	0.81	10.67	8.9
11 PW	DAS	0.4	0.54	0.56	0.54	12.48	11.25
	ICA	0.4	0.41	0.56	0.51	12.6	11.4
75 PW	DAS	0.4	0.56	0.56	0.56	15.55	12
	ICA	0.4	0.42	0.56	0.53	15.96	12.1

5.3.2 *In vivo* data

In real ultrasound tissues, there are more sources of degradation in image quality. In order to make sure that the proposed method also works on *in vivo* data, the results of beamforming on real carotid images of PICMUS dataset are provided in Fig. 5.7. Visual comparison of beamformed images with different number of angles reveals that ICA outperforms classical DAS in both cross as well as longitudinal sections. Furthermore, ICA results in sharper images with a better contrast.

5.3.3 Comparison with other adaptive methods

As mentioned before, our focus in current chapter is on beamforming of the received signals. So, comparison with other adaptive approaches is of crucial importance. In this way, we present the results of five well-known approaches, namely MV [73], EMV [2], CF [145], generalized CF (GCF) [145], and PCF [3]. The comparison with these methods was not possible without using codes provided by Rindal *et al.* [146] in ultrasound toolbox repository (`http://www.ustb.no/publications/dynamic_range/`). The parameters used in the MV and EMV methods are specified next to enable interested readers to reproduce the presented results. The subarray size equals half of the number of crystal elements (64 in our case), the temporal averaging factor is given by 1.5, and

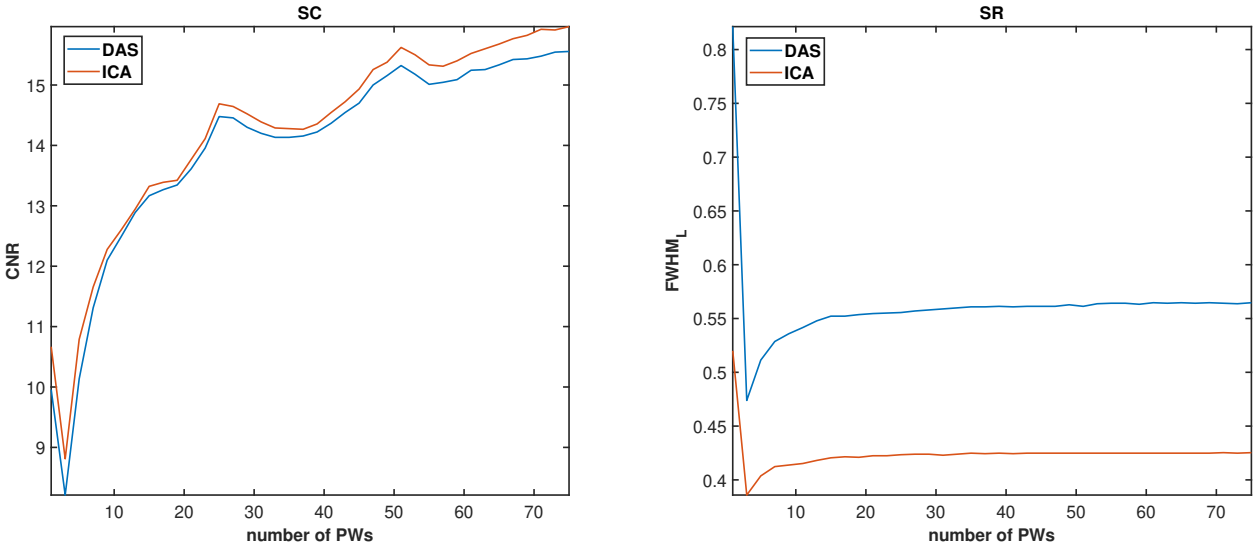

Figure 5.6: Quantitative comparison of beamforming results using different number of plane waves. Left column indicates SC case while right column corresponds to SR case.

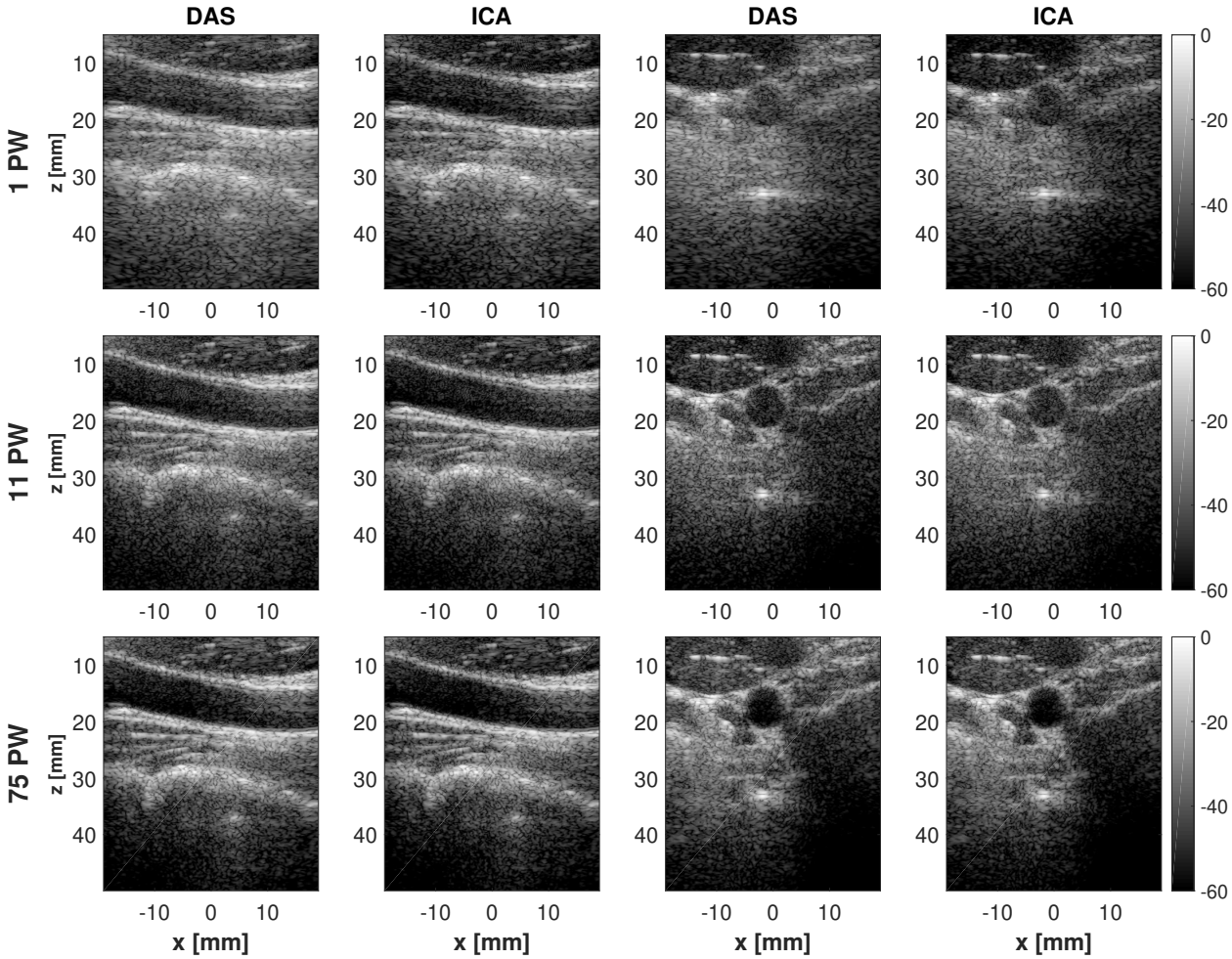

Figure 5.7: Beamforming results on *in vivo* data using 1, 11, and 75 plane waves. Two columns in left indicate cross-section images while left two columns correspond to longitudinal-section. Rows denote different number of transmitted plane waves used in beamforming.

the regularization factor (i.e., the diagonal loading) is 0.01. In the EMV approach, all eigenvectors of the covariance matrix are considered in creating the signal subspace. Fig. 5.8 shows the result of different adaptive beamforming algorithms on a single 0° plane wave of simulated and experimental data. The quantitative comparison is provided in Table 5.2. The EMV method outperforms all other methods, even our proposed method, in terms of indices. However, methods based on the MV are very time consuming and are not practical for online applications. In terms of computational time, our proposed method typically takes 75 milliseconds to estimate the apodization window while MV and EMV take a few minutes, and CF methods take one second. The approaches based on CF outperform the proposed approach in terms of FWHM index while are worse in terms of contrast.

5.4 Discussion

If we do not consider the f# in constructing the observation matrix, all pixels contribute equally in specifying the elements' weights. In practice, the backscattered signals from pixels located at

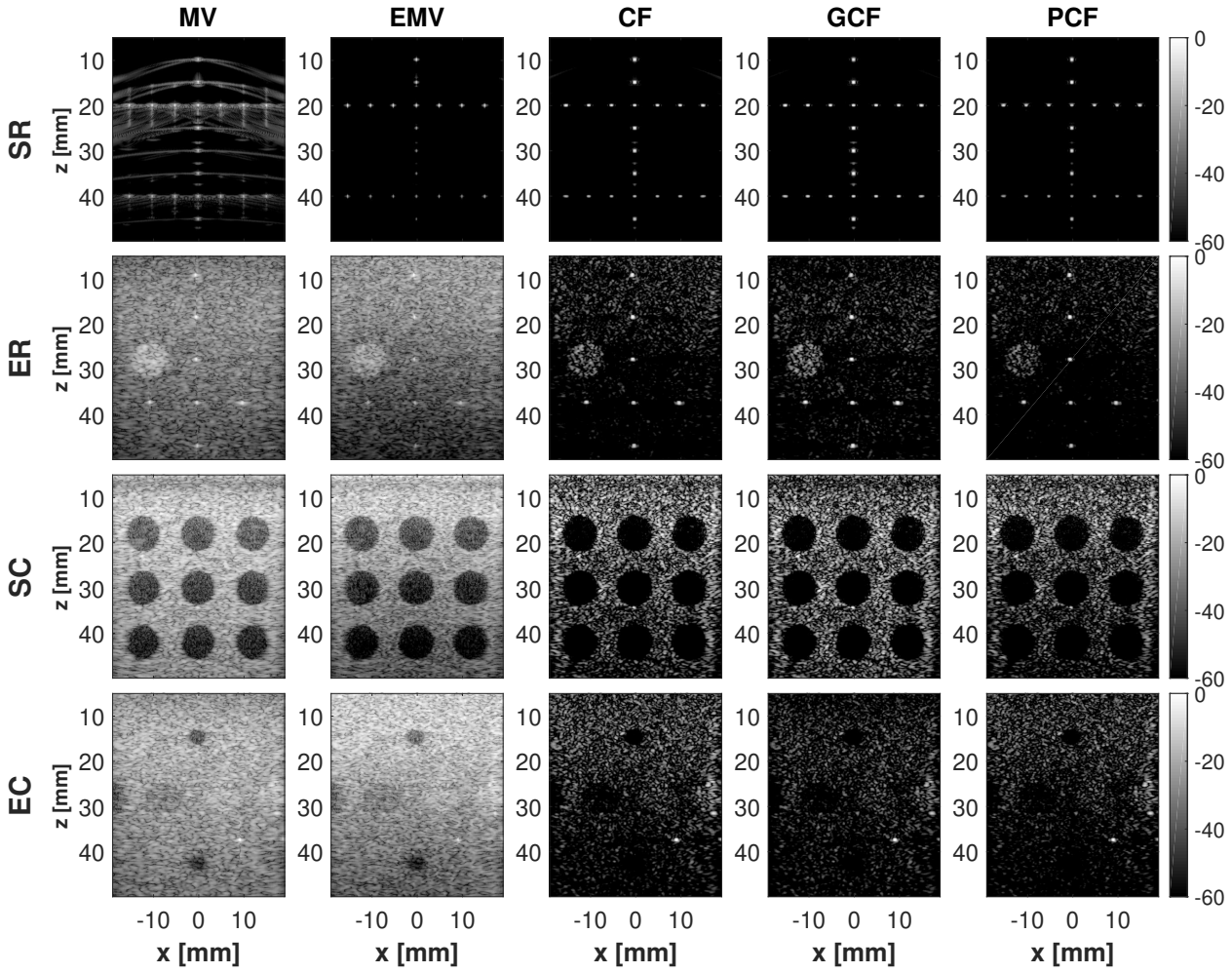

Figure 5.8: Results of other adaptive beamforming methods on the single 0° plane wave. Columns indicate different image data sets while rows correspond to different adaptive beamforming methods.

shallow regions are recorded precisely by a small subset of elements close to the pixel location. The backscattered signals originating from the subordinate pixels, however, can be properly recorded by most of the elements. Therefore, it makes more sense to restrict the trace of each pixel only to its corresponding elements. Otherwise, ICA fails to estimate the source and apodization windows, leading to images that are even lower in quality than DAS.

Using a part of samples of each channel which only correspond to the middle part of the final image is important from two aspects. First, it removes the effect of incomplete data of borders on the ICA performance. Second, the FastICA algorithm converges faster as it works with a lower amount of data. Note that the estimated apodization weights and its specifications such as width of main lobe or the amount of side lobe attenuation in each dataset are different. So, it can be concluded that there is not a unique solution that works for all data. It is worth mentioning that although the f# is applied before ICA, it does not imply that some of the probe elements are ignored completely. More specifically, as shown in Fig. 5.2, the entire aperture is used in constructing the observation matrix X. The trace of considered pixels, therefore, may be in some or all of r_i depending on the location of the pixel.

Table 5.2: Quantitative results of other adaptive beamforming methods in terms of CNR and FWHM indexes for simulation and real phantom experiments.

dataset		SR		ER		SC	EC
index		$FWHM_A$	$FWHM_L$	$FWHM_A$	$FWHM_L$	CNR	CNR
1 PW	MV	0.41	0.1	0.59	0.43	11.1	7.95
	EMV	0.39	0.09	0.58	0.33	12	8.1
	CF	0.32	0.44	0.48	0.47	8.2	6.3
	GCF	0.32	0.43	0.48	0.47	8.1	6.3
	PCF	0.29	0.37	0.46	0.41	6.9	5.2
11 PW	MV	0.43	0.1	0.59	0.29	11.4	9.8
	EMV	0.4	0.09	0.56	0.28	15.2	11.5
	CF	0.37	0.37	0.55	0.37	11.9	10.2
	GCF	0.38	0.36	0.55	0.37	11.8	10.2
	PCF	0.37	0.3	0.55	0.31	11	9.05
75 PW	MV	0.43	0.1	0.58	0.31	14.7	11
	EMV	0.4	0.09	0.56	0.29	17	10.4
	CF	0.4	0.38	0.56	0.38	14.05	10
	GCF	0.4	0.38	0.56	0.38	13.9	10
	PCF	0.39	0.29	0.56	0.32	14.13	10.3

The algorithm can be separately applied to estimate weights associated with different insonification angles. However, the weights for different angles are not much different, overall improvement is negligible, and processing time is increased by a multiple equal to the number of angles. The angular apodization can also be estimated using ICA for CPWC. However, the main focus of this chapter was apodization of the received signals. To limit the sources of improvement, the angular weights are not used which also makes the comparison with other approaches possible.

As for the agreement between ICA assumptions with the nature of our problem, it is shown in [147] that the underlying statistics of ultrasound pressure field is Gaussian. However, as mentioned in Section 5.1.1 and shown in [141], the ICA can still be used even if only one of the independent components is Gaussian. Therefore, we can use the ICA algorithm to estimate the desired discretized map of scatterers as the only Gaussian component. In future, we plan to use IQ data using the FastICA algorithm developed for complex-valued data [148].

5.5 Conclusions

We have proposed a new beamforming approach for ultrasound plane-wave imaging based on ICA. Beamforming has been formalized as the estimation of one independent image out of several non-independent observation and the apodization weights have been estimated based on collected data. The images of one single plane-wave transmission as well as multiangle plane-wave acquisitions have been successfully reconstructed. Results show that the proposed method simultaneously improves the resolution and contrast while the resulting image is also visually appealing.

Chapter 6

Inverse Problem of Ultrasound Beamforming with Denoising-Based Regularized Solutions

This chapter is based on our published paper [149].

Medical ultrasound probe is made of several piezoelectric elements used for the transmission of non-invasive acoustic waves into the medium and also the reception of the backscattered signals. To insonify the medium with a desired wave, the excitation pulses of transducer elements as well as their firing times are adjusted in a process called transmit beamforming [9]. The backscattered echoes from each location of the medium are traced back in receive beamforming to reconstruct a spatial map of the tissue echogenicity [9].

Although receive beamforming is an ill-posed inverse problem, delay-and-sum (DAS) algorithm is commonly used for real-time ultrasound imaging in commercial scanners. DAS simply provides the backprojection solution of the inverse problem of beamforming [150]. Nevertheless, DAS uses a predefined apodization window for the entire image which limits the resulting quality by the well-known trade-off between the level of side lobes and the width of the main lobe in the frequency domain. Adaptive beamforming methods have been developed to effectively determine the apodization weights based on the echo signals [73, 151, 152].

Inverse problem formulation of ultrasound beamforming is another alternative to DAS, wherein a measurement model is considered for the synthesis of the desired image [153, 6, 125, 131, 132]. Linear models relate each sample of received channel data to pixels of the image to be recovered through a weighting matrix. While linear models are simple and provide a plausible approximation of the image under scrutiny, the weighting matrix is usually of size several hundreds of thousands which makes the model memory intensive. Despite its high dimension, the weighting matrix is usually sparse and thus easy to store. However, the high dimension of the inverse problem to be solved requires optimization algorithms and representations (e.g., operators) that do not include matrix inversion [131, 132, 125, 154].

Similar to most inverse problems in computational imaging, the existing inverse problem-based beamformers in ultrasound imaging use regularization functions derived from *a priori* statistics of the ultrasound image. The most used are Gaussian models, turning into an ℓ_2-norm regularization term, or Laplacian promoting sparsity through the ℓ_1-norm [131, 132, 125, 154, 155]. While the

latter is well-adapted to reconstruct ultrasound images with high resolution and contrast, sparse solutions have shown a poor performance in preserving the speckle texture, which is an important feature for applications such as motion estimation and tissue classification.

As an alternative to standard regularization functions used to solve image reconstruction or restoration problems, an important class of methods has been proposed in the computational imaging literature [156, 157, 158, 159]. The main idea is to use denoising algorithms as regularizers. Specifically, Venkatakrishnan *et al.* proposed an interesting idea termed plug-and-play (PnP) [157] based on the alternating direction method of multipliers (ADMM) which allows decoupling the measurement model and the regularization terms. It has been shown that the proximity operation related to the regularization term can be replaced by an image denoising algorithm [157]. This idea has been increasingly applied to a number of applications [156, 158, 160, 161]. Nevertheless, the explicit objective function of PnP approach is unknown, and this issue strongly limits studying theoretical convergence properties and interpretations. New insights on characterization of proximity operator and the proof of convergence for the PnP algorithm under certain conditions have been presented in [162, 163, 164].

Recently, Romano *et al.* proposed a skillful way to plug-in denoising algorithms when solving imaging inverse problems called regularization by denoising (RED) [159]. The explicit regularizer of RED is designed to enforce the orthogonality of the image and what a denoiser removes from the image. It has been shown that for locally homogeneous denoisers, the gradient expression of regularizer can be easily found, and several iterative algorithms were proposed to find the optimal solution [159]. Reehorst and Schniter [165] shed more light on the RED algorithm and have shown that besides local homogeneity, the denoising algorithm must also be Jacobian symmetric in order to be explained by an explicit regularization term. Although RED works very well in practice, it has been shown that common denoisers lack Jacobian symmetry property. Therefore, in [165] RED has been explained in a novel framework called score-matching by denoising (SMD).

Herein, inspired by the success of PnP and RED algorithms in various medical imaging inverse problems [166, 161, 167, 168], we devise a general framework for the inverse problem of ultrasound beamforming based on the ADMM. We use a linear forward model for the image under scrutiny, and the basic solution is found by considering ℓ_1-norm regularizer which we will refer to henceforth as ADMM solution. Moreover, the proposed framework is extended with both PnP and RED algorithms refereed to henceforth as PnP and RED solutions, respectively. Furthermore, the source codes for Matlab implementations of the proposed algorithms are publicly available in these links: `https://github.com/Sobhan-Goudarzi/Denoising-Based-Ultrasound-Beamforming` and `code.sonography.ai`.

6.1 Related work

During the past few years, inverse problem formulations have attracted a growing interest in the field of medical ultrasound imaging. They have been used in different problems such as deconvolution [169, 170], despeckling [171, 172], compressive sensing (CS) [173, 174, 175, 176, 177], and beamforming [131, 132, 125]. In particular, CS applications perhaps popularized the inverse problem-based approaches in ultrasound imaging. The CS reconstruction of radio frequency (RF) channel data was performed in [173]. Afterward, CS was applied on beamformed envelope data to reduce the size of stored data [174]. CS has also been used to reconstruct a high-quality ultrasound image by using a reduced number of transducer elements [175, 176] or sub-Nyquist sampled data [177]. Later, beamforming has been formulated as an inverse problem in order to improve the quality of ultrasound images. Szasz *et al.* assumed a linear model between the RF

channel data and the desired image [131]. Each image depth was then reconstructed by solving a regularized inverse problem, with applications to both plane-wave [178] and focused imaging [131]. This approach was further extended by adding more regularization terms and reconstructing all image depths concurrently [132]. Besson *et al.* developed two matrix-free formulations to mitigate the memory and computational requirements of the inverse problem of ultrafast imaging [125].

Deep learning has become another option for solving the ultrasound beamforming problem [179, 46, 48, 180, 20]. Deep models make ill-posed inverse problems tractable thanks to their great potential for approximating non-linear mapping functions between high dimensional training pairs. Convolutional neural networks (CNNs) were adapted to estimate high quality In-phase/Quadrature (IQ) data from delayed RF data [77]. Generative adversarial networks (GANs) were used to mimic eigenspace-based minimum variance (EMV) beamforming [181]. The challenge on ultrasound beamforming with deep learning (CUBDL) was held in conjunction with the 2020 IEEE International Ultrasonics Symposium (IUS) [82]. In [71], deep learning was adapted to construct a general ultrasound beamformer and successfully taken apart in this challenge. This method was designed to mimic minimum variance (MV) beamforming using MobileNetV2 network architecture, and was ranked first in terms of image quality. Overall, considering the network size as well, it was jointly ranked first with another submission [83]. Recently, self-supervised learning [182] as well as Complex CNNs [183] were adapted for plane-wave beamforming. Deep learning has also been used to reconstruct ultrasound images from sub-sampled data [184, 185].

6.2 Inverse problem of ultrasound beamforming

The goal of ultrasound imaging is to form a high-quality spatial map of the medium echogenicity. To do so, L piezoelectric elements of an ultrasound probe transmit an acoustic wave into the medium, and the backscattered waves are collected with N crystal elements of the same probe. Depending on the imaging technique (e.g., line-per-line, plane-wave, or synthetic aperture imaging) and the probe type (e.g., linear, curvilinear, or phased array), this process may be repeated several times to form a single image. Fig. 6.1 shows an example for N-element linear probe, with a transducer pitch of p, in which the backscattered signals are recorded with a specific sampling frequency (f_s), and the beamforming grid is broken into certain numbers of pixels in the axial (i.e., the wave propagation direction) and lateral directions (z, x) with the pixel sizes of $d_z = \frac{c}{2f_s}$ and $d_x = p$, respectively. The speed of sound in the medium is assumed constant and denoted by c. If the time offset following a transmit event equals zero, the actual time corresponding to m^{th} sample of recorded signals is $t = (m-1)/f_s$, where $m = \{1, 2, ..., M\}$.

During the reception, echoes from different pixels might simultaneously arrive at a transducer element and lead to a single output sample if and only if the sum of propagation times for transmitted wave reaching them (τ_t) and getting back to that element (τ_r) was the same. Considering digitization error, all pixels respecting the following condition contribute to that sample of element's output:

$$\mid t - \tau \mid \leq \frac{1}{f_s}, \tag{6.1}$$

where $\tau = \tau_t + \tau_r$ is the propagation delay of each pixel and depends on its location, the probe geometry, the type of transmitted ultrasound wave (e.g., focused, plane-wave, or spherical wave), and assumed speed of sound. As illustrated in Fig. 6.1, Eq. (6.1) results in an elliptical region with varying weights. Therefore, each sample of the RF channel data can be linearly modeled as a combination of the pixels' values in the desired image. The forward model can be written as

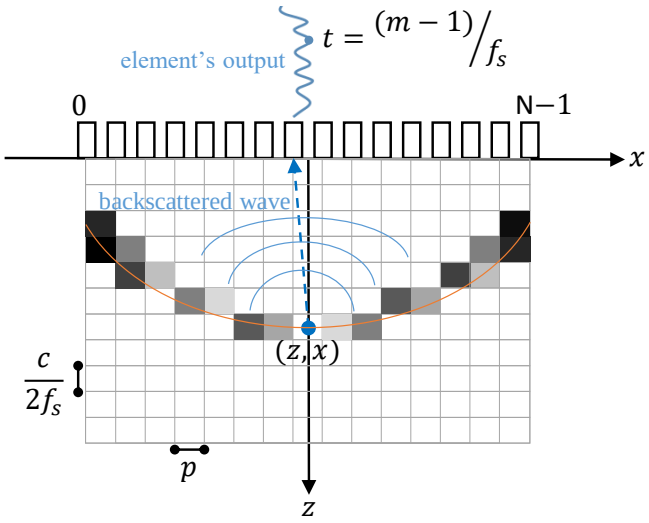

Figure 6.1: The illustration of image pixels' contribution into a single sample of pre-beamformed data.

follows:

$$\mathbf{y} = \Phi\mathbf{x} + \nu, \tag{6.2}$$

where $\mathbf{y}, \mathbf{x} \in \mathbb{R}^{MN}$ are the vectorized versions of pre-beamformed data and desired image, respectively. $\Phi \in \mathbb{R}^{MN \times MN}$ stands for the weighting matrix, and ν is the electronic noise affecting the raw data which has been shown to be well-approximated by additive white Gaussian noise (AWGN) [186].

The way, used in here, for designing matrix Φ has been inspired by the work in [132]. This simple method is based on the assumption of linear propagation in the medium and does not incorporate the Point Spread Function (PSF) of the probe. As shown in Fig. 6.1, each row of matrix Φ includes the contributions of image pixels into a single sample of pre-beamformed data. More specifically, the pixels are weighted using the following equation:

$$\Phi(i,j) = \begin{cases} 1 - \frac{|t_i - \tau_j|}{t_{max}} & \mid t_i - \tau_j \mid \leq \frac{1}{f_s} \\ 0 & \mid t_i - \tau_j \mid > \frac{1}{f_s} \end{cases}, \tag{6.3}$$

where t_{max} is the maximum absolute difference between the actual time (t_i) corresponding to a sample of element's output and the propagation delays (τ_j) of pixels which contribute to that sample (i.e., only the ones respecting the $\mid t_i - \tau_j \mid \leq \frac{1}{f_s}$ condition). Matrix Φ is highly sparse because only a small portion of pixels satisfy Eq. (6.1). Moreover, Φ is data independent and can be precalculated based on the known imaging settings. Finally, matrix Φ is multiplied with a reception apodization matrix, commonly used in DAS beamforming, in which the directionality of transducer elements is taken into account and the f-number is fixed for the entire image. It has to be emphasized that matrix Φ does not necessarily need to be a square matrix because it can be defined for any grid partitioning not equal to pre-beamformed data.

A popular approach, adopted here, to invert Eq. (6.2) and recover the beamformed image $\mathbf{x}$ is to use a variational approach [187], wherein we pose and solve the following optimization problem:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \frac{1}{2} \parallel \mathbf{y} - \Phi\mathbf{x} \parallel_2^2 + \mu \parallel \mathbf{x} \parallel_1, \tag{6.4}$$

where the first term is the data loss that penalizes mismatch to the observed RF channel data, and the second term is the regularization term that promotes the sparsity of beamformed image $\mathbf{x}$. Another way of thinking about this optimization problem is to use a Bayesian prospective in which Eq. (6.4) is equivalent to maximum *a posteriori* estimation of $\mathbf{x}$ with a zero-mean Laplacian prior. Interested readers about the probabilistic interpretation of regularization are referred to [188]. The constant hyperparameter μ controls the contribution of the data fidelity and sparse regularization terms. Note that the choice of the ℓ_1-norm to promote sparsity has been extensively used in ultrasound image reconstruction [131, 178, 132]. However, other regularization have been also used, such as wavelet frames [125], ℓ_2-norm [131], ℓ_p-norms [155], etc. For simplicity, we focus here on ℓ_1-norm and extend it to newly proposed priors.

The objective function presented in Eq. (6.4) is convex but the ℓ_1 term is nondifferentiable and the problem does not have a closed-form solution. First-order proximal splitting algorithms [189] that operate individually on each term are well suited for this optimization. Herein, ADMM is adopted to find the solution of Eq. (6.4), whose convergence has been proven for convex optimization problems [190].

In the reminder of this section, details on solving Eq. (6.4) using ADMM are first outlined. Afterward, beamforming using PnP algorithm is introduced in Section 6.2.2. Finally, Section 6.2.3 describes how Eq. (6.4) is modified and solved based on the RED algorithm.

6.2.1 ADMM solution

Split-variable ADMM is based on optimizing each term of the objective function separately, which is very useful in practice when a single optimization approach cannot be used for all terms [190]. To do so, the independent variable $\mathbf{x}$ is splitted into two variables $\mathbf{u}$ and $\mathbf{v}$ with the constraint that $\mathbf{u} = \mathbf{v}$. Consequently, the new, but fully equivalent, form of Eq. (6.4) is as follows:

$$(\hat{\mathbf{u}}, \hat{\mathbf{v}}) = \underset{(\mathbf{u},\mathbf{v})}{\operatorname{argmin}} \frac{1}{2} \parallel \mathbf{y} - \Phi\mathbf{u} \parallel_2^2 + \mu \parallel \mathbf{v} \parallel_1 \;\; s.t. \; \mathbf{u} = \mathbf{v}, \tag{6.5}$$

where the corresponding unconstrained problem can be written using the augmented Lagrangian approach as following:

$$(\hat{\mathbf{u}}, \hat{\mathbf{v}}, \hat{\lambda}) = \underset{(\mathbf{u},\mathbf{v},\lambda)}{\operatorname{argmin}} \frac{1}{2} \parallel \mathbf{y} - \Phi\mathbf{u} \parallel_2^2 + \mu \parallel \mathbf{v} \parallel_1 - \lambda^T(\mathbf{u} - \mathbf{v}) + \frac{\beta}{2} \parallel \mathbf{u} - \mathbf{v} \parallel_2^2, \tag{6.6}$$

where $\lambda \in \mathbb{R}^{MN}$ is the Lagrange multiplier, and $\beta > 0$ is the weight of penalty term which penalizes violation from the constraint.

The equivalent but more compact form of Eq. (6.6) is as follows [191]:

$$(\hat{\mathbf{u}}, \hat{\mathbf{v}}, \hat{\lambda}) = \underset{(\mathbf{u},\mathbf{v},\lambda)}{\operatorname{argmin}} \frac{1}{2} \parallel \mathbf{y} - \Phi\mathbf{u} \parallel_2^2 + \mu \parallel \mathbf{v} \parallel_1 + \frac{\beta}{2} \parallel \mathbf{u} - \mathbf{v} + \frac{\lambda}{\beta} \parallel_2^2 . \tag{6.7}$$

As mentioned before, ADMM separates the minimization of each term in Eq. (6.7), and finds its solution through an iterative process as summarized in Algorithm 6.1. After setting the hyperparameters μ and β, the iterative algorithm is initialized with arbitrary values for the new variables (i.e., $\mathbf{u}$ and $\mathbf{v}$) as well as the Lagrange multiplier (λ). In each iteration, the cost function presented in Eq. (6.7) is calculated, and once its relative error for two consecutive iterations becomes smaller than a small constant threshold ϵ, the algorithm ends. The proposed ADMM solution includes two main steps as follows.

Algorithm 6.1 Ultrasound beamforming using ADMM

1: **Input:** Φ, $\mathbf{y}$
2: **Set:** $\mu > 0$, $\beta > 0$, $\mathbf{u}^0$, $\mathbf{v}^0$, λ^0, ϵ
3: **While** stopping criterion $> \epsilon$ **do**
4: $\mathbf{u}^{i+1} = \text{argmin}_{\mathbf{u}} \frac{1}{2} \parallel \mathbf{y} - \Phi\mathbf{u} \parallel_2^2 + \frac{\beta}{2} \parallel \mathbf{u} - \mathbf{v}^i + \frac{\lambda^i}{\beta} \parallel_2^2$
5: $\mathbf{v}^{i+1} = \text{argmin}_{\mathbf{v}} \mu \parallel \mathbf{v} \parallel_1 + \frac{\beta}{2} \parallel \mathbf{u}^{i+1} - \mathbf{v} + \frac{\lambda^i}{\beta} \parallel_2^2$
6: $\lambda^{i+1} = \lambda^i + \beta(\mathbf{u}^{i+1} - \mathbf{v}^{i+1})$
7: **End**

6.2.1.1 Beamforming update

The first step corresponds to the minimization of least-squares term written in line 4 of Algorithm 6.1. The optimal solution of this cost function can be easily found by setting the gradient to zero, given that it is an unconstrained and differentiable convex problem. By doing so, the following closed-form solution is obtained:

$$\mathbf{u}^{i+1} = (\Phi^T\Phi + \beta J)^{-1}(\Phi^T\mathbf{y} + \beta\mathbf{v}^i - \lambda^i), \tag{6.8}$$

where J is a matrix of ones with the same size as $\Phi^T\Phi$. This solution, however, involves large matrix inversion which is intractable in practice because $\Phi^T\Phi$ is a square matrix of size several hundreds of thousands. $\Phi^T\Phi$ is not even a diagonal matrix, nor one that can be diagonalizable through Fourier transform. Therefore, numerical methods are used to solve this issue and find the solution of this step. Herein, the limited-memory BFGS solver[1] is used to find the optimal solution. Limited-memory BFGS is a highly efficient quasi-Newton method for unconstrained optimization of differentiable real-valued high-dimensional functions that achieves quadratic convergence for many problems [192].

6.2.1.2 Sparsity and Lagrange multiplier updates

The line 5 of Algorithm 6.1 corresponds to the optimization of sparsity constraint. It is commonly referred to as proximal mapping of the ℓ_1-norm as following [191]:

$$prox_{\mu\|.\|_1/\beta}(\mathbf{u}^{i+1} + \frac{\lambda^i}{\beta}) = \text{argmin}_{\mathbf{v}} \mu \parallel \mathbf{v} \parallel_1 + \frac{\beta}{2} \parallel \mathbf{u}^{i+1} - \mathbf{v} + \frac{\lambda^i}{\beta} \parallel_2^2, \tag{6.9}$$

where the objective function is strictly convex, and its optimum solution can be found using shrinkage function [191], which operates as the soft-thresholding operator:

$$\mathbf{v}^{i+1} = soft_{\mu/\beta}(\mathbf{u}^{i+1} + \frac{\lambda^i}{\beta}) = max\{|\mathbf{u}^{i+1} + \frac{\lambda^i}{\beta}| - \frac{\mu}{\beta}, 0\} sign(\mathbf{u}^{i+1} + \frac{\lambda^i}{\beta}). \tag{6.10}$$

Finally, the 6th line of Algorithm 6.1 entails updating the Lagrangian multiplier.

6.2.2 PnP solution

Our primary motivation for using ADMM is to construct a flexible framework which can be extended to PnP and RED algorithms. By closely looking at step 5 of Algorithm 6.1, it can be associated to

[1]MATLAB implementation is publicly available in this link: `http://www.cs.ubc.ca/~schmidtm/Software/minFunc.html`

Algorithm 6.2 Ultrasound beamforming using PnP

1: **Input:** Φ, $\mathbf{y}$
2: **Set:** $\beta > 0$, $\mathbf{u}^0$, $\mathbf{v}^0$, λ^0, ϵ
3: **While** stopping criterion $> \epsilon$ **do**
4: $\mathbf{u}^{i+1} = \text{argmin}_{\mathbf{u}} \frac{1}{2} \parallel \mathbf{y} - \Phi\mathbf{u} \parallel_2^2 + \frac{\beta}{2} \parallel \mathbf{u} - \mathbf{v}^i + \frac{\lambda^i}{\beta} \parallel_2^2$
5: $\mathbf{v}^{i+1} = \mathcal{F}(\mathbf{u}^{i+1} + \frac{\lambda^i}{\beta})$
6: $\lambda^{i+1} = \lambda^i + \beta(\mathbf{u}^{i+1} - \mathbf{v}^{i+1})$
7: **End**

a variational denoising problem wherein $(\mathbf{u}^{i+1} + \frac{\lambda^i}{\beta})$ is the observation and $\mathbf{v}$ is the clean image to be estimated. Thanks to the modular structure of the ADMM, step 5 can be accomplished not only using the soft-thresholding function explained before, but any other approach which removes the noise term of the observation is also applicable. This point is the core idea of PnP approach [157] in which the choice of regularizer (ℓ_1-norm in Eq. (6.4)) is replaced by the choice of any sophisticated denoising algorithm.

Motivated by the success and wide applications of nonlocal means (NLM) denoiser in ultrasound imaging [193], the shrinkage function is replaced herein with NLM algorithm, which is grounded in two reasons. First, NLM has a low number of hyperparameters which are easy to tune. Second, the noise type of ultrasound RF data is additive Gaussian mainly brought about by the sensor noise and the acquisition card [186]. Therefore, the Gaussian-weighted Euclidean distance can be reliably used within NLM algorithm to assess the similarity between image patches.

Our PnP beamforming approach is summarized in Algorithm 6.2, wherein the NLM denoiser, denoted by $\mathcal{F}$, is applied to the observation $(\mathbf{u}^{i+1} + \frac{\lambda^i}{\beta})$ in each iteration. Except for line 5, the rest of the algorithm is exactly the same as ADMM while the hyperparameter μ does not exist anymore.

6.2.3 RED solution

The denoiser used in the PnP algorithm does not relate to an explicit regularizer, and as such, the loss function is not explicitly defined. Romano *et al.* recently proposed another way, called RED, to exploit denoising algorithms for regularization [159] in which an explicit regularizer is defined as follows:

$$\rho(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T(\mathbf{x} - \mathcal{F}(\mathbf{x})), \tag{6.11}$$

where ρ is designed to enforce the orthogonality of the image ($\mathbf{x}$) and what a denoiser removes from the image ($\mathbf{x} - \mathcal{F}(\mathbf{x})$). The main advantage of having an explicit regularizer in RED compared to PnP is that its theoretical convergence properties can be analyzed and different optimization

Algorithm 6.3 Ultrasound beamforming using RED

1: **Input:** Φ, $\mathbf{y}$
2: **Set:** $\mu > 0$, $\beta > 0$, $\mathbf{u}^0$, $\mathbf{v}^0$, λ^0, ϵ
3: **While** stopping criterion $> \epsilon$ **do**
4: $\mathbf{u}^{i+1} = \text{argmin}_{\mathbf{u}} \frac{1}{2} \parallel \mathbf{y} - \Phi\mathbf{u} \parallel_2^2 + \frac{\beta}{2} \parallel \mathbf{u} - \mathbf{v}^i + \frac{\lambda^i}{\beta} \parallel_2^2$
5: $\mathbf{v}^{i+1} = \text{argmin}_{\mathbf{v}} \frac{\mu}{2}\mathbf{v}^T(\mathbf{v} - \mathcal{F}(\mathbf{v})) + \frac{\beta}{2} \parallel \mathbf{u}^{i+1} - \mathbf{v} + \frac{\lambda^i}{\beta} \parallel_2^2$
6: $\lambda^{i+1} = \lambda^i + \beta(\mathbf{u}^{i+1} - \mathbf{v}^{i+1})$
7: **End**

algorithms can be used to solve the problem [165]. Furthermore, Romano *et al.* has shown numerical evidence that denoising algorithms are locally homogeneous under certain conditions which helps to compute the gradient expression of regularizer $\rho_{red}(.)$ as follows [159]:

$$\nabla\rho(\mathbf{x}) = \mathbf{x} - \mathcal{F}(\mathbf{x}). \tag{6.12}$$

Herein, the proposed ADMM framework for ultrasound beamforming is extended to the RED algorithm. To do so, ℓ_1-norm in Eq. (6.4) is substituted by ρ in Eq. (6.11), and the resulting optimization problem is as follows:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \frac{1}{2} \parallel \mathbf{y} - \Phi\mathbf{x} \parallel_2^2 + \frac{\mu}{2}\mathbf{x}^T(\mathbf{x} - \mathcal{F}(\mathbf{x})), \tag{6.13}$$

where its iterative solution is summarized in Algorithm 6.3. The only difference compared to original ADMM is the fifth line involving the new regularization term. Since ρ is differentiable, the solution $\mathbf{v}^{i+1}$ must obey the fixed point relationship:

$$\begin{aligned}\mu\nabla\rho_{red}(\mathbf{v}) - \beta(\mathbf{u}^{i+1} - \mathbf{v} + \frac{\lambda^i}{\beta}) = \mu(\mathbf{v} - \mathcal{F}(\mathbf{v})) - \beta(\mathbf{u}^{i+1} - \mathbf{v} + \frac{\lambda^i}{\beta}) = 0 \\ \Leftrightarrow \mathbf{v^{i+1}} = \frac{\mu}{\mu+\beta}\mathcal{F}(\mathbf{v^{i+1}}) + \frac{\beta}{\mu+\beta}\mathbf{u^{i+1}} + \frac{1}{\mu+\beta}\lambda^{\mathbf{i}},\end{aligned} \tag{6.14}$$

where the solution $\mathbf{v}^{i+1}$ is a function of its denoised version $\mathcal{F}(\mathbf{v^{i+1}})$. Therefore, an approximation of $\mathbf{v}^{i+1}$ can be obtained by iterating:

$$\mathbf{z^k} = \frac{\mu}{\mu+\beta}\mathcal{F}(\mathbf{z^{k-1}}) + \frac{\beta}{\mu+\beta}\mathbf{u^{i+1}} + \frac{1}{\mu+\beta}\lambda^{\mathbf{i}}. \tag{6.15}$$

over $k = 1, ..., K$ iterations with sufficiently large K. The previous value is used to initialize $\mathbf{z}$ (i.e., $\mathbf{z^0} = \mathbf{v^i}$). Similar to the PnP approach, NLM denoiser is used in the RED algorithm proposed in this work for ultrasound imaging.

6.3 Experiments

The experimental part is designed to provide a clear understanding of the advantages as well as the limitations of the proposed method. In this section, first, details regarding datasets on which the proposed method is evaluated and compared with other approaches are explained. Then, different criteria used for quantitative evaluation of results are introduced. Herein, our dataset is exactly the same as what we have explained in Section 5.2.1.

6.3.1 Evaluation metrics

The reconstructed images using different beamforming methods are evaluated and compared in terms of resolution and contrast which are the main specialized ultrasound assessment indexes. Moreover, the performance of our method in preserving the speckle quality is also investigated.

In ultrasound imaging, speckle is an important feature, and as such, it is crucial to preserve it during reconstruction. For the fully developed speckle regions, the intensity of resulting envelope image follows a Rayleigh distribution. Herein, a Kolmogorov–Smirnov (KS) test, designed by PICMUS organizers, is used to verify whether the data follows a Rayleigh distribution or not. In a region of image, the significance level of $\alpha = 0.05$ is considered to decide whether the speckle statistics are preserved or not.

Table 6.1: Quantitative results in terms of resolution and contrast indexes for simulation and real phantom experiments. The KS columns indicate whether the method preserves speckle texture or not, which are indicated by ✓and ✗ marks, respectively.

dataset	SR		ER			SC			EC		
index	$FWHM_A$	$FWHM_L$	$FWHM_A$	$FWHM_L$	KS	CNR	gCNR	KS	CNR	gCNR	KS
DAS	0.4	0.47	0.48	0.8	✓	10.25	0.89	✓	8.8	0.87	✓
DAS_CPWC	0.4	0.4	0.49	0.55	✓	17.53	0.99	✓	13.25	0.97	✓
ADMM	0.38	0.39	0.47	0.74	✗	9.01	0.88	✗	6.85	0.72	✗
PnP	0.29	0.43	0.48	0.75	✓	14.87	0.93	✓	16.45	0.99	✓
RED	0.37	0.46	0.48	0.76	✓	15.48	0.94	✓	14.7	0.98	✓
RED_CPWC	0.36	0.26	0.46	0.36	✓	18	1	✓	15	1	✓
EMV	0.4	0.1	0.59	0.42	✓	11.21	0.93	✓	8.1	0.83	✓
PCF	0.3	0.38	0.46	0.41	✗	5.64	0.76	✗	3.2	0.68	✗
MNV2	0.42	0.27	0.53	0.77	✓	10.48	0.89	✓	7.8	0.83	✓
Stolt	0.42	1.1	0.55	0.41	✓	2.1	0.55	✗	6.55	0.78	✓
UFSB	0.4	0.85	0.55	0.52	✓	7.3	0.78	✗	5.96	0.76	✓

6.4 Results

This section is dedicated to the performance evaluation of proposed ADMM, PnP, and RED approaches. The images, introduced in Section 5.2.1, are reconstructed by solving the inverse problem of beamforming using the algorithms 6.1, 6.2, and 6.3. Furthermore, the advantages of our methods are demonstrated in comparison with other approaches for solving the problem of ultrasound beamforming. Among the time-domain approaches, the classical DAS, EMV [2], and phased coherence factor (PCF) [3] are selected for the sake of comparison. The results of Fourier domain technique based on Stolt's migration [5], and ultrasound Fourier slice beamforming (UFSB) [4] are also included. The comparison is completed by incorporating the results of our deep learning method, referred to as MNV2 [71]. Moreover, the beamforming methods are also compared in terms of the reconstruction time. Finally, the sensitivity of proposed methods to parameter selection and algorithm initialization is also analyzed.

Hereafter, the Hanning apodization window with f-number equals to 0.5 is always considered as reception apodization matrix in our method, DAS, and other methods on top of DAS (except for *in vivo* datasets for which the Tukey (tapered cosine) window with constant parameter of 0.25 and f-number equals to 1.75 is considered). The search and comparison window sizes of NLM denoiser are set to 21 and 5, respectively. The reliable built-in MATLAB implementation of NLM is used in which the standard deviation of noise estimated from the image is used as degree of smoothing. This helps to have adaptive denoising with less parameters. Details regarding the method used for the estimation of noise variance can be found in [194]. The threshold for stopping criterion is set to $\epsilon = 10^{-3}$. The initial values equal to zero are selected for all iterative algorithms. The hyperparameters of each method are tuned independently to get the best results. Quantitative indexes are separately calculated for each cyst region or point target in the image, and the average values are reported.

6.4.1 Beamforming results with the proposed approaches

6.4.1.1 Simulation and experimental data

Fig. 6.2 displays the beamforming results from the single 0^o plane-wave as well as Coherent Plane-Wave Compounding (CPWC) on 75 insonifications. The quantitative evaluation is also summarized in Table 6.1. As results confirm, the PnP and RED approaches substantially improve the contrast

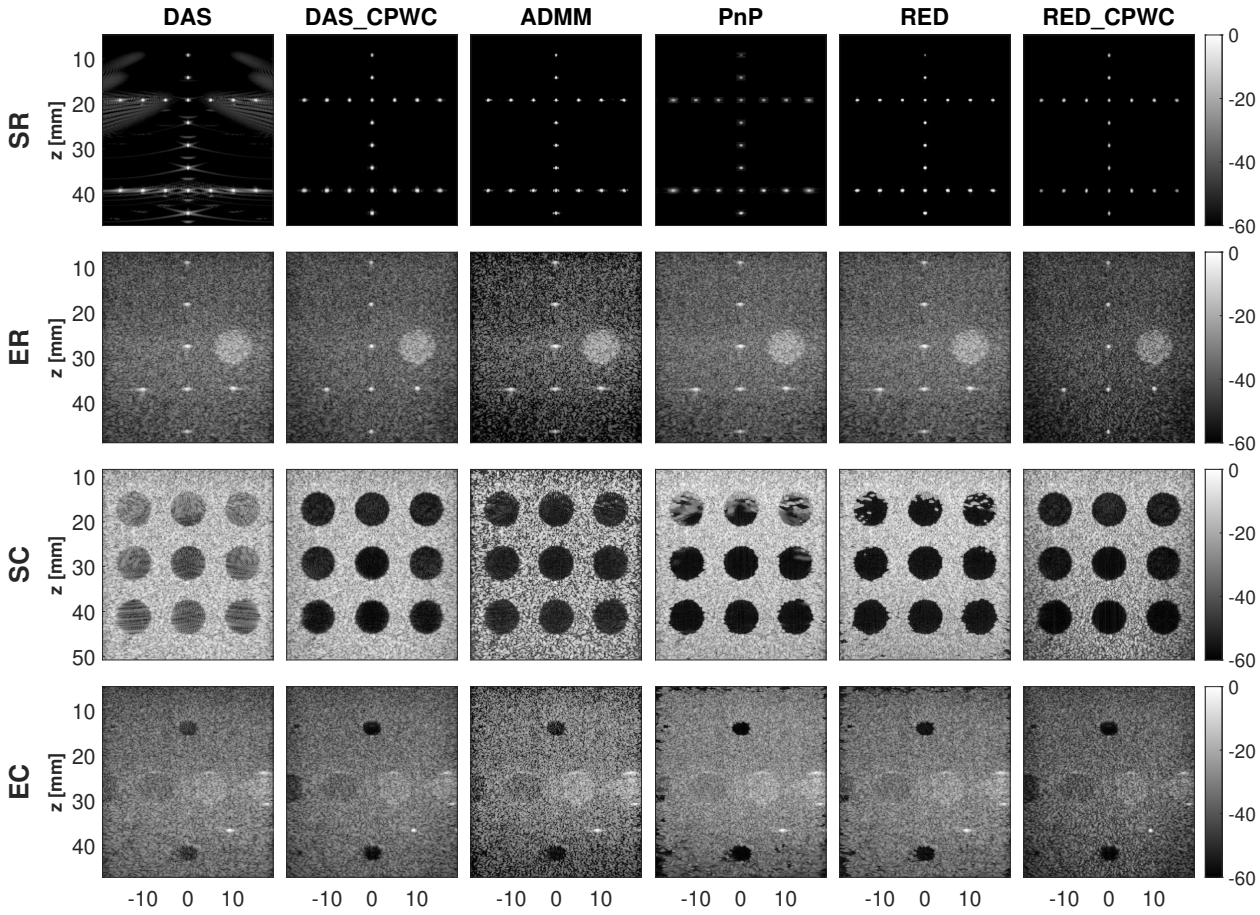

Figure 6.2: The results of simulation and experimental datasets. Rows indicate datasets while columns correspond to different approaches. All results are from a single 0^o plane-wave insonification except for CPWC which is obtained from 75 steered insonifications.

and perfectly suppresses the side-lobe artifacts similar to DAS_CPWC which averages the DAS results over 75 angles. The ADMM approach fails the KS test due to the soft-thresholding operator while PnP as well as RED algorithms replace it with denoising and preserve the speckle statistics.

To apply the proposed approaches on different insonification angles, the process is as follows. First, the matrix Φ is independently calculated for each of the transmitting angles, given that the traveling time of the tilted plane-wave to the scatterers is different. Then, the proposed algorithms are independently performed for each angle. Finally, the resulting RF data (before envelope detection) are averaged to coherently compound all plane waves. Since the procedure is exactly the same for ADMM, PnP, and RED algorithms, we only present the results of RED algorithm, on 75 insonification angles, in order to keep the chapter concise. The RED_CPWC results confirm that the method is extensible to different angles, and CPWC improves the results as compared to a single 0^o insonification.

The proposed methods marginally improve the resolution as compared to DAS. This result is expected because the focus of ADMM, PnP, and RED algorithms is on the regularization term which mainly affects the contrasts of results. The resolution, however, is rooted in the measurement model and mainly depends on the matrix Φ. Nevertheless, the visual inspection of SR image reveals that PnP method makes the point targets blurry due to the averaging of image patches in NLM

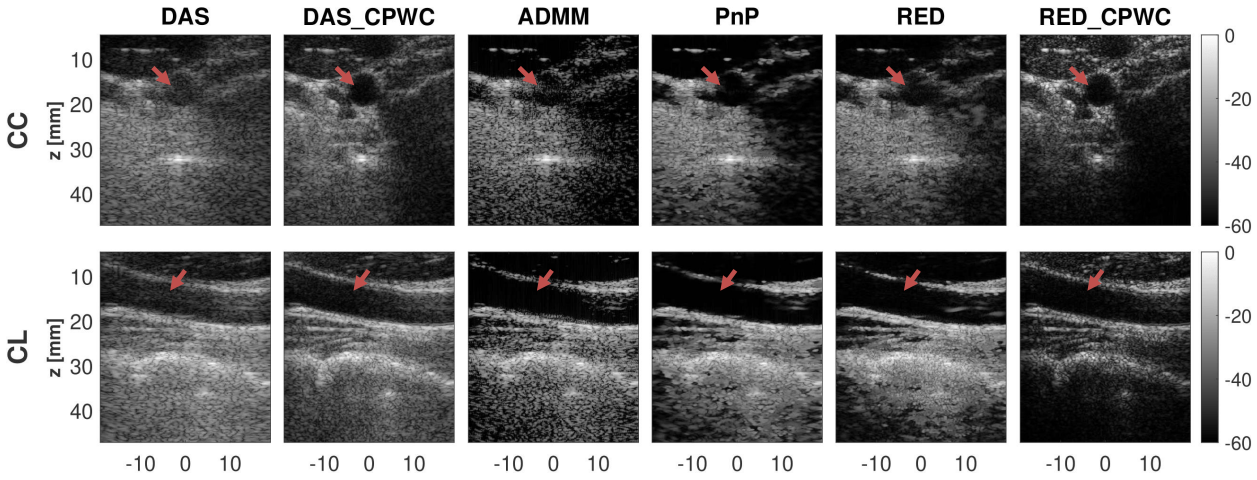

Figure 6.3: The results of *in vivo* datasets. Rows indicate datasets while columns correspond to different approaches. All results are from a single 0^o plane-wave insonification except for CPWC which is obtained from 75 steered insonifications.

denoiser. RED algorithm does not suffer from this problem since the denoiser is used to define an explicit regularizer rather than directly applying on the image.

6.4.1.2 *In vivo* data

Fig. 6.3 illustrates the cross-sectional and longitudinal views from the carotid artery. The visual comparison of different approaches confirm that the clutter artifacts caused by diffuse reverberation from shallow layers are suppressed by proposed algorithms, and dark images of the artery are reconstructed in both views (pointed out by red arrows).

6.4.2 Comparison with existing beamformers

Fig. 6.4 shows the results of existing state of the art beamforming methods with the quantitative results reported in Table 6.1. Overall, it can be verified that EMV beamformer is of the highest axial resolution among all approaches including ours. But the proposed RED algorithm is the best in terms of contrast. Another advantage of RED is that similar contrast indexes are achieved for both simulation and experimental phantoms. However, EMV performance noticeably drops in real experiments as compared to simulation. This difference is mainly brought about by the selection of eigenvectors of the covariance matrix used for making the signal subspace. More specifically, the SR data includes anechoic background from which the point targets can be recovered by only using the principal eigenvectors. However, the reconstruction of ER image requires considering all eigenvector in order not to lose the speckle texture, which makes the results of EMV exactly the same as MV. This difference is also visible for contrast because experimental data contains an additional noise reducing the quality of estimated covariance matrix.

In terms of computational time, Fourier domain beamforming techniques are the fastest options, and the iterative approaches are the slowest ones. Herein, we compare the reconstruction time of different algorithms for EC experiment. The Stolt's migration and UFSB approaches take 100 milliseconds and 2.92 seconds, respectively. Among the time-domain approaches, DAS and PCF have a similar speed and take 1.6 seconds to reconstruct the image. The EMV method is much slower as it needs 8 minutes to form the image mainly due to the covariance matrix estimation

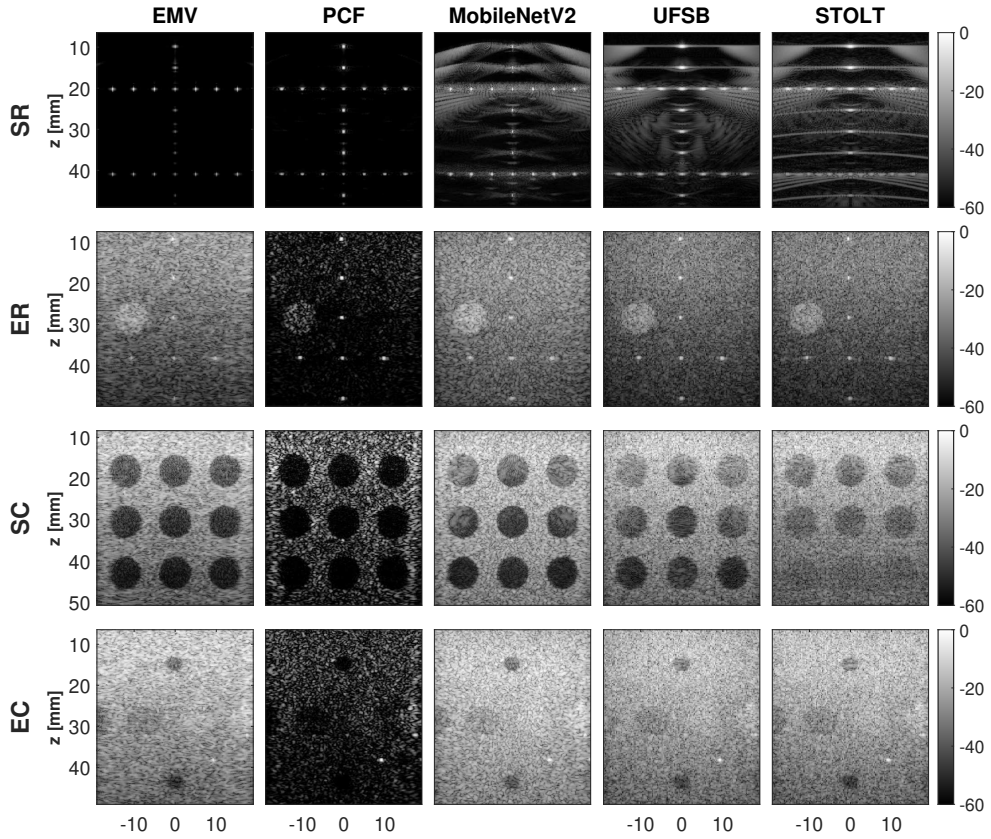

Figure 6.4: Simulation and experimental images reconstructed through previous beamforming methods. Rows indicate datasets while columns correspond to different approaches. All results are from a single 0^o plane-wave insonification.

and its decomposition. MNV2 is the optimized version of MV which reduces the time to 40.2 seconds. The proposed ADMM, PnP, and RED approaches are even slower than EMV mainly due to the numerical optimization method used for optimizing the measurement loss and avoiding intractable matrix inversion. Although all of the proposed approaches converge in less than 15 iterations, the required time is around 20 minutes because each iteration contains an inner iterative algorithm (BFGS). It has to be mentioned that all these numbers are with straightforward Matlab implementations without any runtime optimization.

6.4.3 Sensitivity analysis

As mentioned in Algorithms 6.1, 6.2, and 6.3, initial values must be set for new variables (i.e., $\mathbf{u}$ and $\mathbf{v}$) and the Lagrange multiplier (λ). Moreover, the hyperparameters μ and β, which respectively determine the coefficient of regularizer and penalty term, should be specified.

Fortunately, whatever initialization for new variables and the Lagrange multiplier do not change the final solution since Eq. (6.4) is convex and does not have any local minimum. But the number of iterations required for convergence might alter if outlying values are selected. Herein, the initial points are always set to zero.

As for the hyperparameters, a large β (equals to 1000) is set in order to perfectly accomplish the equality constraints of new variables (Eq. (6.5)) and converge toward equal values for $\mathbf{u}$ and $\mathbf{v}$. While the regularizer coefficient μ is not used within PnP, it specifies the threshold of Eq. (6.10)

in ADMM algorithm. Large μ makes the ADMM results too dark since it wipes out the speckle texture while small μ cause a poor contrast. In RED approach, μ must be set to high values (between 1000 to 5000) as the denoiser residual ($\mathbf{x} - \mathcal{F}(\mathbf{x})$) is of a negligible amplitude. RED has also an extra parameter K which specifies the number of iterations in Eq. (6.15). Although sufficiently large K is required for the exact RED algorithm, our investigations demonstrate that many inner iterations are unnecessary and $K = 1$ is enough to get proper results. This behavior has also been reported in [165].

6.5 Discussion

The efficacy of inverse problem formulation of ultrasound beamforming has already been shown in several studies [131, 132, 125, 178]. The main issue, however, is the loss of speckle texture in the proximal mapping step. The proposed RED algorithm is a reliable solution to this problem. The RED approach not only takes advantage of denoising for image reconstruction but also explicitly minimizes a well-defined regularizer. RED's ability to preserve the speckle information is of crucial significance in image computing applications such as quantitative ultrasound and speckle tracking.

Once the medium is insonified, the spherical waves get backscattered toward the probe elements. Signals originating from pixels located in shallow regions are recorded by a small subset of elements close to the pixel location, while the backscattered signals originating from the subordinate pixels are properly recorded by most of the elements. That is why matrix Φ is multiplied with a reception apodization matrix to keep the f-number fixed for the entire image depths. Therefore, the anechoic cysts of shallow regions are reconstructed with a small part of probe elements which causes a lower contrast and brings some artifacts, as seen in the results of SC datasets. Nevertheless, those artifacts disappear in the result of CPWC where more data is available.

The modular property of ADMM helped us design a single framework for three different approaches of solving the inverse problem of ultrasound beamforming. Besides changing the prior term, ADMM provided the possibility of optimizing the measurement loss using limited-memory BFGS in which the Hessian matrix is approximated. This property is essential in our case wherein the calculation of Hessian is intractable due to the large size of matrix Φ.

Although we have only presented the results for NLM denoiser, another important advantage of the proposed PnP and RED approaches is to open a way to incorporate any denoiser algorithm in the inverse problem of beamforming which is our plan for future work. This point is worth noting because the range of denoiser options is much larger than the choice of regularization functions.

Although using the noise standard deviation, estimated from the image, as degree of smoothing helps to have adaptive denoising with less parameters, it may cause loss of structural details in the low quality regions of the image. As seen in top regions of the in vivo datasets as well as top and left edges of the EC dataset, the image quality is low and the noise standard deviation is overestimated in those regions, which results in stronger smoothing and removes some structural details. This point is unavoidable unless the degree of smoothing is controlled manually, or a better noise variance estimation algorithm is adopted. The latter represents an interesting avenue for future work.

Achieving the same contrast as CPWC can be considered as a step toward eliminating the necessity of transmitting several plane-waves with different angles. However, the improvement in resolution is minor and we still need to extend the measurement model and the way matrix Φ is defined, which is the subject of our future research.

Our formulation can be applied to other imaging techniques (such as focused and synthetic aperture imaging) or even other types of ultrasound transducer (such as convex and phased

array). Herein, we only consider reporting the results on benchmark PICMUS dataset since it is available online which makes the comparison with previous approaches easier. It also facilitates reimplementing the methods and verifying the results.

6.6 Conclusions

Denoising algorithms have been recently adopted in solving the inverse problem of imaging. Herein, we proposed a novel framework for incorporating denoisers in medical ultrasound beamforming. Our framework is based on the ADMM wherein a linear forward model is used for the image under scrutiny, and three solutions are found by considering Laplacian, PnP, and RED priors. The results show that the proposed RED approach gives the best images quality with a high contrast while the speckle information is also preserved.

Chapter 7

A Unifying Approach to Inverse Problems of Ultrasound Beamforming and Deconvolution

This chapter is based on our published paper [154].

Although a better beamformer improves the system PSF, ultrasound imaging still has a non-ideal PSF due to many factors such as the limited bandwidth of piezoelectric crystal elements, the physical phenomena of acoustic wave propagation in the tissue, etc. Under the assumption of weak scattering for soft tissues and using the first-order Born approximation, the ultrasound Radio-Frequency (RF) data can be linearly modeled as the result of convolution between the ground-truth Tissue Reflectivity Function (TRF) and the PSF of the ultrasound imaging system [195, 196]. Therefore, another line of research has been devoted to mitigating the adverse effect of non-ideal PSF using the deconvolution approach [195, 197, 169]. To our knowledge, deconvolution has only been applied as a post-processing approach after reconstructing the ultrasound image based on existing beamforming procedures, essentially DAS.

As shown in Fig. 7.1, the current chapter is motivated by the novel idea of unifying the beamforming and deconvolution steps together to simultaneously take advantage of both methods in image reconstruction. The proposed framework is a joint inverse problem including two linear models of beamforming and deconvolution plus an additional sparsity constraint. In other words, our method is designed to estimate the desired image directly and concurrently minimize the adverse effect of the PSF. The resulting optimization problem is solved using split-variable Alternating Direction Method of Multipliers (ADMM) algorithm [190, 198] as it allows the minimization of each term of the objective function separately.

7.1 Related work

In the rich body of literature on the topic of ultrasound beamforming and deconvolution, the most recent and relevant studies are reviewed in this section.

Beamforming approaches can be categorized into four main groups. The first set is time-domain methods, among which DAS is the most popular non-adaptive method. Filtered-delay multiply and sum (F-DMAS) was proposed in [152]. The algorithm is based on a pairwise multiplication of delayed RF signals before summation. Another extension to DAS has recently been proposed based

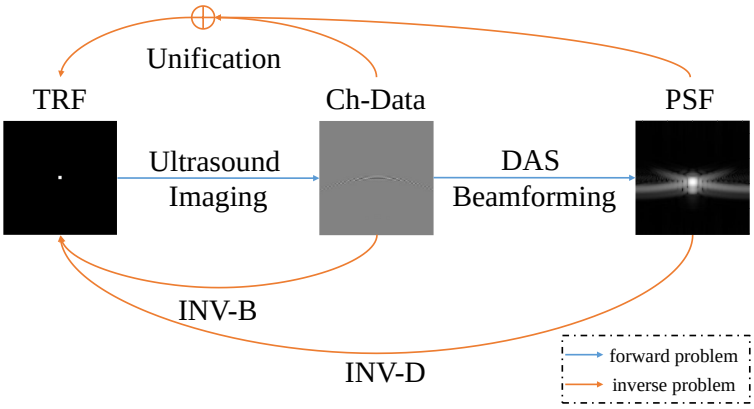

Figure 7.1: The illustration of the proposed unifying approach. Inverse problem of beamforming (INV-B) directly estimates the desired image from pre-beamformed channel data (Ch-Data) while inverse problem of deconvolution (INV-D) recovers the Tissue Reflectivity Function (TRF) from the DAS output. Herein, both inverse problems are jointly solved in order to reconstruct the desired TRF.

on null subtraction imaging [199], wherein envelope images reconstructed by different apodization windows are linearly combined in order to overcome the classical trade-off and have both a low side-lobe level and a narrow main-lobe.

Minimum Variance beamforming (MVB) [200] is the most potent approach among adaptive algorithms that mainly improves the lateral resolution perpendicular to the wave propagation called axial direction in ultrasound imaging [201]. In MVB, the main challenge is estimating the covariance matrix from the data, making it computationally expensive [73]. The MVB has also been extended using singular value decomposition (SVD) of the covariance matrix to improve the contrast [2]. But there is no clear criterion for removing small eigenvalues, and a part of speckle texture might be omitted in this method. To speed up MVB, Nilsen and Hafizovic [202] proposed beamspace MVB. Their method is based on the extraction of spatial statistics from a set of orthogonal beams formed in a different direction. This idea has later been used to extend MVB using multibeam covariance matrices [203]. A fast version of MVB has been developed based on principal component analysis (PCA) [204] as well as Legendre polynomials [205]. In [117], the apodization weights are estimated using independent component analysis (ICA).

There is another type of adaptive method, in the time domain, based on the coherence factor (CF), which is defined as the ratio of coherent to incoherent energy across the aperture [206]. CF was used as an adaptive weight on top of DAS to improve the image quality [151]. Generalized CF (GCF) was derived from the spatial spectrum and defined as the ratio between the energy of a predefined low-frequency range (the coherent portion of RF data) to the total spectral energy [145]. Subsequently, phased CF (PCF) was proposed based on the phase, rather than amplitude information of aperture [3].

The second group of beamforming methods is implemented using the Fourier transform. The pioneer studies were based on synthetic aperture focusing [207, 208]. Later on, the Fourier beamforming was extended for plane-wave Imaging [128] and implemented through different strategies [129, 130, 209, 5]. Wagner *et al.* proposed compressed beamforming that works on the sub-Nyquist RF data [153]. Consequently, this idea has been extended as a general beamformer in Fourier domain [177, 4, 210].

Szasz *et al.* proposed the third group of beamforming methods [131], which assume a linear

model between the observed data (i.e., the RF channel data) and the desired image to be recovered. Beamforming is then performed by solving a regularized inverse problem for each image depth separately. The results have been presented for focused [131] as well as plane-wave imaging [178]. Subsequently, this idea was further extended by considering more regularizations in the objective function and reconstructing all image depths jointly [132]. Two matrix-free formulations have been proposed in [125] that are both faster and more memory efficient than other inverse problem formulations. Recently, denoising-based regularization terms have been adapted in the inverse problem of ultrasound beamforming [149], which noticeably improves the contrast and preserves the speckle statistics.

The last group of ultrasound beamforming methods is based on deep learning [46, 48, 82, 20]. While deep models have great potential for estimating non-linear mapping functions between high dimensional input-output pairs and solving ill-posed problems, deep beamformers are subject to the following limitations. First, deep learning requires a massive amount of training data which is commonly unavailable. Second, the training ground truth is not known specifically for *in vivo* data. Third, there is a noticeable performance reduction on test data due to domain shift between training and test data. The challenge on ultrasound beamforming with deep learning (CUBDL) was organized in conjunction with the 2020 IEEE International Ultrasonics Symposium (IUS) [82]. In [71], a general ultrasound beamformer was designed, based on deep learning, to mimic MVB. In terms of image quality, it was ranked first. Overall, considering the network size as well, it was jointly ranked first with another submission [83].

Finally, in ultrasound image deconvolution, Taxt *et al.* proposed a 2-D blind homomorphic approach wherein the PSF is estimated in the complex cepstrum domain followed by Wiener filtering for the deconvolution [211]. An approach based on parametric inverse filtering was proposed in [197]. Subsequently, Yu *et al.* introduced a single-input multiple-output (SIMO) channel model for deconvolution of ultrasound images [212]. Two frameworks of Compressive Sensing (CS) and deconvolution were combined in [169] and the resulting method is called compressive deconvolution. An analytical model for the spatially-varying PSF in ultrasound imaging is proposed in [213]. A physical model for the nonstationary blur in plane- and diverging-wave imaging is proposed in [6]. Recently, the nonlinearity of ultrasound wave propagation in the tissue was considered in the deconvolution problem, and the enhanced image was reconstructed by the minimization of a joint cost function including the deconvolution models for both fundamental and harmonic RF images [170]. Most of the deconvolution studies reviewed here are categorized as blind methods as they are based on the estimation of PSF of the imaging system, while non-blind methods [214] assume that the PSF is known (e.g., through experimental measurement).

7.2 Background

7.2.1 Inverse problem of ultrasound beamforming

The purpose of ultrasound beamforming is to reconstruct a high-quality spatial map of the medium echogenicity. Without loss of generality, let us consider a N-element ultrasound linear probe with a transducer pitch of p, as shown in Fig. 6.1, from which L piezoelectric elements transmit an acoustic wave into a medium with the constant sound speed of c. We also assume that backscattered signals are recorded with all elements of the same probe with a specific sampling frequency (f_s). To form a single image, this process may be repeated several times depending on the probe type (e.g., linear, phased array, or curvilinear) and the imaging technique (e.g., plane-wave, line-per-line, or synthetic aperture imaging). The beamforming grid is partitioned with a pixel sizes of $d_z = \frac{c}{2f_s}$ in the axial (i.e., the wave propagation direction) and $d_x = p$ in the lateral directions.

If there is no time offset after wave transmission (i.e., the elements immediately start to collect the backscattered waves), m^{th} sample of the elements' outputs corresponds to the actual time of $t = (m-1)/f_s$, where $m = \{1, 2, ..., M\}$. In order to trace back the echoes corresponding to each pixel, the associated time delay (τ) equal to the sum of two-way propagation times of transmitted wave reaching that pixel (τ_t) and getting back to the transducer elements (τ_r) needs to be applied to each signal recorded by piezoelectric elements of the probe. Considering digitization error, the following condition determines all pixels contributing to a sample of element's output:

$$\mid t - \tau \mid \leq \frac{1}{f_s}, \tag{7.1}$$

where $\tau = \tau_t + \tau_r$ depends on pixel location, assumed speed of sound, the type of transmitted ultrasound wave (e.g., plane-wave, focused, or spherical wave), and the probe geometry. Eq. (7.1) leads to an elliptical region with different weights as illustrated in Fig. 6.1. Therefore, a linear forward model between each sample of the RF channel data and the pixels' values in the desired image can be written as follows:

$$\mathbf{y}_{ch} = \Phi \mathbf{x} + \nu, \tag{7.2}$$

where $\mathbf{x}, \mathbf{y}_{ch} \in \mathbb{R}^{MN}$ are the vectorized versions of the desired image and the collected pre-beamformed channel data, respectively. $\Phi \in \mathbb{R}^{MN \times MN}$ is the weighting matrix, and ν stands for the additive white Gaussian noise (AWGN).

The way matrix Φ is designed has been described in our previous work [149]. In this method, the linear propagation in the medium is assumed and PSF of the probe is not incorporated. As Fig. 6.1 and Eq. (7.2) depict, rows of matrix Φ contain the contributions of image pixels into samples of RF channel data. For every single sample of pre-beamformed data, the pixels' contributions are determined using the following equation:

$$\Phi(i,j) = \begin{cases} 1 - \frac{|t_i - \tau_j|}{t_{max}} & \mid t_i - \tau_j \mid \leq \frac{1}{f_s} \\ 0 & \mid t_i - \tau_j \mid > \frac{1}{f_s} \end{cases}, \tag{7.3}$$

where t_i is the actual time corresponding to a sample of element's output, and τ_j is the propagation delays of pixels which contribute to that sample (i.e., only the ones respecting the condition of Eq. (7.1)). t_{max} is the maximum absolute difference between t_i and τ_j. Since only a small portion of pixels satisfy Eq. (7.1), matrix Φ becomes highly sparse. Furthermore, data independent matrix Φ can be precalculated based on the known imaging settings. A reception apodization matrix, commonly used in DAS beamforming, is also multiplied with matrix Φ in order to take into account the directionality of transducer elements and fix the f-number for the entire image depths. It has to be mentioned that matrix Φ is not necessarily a square matrix because it can be determined for any grid partitioning not equal to pre-beamformed data. More details regarding the construction of matrix Φ can be found in [149].

The most straightforward inverse problem formulation of ultrasound beamforming is to estimate $\mathbf{x}$ by solving the following least-squares optimization problem:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \parallel \mathbf{y}_{ch} - \Phi \mathbf{x} \parallel_2^2 . \tag{7.4}$$

In contrast to common beamforming approaches (such as DAS, MVB, CF, etc), there is no need to estimate the apodization weights, and the beamformed image is directly reconstructed. It has been shown in [131, 132] that Eq. (7.4) does not solely provide the best results, and additional constrains should be considered. Furthermore, Eq. (7.4) can be sequentially solved for each depth of the image

(as performed in [131]) or solved at once for the whole image depths (as performed in [132]). Herein, we follow the second approach because of two main reasons. First, independently solving the inverse problem for each depth increases the computational cost of the algorithm noticeably. Second, we have observed border artifacts on the resulting image when different depths are reconstructed separately.

7.2.2 Deconvolution of ultrasound images

Under the assumption of weak scattering for soft tissues and using the first-order Born approximation, the linear model, given in Eq. (7.2), can also be used to express beamformed RF image as the result of convolution between the TRF and the PSF of the ultrasound imaging system as following [169, 195, 196]:

$$\mathbf{y}_{DAS} = H\mathbf{x} + \nu, \tag{7.5}$$

where $\mathbf{y}_{DAS}$ is the RF image resulting from DAS beamforming. H is a block circulant with circulant block (BCCB) matrix formed based on the PSF and accouting for circulant convolution. Although the assumption of the convolution model may not be valid in practice (especially for biological tissues), it has long been shown that the linear model is a good approximation and helps to reduce the adverse effect of PSF through deconvolution. Same as before (i.e., Eq. (7.4)), we consider the inverse problem formulation of deconvolution for finding the desired TRF.

The PSF is usually spatially variant in the ultrasound images mainly due to wave divergence, attenuation, and a limited number of crystal elements in the lateral direction. There are a few settings such as time gain compensation (TGC) and transmitting several focused beams (in line-per-line imaging) that help to have less variation in PSF across the image [20]. Therefore, a part of deconvolution studies solves the problem with the assumption of having a spatially-invariant PSF [169, 197, 170]. Herein, the experiments are based on plane-wave imaging and we have also considered a fixed PSF in our formulation.

7.2.3 Basics of ADMM

As we use ADMM for solving the proposed optimization problem, a short overview of the method is provided in this section. More details of the ADMM algorithm can be found in [190].

Let us assume that our goal is to solve the following constrained optimization problem:

$$(\hat{\mathbf{u}}, \hat{\mathbf{v}}) = \underset{(\mathbf{u},\mathbf{v})}{\operatorname{argmin}} \{f(\mathbf{u}) + g(\mathbf{v})\} \;\; s.t. \;\; A\mathbf{u} + B\mathbf{v} = \mathbf{c}, \tag{7.6}$$

where $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, and $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$ are closed convex functions. A and B are known matrices, and $\mathbf{c}$ is a given constant vector. To solve the corresponding unconstrained problem, the augmented Lagrangian function can be written as:

$$\mathcal{L}(\mathbf{u}, \mathbf{v}, \lambda) = f(\mathbf{u}) + g(\mathbf{v}) - \lambda^T(A\mathbf{u} + B\mathbf{v} - \mathbf{c}) + \frac{\beta}{2} \parallel A\mathbf{u} + B\mathbf{v} - \mathbf{c} \parallel_2^2 . \tag{7.7}$$

The penalty term with parameter $\beta > 0$ is added to enforce the constraint, and $\lambda \in \mathbb{R}^n$ is the Lagrange multiplier.

Eq. (7.7) can be written in an equivalent but more compact form as follows [191]:

$$\mathcal{L}(\mathbf{u}, \mathbf{v}, \lambda) = f(\mathbf{u}) + g(\mathbf{v}) + \frac{\beta}{2} \parallel A\mathbf{u} + B\mathbf{v} - \mathbf{c} + \frac{\lambda}{\beta} \parallel_2^2 . \tag{7.8}$$

The standard split-variable ADMM algorithm finds the solution of Eq. (7.8) through an iterative process as following [190]:

$$\begin{cases} \mathbf{u}^{i+1} = \text{argmin}_{\mathbf{u}}\, \mathcal{L}(\mathbf{u}, \mathbf{v}^i, \lambda^i) \\ \mathbf{v}^{i+1} = \text{argmin}_{\mathbf{v}}\, \mathcal{L}(\mathbf{u}^{i+1}, \mathbf{v}, \lambda^i) \\ \lambda^{i+1} = \lambda^i + \beta(A\mathbf{u}^{i+1} + B\mathbf{v}^{i+1} - \mathbf{c}) \end{cases} . \tag{7.9}$$

As used in the next section, split-variable ADMM minimizes each term of the cost function separately. This property is beneficial in practice when a single optimization approach is not appropriate for both f and g functions. While the alternating minimization of f and g is much easier, it has been proven that ADMM iterations converge in convex optimization problems [190].

7.3 Proposed joint beamforming-deconvolution algorithm

The main idea of the current work is to find the desired TRF by solving a joint inverse problem of beamforming and deconvolution. Using the same variables introduced in Section 7.2, the proposed optimization problem is as follows:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\text{argmin}} \frac{\gamma_D}{2} \parallel \mathbf{y}_{DAS} - H\mathbf{x} \parallel_2^2 + \frac{\gamma_B}{2} \parallel \mathbf{y}_{ch} - \Phi\mathbf{x} \parallel_2^2 + \mu \parallel \mathbf{x} \parallel_1 . \tag{7.10}$$

In addition to least-square terms for beamforming and deconvolution, the ℓ_1-norm regularization term is also considered to enforce the sparsity of the solution, which is a common choice in ultrasound imaging [131, 178, 132], but other regularization terms such as wavelet frames [125], ℓ_2-norm [131], ℓ_p-norms [155] can also be used with the proposed framework. γ_D, γ_B, and μ are constant hyperparameters controlling the contribution of the deconvolution, beamforming, and sparse regularization terms, respectively. It is obvious that the objective function in Eq. (7.10) is convex. The L_1 term, however, makes it nondifferentiable without a closed-form solution. Therefore, we split the independent variable $\mathbf{x}$ into three equivalent variables $\mathbf{u}$, $\mathbf{z}$, and $\mathbf{w}$ and consider the equality constraints. Hence, the new, but equivalent, form of Eq. (7.10) is as following:

$$(\hat{\mathbf{u}}, \hat{\mathbf{w}}, \hat{\mathbf{z}}) = \underset{(\mathbf{u},\mathbf{w},\mathbf{z})}{\text{argmin}} \frac{\gamma_D}{2} \parallel \mathbf{y}_{DAS} - H\mathbf{u} \parallel_2^2 + \frac{\gamma_B}{2} \parallel \mathbf{y}_{ch} - \Phi\mathbf{z} \parallel_2^2 + \mu \parallel \mathbf{w} \parallel_1 \; s.t. \begin{cases} \mathbf{u} = \mathbf{z} \\ \mathbf{u} = \mathbf{w} \end{cases} . \tag{7.11}$$

By looking closely at Eq. (7.11), it can be considered as a specific form of the general formulation presented in Eq. (7.6) with the following correspondences:

$$\begin{cases} f(\mathbf{u}) = \frac{\gamma_D}{2} \parallel \mathbf{y}_{DAS} - H\mathbf{u} \parallel_2^2 \\ g(\mathbf{v}) = \frac{\gamma_B}{2} \parallel \mathbf{y}_{ch} - \Phi\mathbf{z} \parallel_2^2 + \mu \parallel \mathbf{w} \parallel_1 \\ \mathbf{v} = \begin{bmatrix} \mathbf{w} \\ \mathbf{z} \end{bmatrix}, \lambda = \begin{bmatrix} \lambda_{\mathbf{1}} \\ \lambda_{\mathbf{2}} \end{bmatrix} \\ A = \begin{bmatrix} I \\ I \end{bmatrix}, B = \begin{bmatrix} -I & 0 \\ 0 & -I \end{bmatrix}, \mathbf{c} = 0 \end{cases} , \tag{7.12}$$

where I refers to an identity square matrix of size n, therefore, the augmented Lagrangian function of Eq. (7.11) is exactly the same as what was presented previously in Eq. (7.8), and its solution can be found using the split-variable ADMM approach.

Algorithm 7.1 describes the proposed solution of our cost function in the ADMM framework. Different terms of the Eq. (7.10) are minimized separately in each iteration. The algorithm is

Algorithm 7.1 ADMM Algorithm for solving Eq. (7.10)

1: **Input:** H, Φ, $\mathbf{y}$
2: **Set:** $\gamma_D > 0$, $\gamma_B > 0$, $\mu > 0$, $\beta > 0$, $\mathbf{u}^0$, $\mathbf{v}^0$, λ^0, ϵ
3: **While** stopping criterion $> \epsilon$ **do**
4: $\mathbf{u}^{i+1} = \text{argmin}_{\mathbf{u}} \frac{\gamma_D}{2} \parallel \mathbf{y}_{DAS} - H\mathbf{u} \parallel_2^2 + \frac{\beta}{2} \parallel A\mathbf{u} + B\mathbf{v}^i + \frac{\lambda^i}{\beta} \parallel_2^2$
5: $\mathbf{z}^{i+1} = \text{argmin}_{\mathbf{z}} \frac{\gamma_B}{2} \parallel \mathbf{y}_{ch} - \Phi\mathbf{z} \parallel_2^2 + \frac{\beta}{2} \parallel \mathbf{u}^{i+1} - \mathbf{z} + \frac{\lambda_2^i}{\beta} \parallel_2^2$
6: $\mathbf{w}^{i+1} = \text{argmin}_{\mathbf{w}} \mu \parallel \mathbf{w} \parallel_1 + \frac{\beta}{2} \parallel \mathbf{u}^{i+1} - \mathbf{w} + \frac{\lambda_1^i}{\beta} \parallel_2^2$
7: $\lambda^{i+1} = \lambda^i + \beta(A\mathbf{u}^{i+1} + B\mathbf{v}^{i+1})$
8: **End**

initialized by setting the hyperparameters γ_D, γ_B, μ, and β. A small constant value ϵ is chosen as the threshold for the stopping criterion. The iterative optimization procedure is started with arbitrary initial values for the Lagrange multiplier (λ) and the new variables (i.e., $\mathbf{u}$, $\mathbf{w}$, and $\mathbf{z}$). The proposed ADMM solution can be summarized in three steps as follows.

7.3.1 Deconvolution update

In this step, the solution of the deconvolution term ($\mathbf{u}$) is found by minimizing the corresponding subproblem written in line 4 of Algorithm 7.1. Since this cost function is convex and differentiable, the solution can be easily found by taking the gradient and setting it to zero. By doing so, the following closed-form solution is derived:

$$\mathbf{u}^{i+1} = (\gamma_D H^T H + 2\beta J)^{-1}(\gamma_D H^T \mathbf{y}_{DAS} + \beta\mathbf{w}^i + \beta\mathbf{z}^i - \lambda_1^i - \lambda_2^i), \tag{7.13}$$

where J is a matrix of ones with the same size as $H^T H$. Eq. (7.13) can also be solved in the Fourier domain. This implementation has been successfully used in [170, 169] to reduce the computational complexity of the solution in each iteration.

7.3.2 Beamforming update

The second step corresponds to the minimization of beamforming term written in line 5 of Algorithm 7.1. Same as step 1, if we set the gradient of the cost function to zero, we arrive at the following analytical solution:

$$\mathbf{z}^{i+1} = (\gamma_B \Phi^T \Phi + \beta J)^{-1}(\gamma_B \Phi^T \mathbf{y}_{ch} + \beta\mathbf{u}^{i+1} + \lambda_2^i), \tag{7.14}$$

where $\Phi^T\Phi$ is usually a square matrix of size several hundreds of thousands, and not even diagonal, nor one that can be diagonalizable through Fourier transform. Therefore, Eq. (7.14) is intractable in practice as involves large matrix inversion. That is why a numerical method is adopted to tackle this problem and find the solution of this step. Herein, the optimal solution is found using limited-memory Broyden–Fletcher–Goldfarb–Shanno (BFGS) solver[1], which is a quasi-Newton approach. Limited-memory BFGS achieves quadratic convergence for many problems [192] and is highly efficient for unconstrained optimization of differentiable real-valued high-dimensional functions. It has to be emphasized that the modular property of ADMM provides the possibility of solving the beamforming term using limited-memory BFGS in which the Hessian matrix is approximated. This property is worth noting in our problem wherein the large size of matrix Φ makes the calculation of Hessian intractable.

[1]MATLAB implementation is publicly available in this link: `http://www.cs.ubc.ca/~schmidtm/Software/minFunc.html`

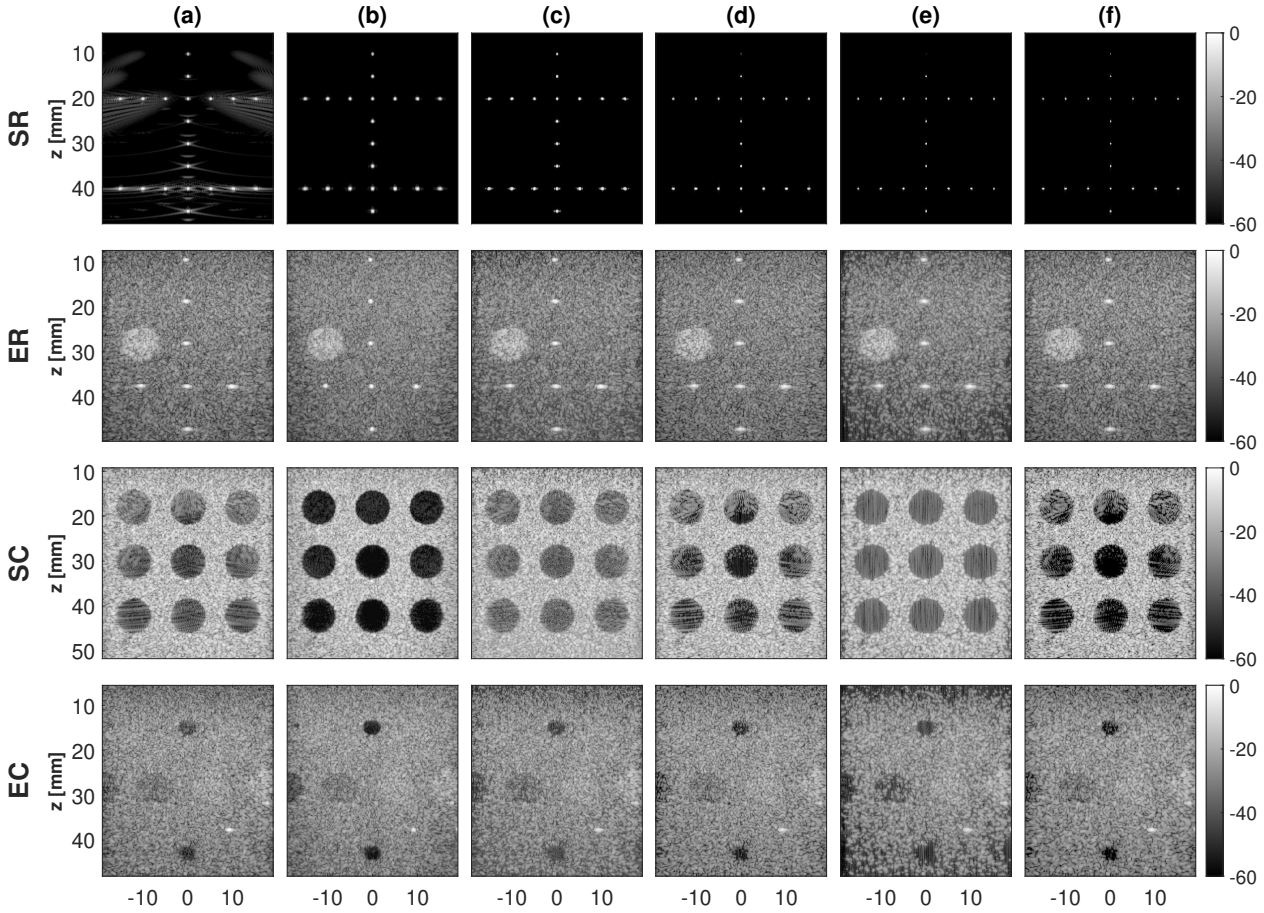

Figure 7.2: Simulation and experimental images reconstructed through different methods. Rows indicate datasets while columns correspond to different approaches. (a) DAS. (b) CPWC. (c) The inverse problem of beamforming. (d) The inverse problem of deconvolution. (e) Sequential approach. (f) The proposed joint formulation. CPWC is obtained from 75 steered insonifications. All other results are from a single 0^o insonification.

7.3.3 Sparsity and Lagrange multiplier updates

The final step of our method entails the optimization of sparsity constraints and updating the Lagrangian multiplier. The minimization problem presented in line 6 of Algorithm 7.1 is commonly called as the proximal mapping of the L_1 norm as follows [191]:

$$prox_{\mu\|.\|_1/\beta}(\mathbf{u}^{i+1} + \frac{\lambda_1^i}{\beta}) = \text{argmin}_{\mathbf{w}}\, \mu \parallel \mathbf{w} \parallel_1 + \frac{\beta}{2} \parallel \mathbf{u}^{i+1} - \mathbf{w} + \frac{\lambda_1^i}{\beta} \parallel_2^2 . \tag{7.15}$$

Eq. (7.15) is the minimization of a strictly convex function, and its unique minimizer can be presented in terms of shrinkage function [191], which acts as a soft-thresholding operator:

$$\mathbf{w}^{i+1} = soft_{\mu/\beta}(\mathbf{u}^{i+1} + \frac{\lambda_1^i}{\beta}) = max\{|\mathbf{u}^{i+1} + \frac{\lambda_1^i}{\beta}| - \frac{\mu}{\beta}, 0\} sign(\mathbf{u}^{i+1} + \frac{\lambda_1^i}{\beta}). \tag{7.16}$$

Finally, the Lagrangian multiplier needs to be updated using the equation in line 7 of Algorithm 7.1.

In each iteration of Algorithm 7.1, the original objective function (i.e., Eq. (7.10)) is calculated, and its relative error for two consecutive iterations is used as the stopping criterion. The theoretical

convergence of the split variable ADMM algorithm to a global minimum in any convex optimization problem has been shown [215].

7.4 Experiments

Herein, our dataset is exactly the same as what we have explained in Section 5.2.1.

7.4.1 Evaluation metrics

The images reconstructed using the proposed method are evaluated in terms of two main specialized ultrasound assessment indexes, including resolution and contrast.

Recently, it has been shown that quantitative indexes such as CNR and FWHM are not reliable when the dynamic range of the final image has been transformed [216, 217]. For example, FWHM improves by taking the square of an image while there is no new information content [217]. Therefore, it is recommended to use histogram matching (HM) prior to the visualization and making the quantitative measurements, which lead to fairer comparisons between different methods and would be an acceptable alternative [216, 217]. In the current chapter, HM is applied to all the presented results in the next section except for the SR experiment for which HM brings additional artifacts due to its binary content. Since a homogeneous speckle region of interest (ROI) has a well-behaved log-Rayleigh distribution for B-mode images, ROI-based HM (details can be found in [216]) is applied to the result of each method, and DAS output is considered as the reference image.

7.5 Results

In this section, the results of the proposed joint formulation are compared with the sequential approach that entails beamforming followed by deconvolution. To better understand the effect of each term in the proposed joint cost function, the images reconstructed by only solving the inverse problem of beamforming (i.e., $\gamma_D = 0$ and $\gamma_B = 1$ in Eq. (7.10)) or deconvolution (i.e., $\gamma_D = 1$ and $\gamma_B = 0$in Eq. (7.10)) are also presented in Section 7.5.1. Furthermore, to demonstrate the superiority of the proposed approach in comparison with the previous beamforming methods, the results of Eigenspace-based MV (EMV) [2], PCF [3], Fourier domain technique based on Stolt's

Table 7.1: Quantitative results in terms of resolution and contrast indexes for simulation and real phantom experiments. Bold numbers refer to the best performances among different methods excluding the reference CPWC.

dataset	SR		ER		SC		EC	
index	$FWHM_A$	$FWHM_L$	$FWHM_A$	$FWHM_L$	CNR	gCNR	CNR	gCNR
DAS	0.4	0.47	0.57	0.96	10.25	0.89	7.75	0.8
CPWC	0.4	0.4	0.39	0.39	17.57	0.99	12.9	0.94
Beamforming	0.38	0.4	0.56	0.94	10.12	0.9	7.65	0.78
Deconvolution	0.29	0.34	0.37	0.55	9.92	0.88	8.15	0.79
Sequential	0.26	0.27	0.36	0.5	10.04	0.88	8.1	0.8
Joint	**0.22**	0.26	0.34	0.46	**11.54**	**0.95**	**9.15**	**0.85**
EMV [2]	0.4	**0.1**	0.51	0.46	9.34	0.85	8.25	0.84
PCF [3]	0.3	0.38	0.46	0.41	7.86	0.79	6	0.7
Stolt [5]	0.42	1.12	0.3	**0.28**	2.3	0.55	7.2	0.79
UFSB [4]	0.4	0.85	0.44	0.42	7.85	0.79	7	0.78
PMNB [6]	0.34	0.28	**0.18**	0.42	8.07	0.82	4.55	0.69

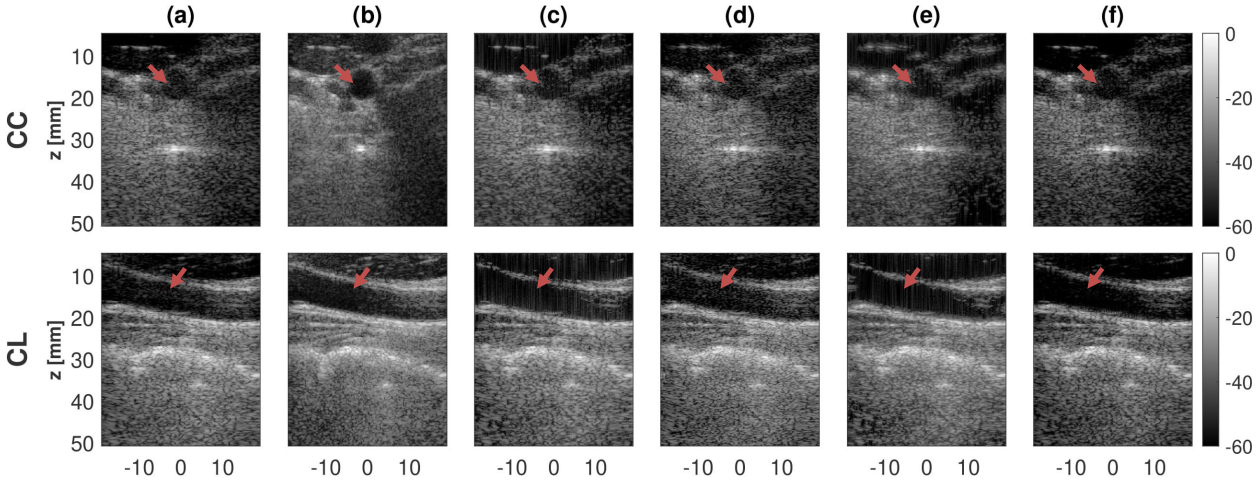

Figure 7.3: Results on *in vivo* data. Rows indicate datasets while columns correspond to different approaches. (a) DAS. (b) CPWC. (c) The inverse problem of beamforming. (d) The inverse problem of deconvolution. (e) Sequential approach. (f) The proposed joint formulation. CPWC is obtained from 75 steered insonifications. All other results are from a single 0^o insonification. Red arrows indicate the carotid artery, which ideally should be dark with no clutter artifact.

migration [5], and ultrasound Fourier slice beamforming (UFSB) [4] are included in Section 7.5.2. In order to have a fair comparison with the previous deconvolution methods, we need to keep the same regularization term. In other words, our results should be compared to a ℓ_1 norm regularized deconvolution. As for the spatially-invariant PSF, it basically comes to canceling the beamforming data fidelity term in our algorithm, which is mentioned above. We have also compared our results with the previous method (called PMNB) [6] in which the spatial variance of the PSF is taken into account. The sensitivity analysis of the proposed method to initial points and parameter selection is presented in Section 7.5.3. Finally, in Section 7.5.4, the convergence graphs of the ADMM algorithm in different problems are presented. The convergence graphs of the ADMM and the Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) are compared. Furthermore, the reconstruction times of all methods are reported and discussed for a single experiment.

Hereafter, we consider the Hanning apodization window with $f\# = 0.5$ for DAS and other methods on top of DAS (except for *in vivo* datasets for which the Tukey (tapered cosine) window with constant parameter of 0.25 and f-number equals to 1.75 is considered). $\epsilon = 10^{-3}$ is selected as the threshold for stopping criterion. The iterative algorithms are started with initial values equal to zero. Since the proposed joint formulation and also deconvolution approach require PSF, the method proposed in [218] has been adopted to estimate the unknown PSF from the RF data. The quantitative indexes are calculated independently for different point targets or cyst regions, and the average values are reported. The hyperparameters of each method are set independently to achieve the best results.

7.5.1 The proposed joint formulation

7.5.1.1 Simulation and experimental data

The images reconstructed from the single 0^o plane-wave are presented in Fig. 7.2. As for the reference quality, the results of Coherent Plane-Wave Compounding (CPWC) on 75 insonifications are illustrated in the second column of Fig. 7.2. The proposed joint formulation can successfully

reconstruct high-quality images in terms of both resolution and contrast. Fig. 7.2 shows that the sequential approach has a poor performance because a part of image content get lost in the two consecutive soft-thresholding steps. As for the result of beamforming and deconvolution alone, the artificial improvement in contrast has been revoked by HM. And Fig. 7.2 depicts that only the results of CPWC and the proposed method are robust to HM.

As quantitative results of Table 7.1 confirm, the proposed method gives high axial and lateral resolutions for both simulation and experimental data similar to or even better than CPWC. The highest contrast corresponds to CPWC results since it averages over 75 angles and perfectly suppresses the side-lobe artifacts. For a single 0^o insonification, the proposed approach improves the contrast as compared to other methods. This point can also be seen in the quantitative comparison reported in Table 7.1.

7.5.1.2 *In vivo* data

The proposed method is also evaluated on real data collected from the carotid artery. The visual comparison of different approaches is illustrated in Fig. 7.3. As can be seen in Fig. 7.3, the proposed method is able to suppress the clutter artifacts caused by diffuse reverberation from shallow layers and create a dark image of the artery in both cross-sectional and longitudinal views (pointed out by red arrows).

7.5.2 Comparison with other methods

In this section, the results of previous beamforming and deconvolution methods are shown. We have included an example from each group of classical beamforming methods reviewed in Section 7.1 plus a deconvolution method based on nonstationary PSF [6]. The results of DAS and inverse problem formulation are presented along with the proposed approach in Section 7.5.1.1. The visual comparison of results, presented in Fig. 7.4, indicates that EMV only gives a better lateral resolution for SR dataset as compared to the proposed method. This point can also be understood from the quantitative comparison reported in Table 7.1. PCF works better than DAS in terms of axial resolution, while its results are still worse than the proposed method. In terms of lateral resolution, however, the results are discussible. While EMV substantially improves the lateral resolution of simulation data, the same improvement was not achieved in the real phantom experiment. The reason behind this difference is the number of eigenvectors of the covariance matrix used for creating the signal subspace. More specifically, EMV only needs the principal eigenvectors of the SR data to successfully reconstruct the image because it only includes some point targets. However, if we only consider some of the eigenvectors in the ER case, a part of the speckle texture would be lost. This variation in improvement is also seen in terms of contrast. That is because of the additional noise of experimental data, which reduces the quality of covariance matrix estimation.

In short, the comparison with previous approaches reveals that the proposed method gives the most consistent improvement in image quality for all datasets. Although other methods may provide better qualities in specific cases, their performance drops in other experiments. This point can also be seen for Stolt's migration [5] as well as PMNB [6] approaches, which give a high resolution in ER experiment but has a low performance for other datasets.

7.5.3 Sensitivity analysis

As mentioned in Algorithm 7.1, the proposed method is initialized with the arbitrary values for the Lagrange multiplier (λ) and new variables (i.e., $\mathbf{u}$, $\mathbf{w}$, and $\mathbf{z}$). Furthermore, hyperparameters γ_D,

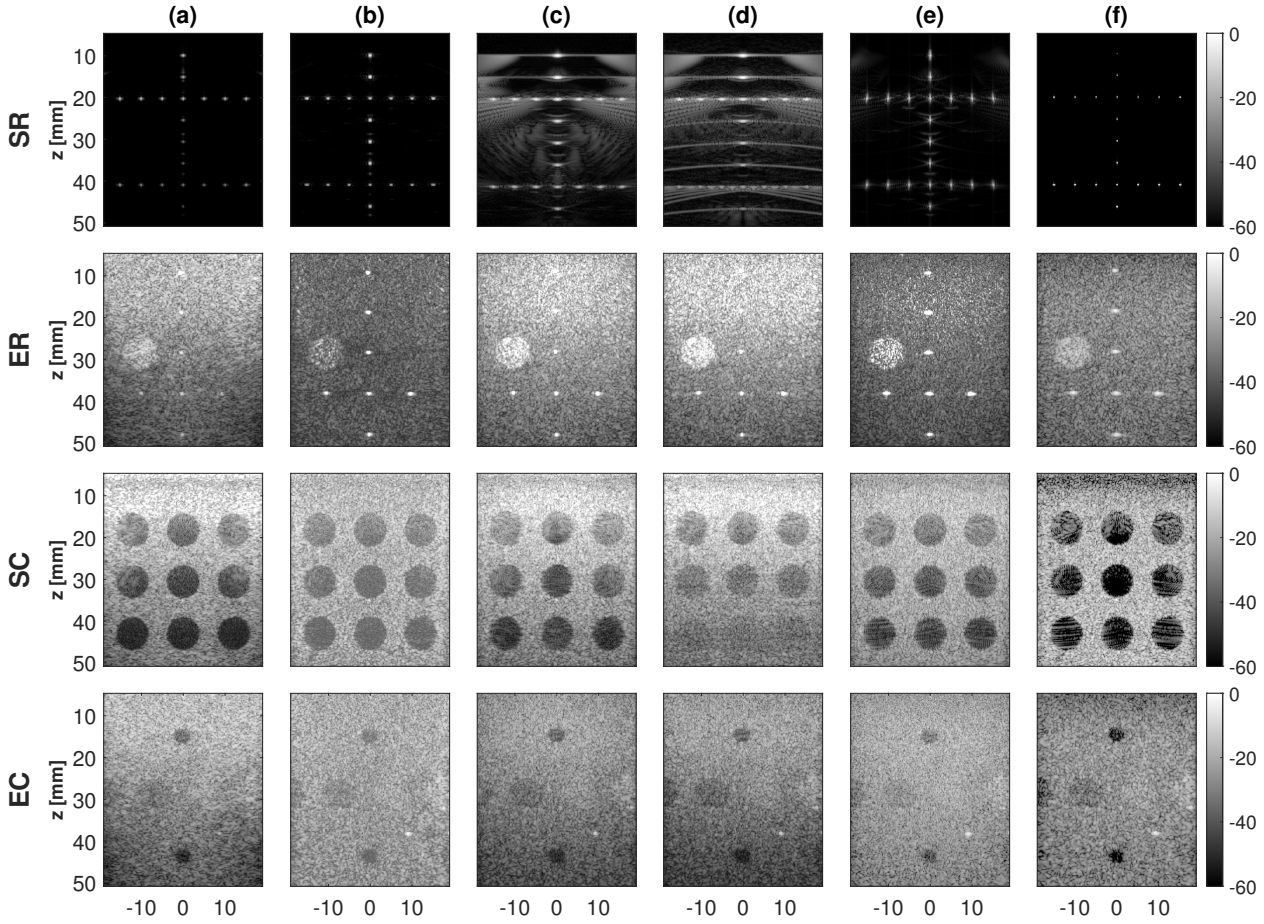

Figure 7.4: Simulation and experimental images reconstructed through previous beamforming and deconvolution approaches. Rows indicate datasets while columns correspond to different approaches. (a) EMV [2]. (b) PCF [3]. (c) UFSB [4]. (d) Stolt's migration [5]. (e) PMNB [6]. (f) The proposed joint formulation. All the results are from a single 0^o insonification. The sharp point targets and dark cyst regions are desired.

γ_B, μ, and β, which respectively specify the weights of deconvolution, beamforming, sparsity, and penalty terms, need to be set.

Generally, the initial values of λ and new variables do not make any difference to the final solution because the proposed objective function is convex, and there is no local minimum. Notwithstanding, the starting points can affect the processing time and the number of iterations for convergence. For all the results presented here, zero is used as the initial value, and the proposed algorithm always converges in less than 40 iterations. The code is implemented in Windows 10 using the MATLAB R2021a programming platform.

The parameter β determines the weight of penalty term in the augmented Lagrangian function. It is necessary to set a large value to β because the split-variable ADMM algorithm must perfectly accomplish the equality constraints of new variables (Eq. (7.11)). Otherwise, a small β allows the algorithm to converge toward different values for $\mathbf{u}$, $\mathbf{w}$, and $\mathbf{z}$ which is not justifiable.

Different hyperparameters' values may completely change the final solution. To demonstrate the effect of each one in the proposed algorithm, the SC image is reconstructed multiple times with different sets of hyperparameters. First, Eq. (7.10) is solved by only considering the beamforming

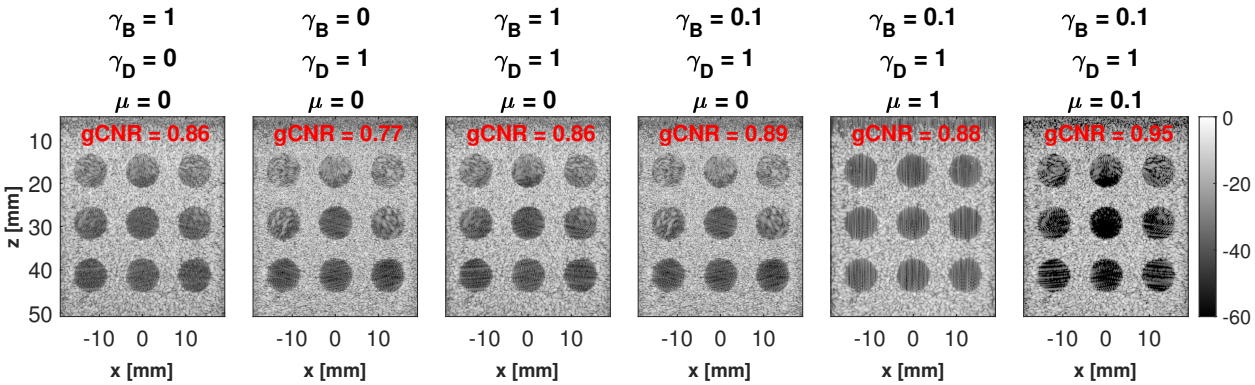

Figure 7.5: The visualization of hyperparameters' effect on the reconstructed Simulation Contrast (SC) image. The proposed method is run multiple times with different combinations of parameters in order to demonstrate the role of each one.

term (i.e., $\gamma_D = 0$ and $\mu = 0$ in Eq. (7.10). As explained in Section 7.2.1, each row of matrix Φ contains the contribution of all pixels that simultaneously arrive to a piezoelectric element. Therefore, as can be seen in the first column of Fig. 7.5, the result is even worse than DAS, mainly due to the off-axis echoes. The second column of Fig. 7.5 only depicts the effect of the deconvolution term (i.e., $\gamma_B = 0$ and $\mu = 0$ in Eq. (7.10) in which the speckle texture is sharper due to the mitigation of PSF. However, the cyst regions still suffer from clutter artifacts. The beamforming and deconvolution terms are combined with the same weights in the third column of Fig. 7.5. The resulting image confirms that the beamforming term was dominant since the solution is similar to the first column. Therefore, the beamforming weight (i.e., γ_B) is reduced to ensure a better balance between both terms. As the fourth column of Fig. 7.5 shows, the resulting image is of better contrast, while the speckle texture is also sharp. In order to reduce clutter artifacts in the cyst regions, the sparsity regularization term is also considered in the fifth column of Fig. 7.5. However, a large μ wipes out the speckle texture and the resulting image becomes too dark. This artificial improvement in contrast index is always revoked by HM. Finally, once the regularization term is also reduced, the resulting image is of high quality and takes the benefits of all three terms simultaneously (the sixth column of Fig. 7.5).

7.5.4 Convergence analysis

Generally, an important point regarding the optimization algorithms used for solving the inverse problem of imaging is the convergence. The theoretical analysis of the convergence for the ADMM algorithm and its computational cost is out of the scope of the current work. Therefore, a plot of stopping criterion versus iteration for different problems solved by ADMM is provided in Fig. 7.6-(a). As mentioned in Section 7.3.3, in each iteration (i) of the proposed algorithm, the objective function (i.e., Eq. (7.10)) is calculated and its relative error for two consecutive iterations is used as the stopping criterion. The proposed joint formulation clearly starts with a higher relative error since it consists of both beamforming as well as deconvolution terms while ADMM converges roughly in 40 iterations for all three problems. In addition to ADMM, Eq. (7.10) is also solved by the Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) [219]. Fig. 7.6-(b) illustrates the convergence of ADMM and FISTA for the reconstruction of the SC dataset. Although FISTA needs more iterations for convergence, its reconstruction time is better than ADMM because the ADMM iterations include the time-consuming BFGS solver.

In order to provide a clear understanding of the speed of different methods, the time taken by each method for the reconstruction of the SC dataset is presented in Fig. 7.7. In short, the slowest algorithm is EMV [2] because of the time-consuming covariance matrix estimation and the independent reconstruction of each pixel of the final image. The proposed joint formulation based on ADMM is the second slow algorithm since it involves blind PSF estimation plus the time-consuming BFGS solver. Fourier domain technique based on Stolt's migration [5] is the fastest algorithm as it takes 100 milliseconds.

7.6 Discussion

The sequential approach and joint formulation comparison confirm that solving each beamforming and deconvolution problem separately does not lead to the same quality. This might be due to a loss of information in the first step of the sequential approach. The same observation has been reported in [169].

Another important advantage of the proposed formulation is the substantial improvement in axial resolution. As reported in Section 7.5.2, other beamforming approaches either do not change or have a lower effect on the axial resolution as compared to our results. This point is crucial because the resolution in the axial direction is usually increased by transmitting pulses with a higher center frequency.

The achieved improvement in image quality can reduce the need to transmit several plane-waves with different angles. Furthermore, the proposed framework provides the possibility of considering the result of any other beamforming approach (not only DAS) in the deconvolution term. In other words, the proposed objective function, in Eq. (7.10), is a linear combination of beamforming and deconvolution terms. Therefore, extra terms based on the result of other beamforming approaches can also be added.

Using ADMM makes the optimization step of the proposed method easy to implement and reduces computational costs. As the proposed algorithm is iterative and the variables are updated serially, parallel implementation is impossible. Therefore, although our method is much faster than a computationally expensive algorithm such as MV, real-time image reconstruction using the proposed method might not be possible. This would be the subject of our future research.

Although the proposed idea can be applied to any imaging technique (i.e., focused, plane-wave,

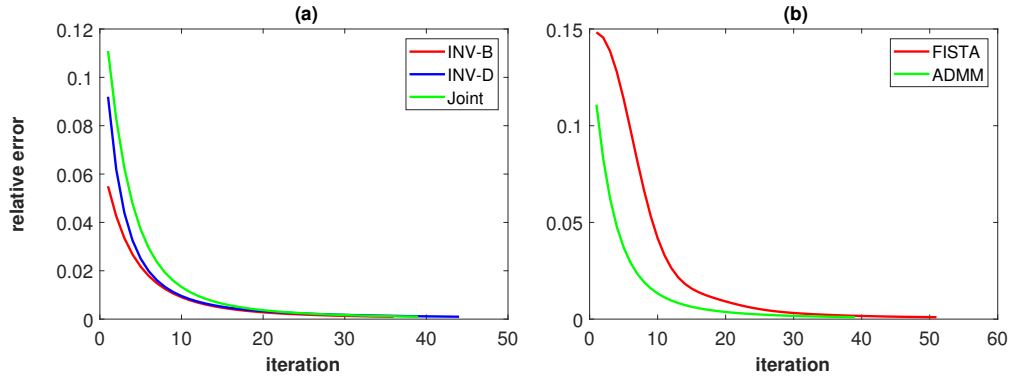

Figure 7.6: (a) The convergence graphs of ADMM for solving the proposed joint formulation, the inverse problem of beamforming (INV-B), and the inverse problem of deconvolution (INV-D) in order to reconstruct the Simulation Contrast (SC) image. (b) The convergence graphs of ADMM and FISTA used for solving the proposed joint formulation in order to reconstruct the SC image.

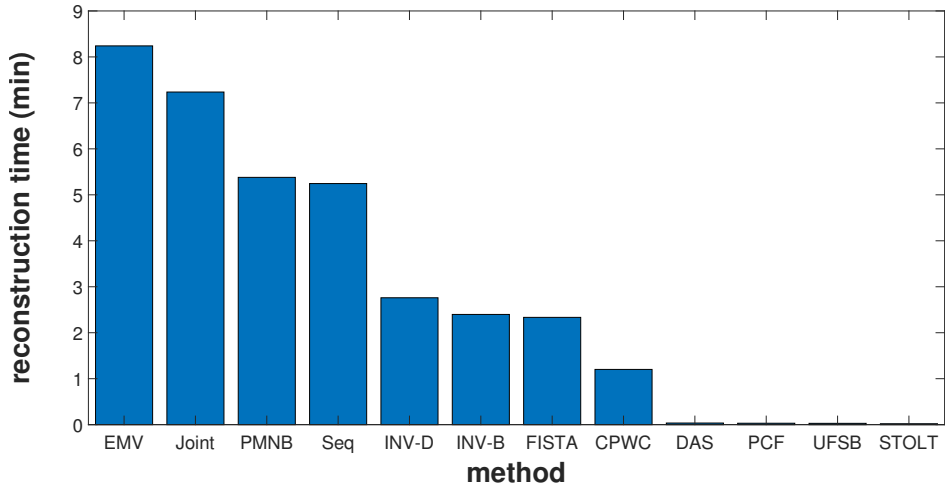

Figure 7.7: The reconstruction times of different methods for the Simulation Contrast (SC) dataset.

and synthetic aperture imaging) or probe type (i.e., linear, convex, and phased array), only the benchmark PICMUS dataset is used here because it is publicly available and the comparison with previous approaches is easier. It also helps the readers to reimplement the algorithm and verify the results easily. The proposed method can also be applied on top of the CPWC. However, to limit the sources of improvement, CPWC results are not used.

It has been previously shown that solving the inverse problem of beamforming gives images with high resolution and contrast [131, 132, 178]. This improvement, however, comes at the expense of speckle information loss. Furthermore, the inverse problem of deconvolution cannot solely improve the image quality. Combining both terms in our proposed objective function helps to achieve a high resolution and contrast while the speckle texture is also preserved. This point can be seen in ER data. The proposed method's ability to preserve the speckle texture is of crucial importance in image computing applications such as speckle tracking and tissue classification. In addition, experienced radiologists often rely on the speckle pattern for diagnosis.

The performance of the proposed method directly depends on the quality of the estimated PSF of the imaging system. Herein, we utilize a common approach used in previous literature [197, 169, 170]. However, any method for PSF estimation can also be used to improve the results. We plan to extend our idea to consider a nonstationary PSF in the model and also take advantage of the harmonic components in RF data for image reconstruction.

7.7 Conclusions

Beamforming and deconvolution have only been used separately in a sequential approach. Herein, we proposed a novel formulation for combining both methods. A regularized inverse problem including two linear models for beamforming and deconvolution plus additional sparsity constraint is solved using the ADMM algorithm. The proposed image reconstruction approach is a joint optimization problem that uses DAS results as an observation. The results show that the proposed iterative method gives ultrasound images with a high resolution and contrast.

Chapter 8

Conclusions and Future Work

8.1 Conclusions

A central question in this thesis was how to reconstruct a high-quality image of tissue echogenicity from RF channel data recorded by the transducer elements. Significant parts of the required information for perfect reconstruction of tissue echogenicity are corrupted due to intrinsic limitations of the medical ultrasound imaging. More specifically, there are several image formation steps before the collection of backscattered data which govern the degree of data corruption. Herein, we focused our investigations on extracting the best possible image quality from the raw channel data. The proposed methods span a wide range of techniques to tackle this problem, from deep learning image reconstruction to inverse problem formulations. Our principal assumption was that a part of the underlying useful information of RF channel data gets lost in common beamforming methods such as DAS, and our goal was to improve the beamforming step.

The proposed methods based on deep learning rely on the ability of neural networks to learn high-dimensional mapping functions from the training data. More specifically, the proposed multi-focus imaging model has shown that the available information from a single-focus transmission can be transformed into a multi-focus image through a GAN model. Besides designing an appropriate architecture, the GAN model benefits from adversarial training, which results in perceptually compelling images. The experimental results on simulation, phantom, and *ex vivo* data confirmed that the proposed approach outperforms other methods in terms of image quality and temporal resolution.

The proposed deep beamformer based on MobileNetV2 is designed to learn a complicated multi-step mathematical function. Therefore, all the necessary preprocessing steps are applied to provide enough information in the network's input. The training results showed that the task is adequately possible, and the network can mimic MVB. Deep approximation makes the beamforming six times faster, while pruning the trained model can further reduce the time by a factor of 2. In terms of memory requirements, pruning the architecture reduced the model size by a factor of 7.5, which is highly important, especially for point of care ultrasound (POCUS) machines.

The reconstruction method proposed in Chapter 4 answers the question of to what extent the PSF of the plane-wave imaging can be improved. More specifically, an equilateral sharp Gaussian PSF is considered for the reconstruction of training targets instead of the conventional ultrasound PSF, which has a strong side- and grating-lobes. The width of Gaussian PSF was specified based on the training performance. Although the formation of imaging PSF results from many complicated factors, the test results have shown that it can be modeled using the proposed idea. Ultrasound simulations were performed to provide the training data since ground-truth in real ultrasound

imaging is often unknown. This approach might be helpful for localization scatterers beyond the resolution of plane-wave imaging.

The proposed adaptive beamformer based on ICA is a novel interpretation of the beamforming process. In other words, the beamforming is translated into a blind source separation problem. We considered the desired tissue reflectivity function of the medium as a single source observed by the transducer elements. The results support that ICA can efficiently extract the relevant apodization window based on the medium content. This approach is also among fast adaptive beamformers, which estimate the apodization weights in less than one second.

Our proposed novelties on the inverse problem formulation of ultrasound beamforming mainly associate with the introduction of recent breakthroughs in high-dimensional convex optimization. More specifically, our linear image formation model is the same as what has already been used in the literature. However, applying denoising algorithms in the form of PnP and RED methods helped us reconstruct images with the highest contrast while preserving the speckle information. Furthermore, combining deconvolution and beamforming in a single objective function results in a high-quality image in terms of both resolution and contrast.

In short, our main objective was to get closer to an optimal balance among competing factors such as resolution, contrast, framerate, and computational complexity in medical ultrasound imaging. Each chapter contains a novel idea that partially solves a specific problem in the image reconstruction pipeline.

8.2 Limitations and Future Work

Our proposed methods can be considered as finding local minima in non-convex optimization problems. In other words, if we look at the problem of interest from a broad perspective, it is easy to comprehend that the advantages of each idea are not always valid and subject to some constraints. Notwithstanding, the limitations of each method pose an avenue for future work and guide us toward a global solution for the task. Herein, restrictions of each idea are mentioned, and the feasible solutions are described as potential avenues for future work.

In Chapter 2, different networks are trained on each depth of B-mode ultrasound images. Therefore, besides training the deep model on RF data, training a single model with separate modules for the reconstruction of the whole image may, in fact, yield superior results because learning multiple related goals often improves the performance of deep networks in the multi-task learning (MTL) framework. Moreover, the reconstruction of multi-focus RF images is useful in applications such as Doppler imaging based on RF data. Transfer learning could have also been optimized by selecting only a part of the layers for fine-tuning. A recent study has shown that layer selection is crucial in deep learning in ultrasound [220].

Our deep beamformer in Chapter 3 is designed to reconstruct each pixel of the final image separately, similar to MVB. This point, however, is computationally expensive during the inference phase, and an interesting room for improvement is to stack the input data corresponding to all pixels in a single tensor and reconstruct the whole image at once. Furthermore, the resulting image quality can also be improved by considering the output of recent extensions of MVB in which not only the contrast index has been improved without sacrificing the speckle statistics, but also the parameters of MVB are automatically selected in an optimal manner [221]. Another important question about deep beamformers is their sensitivity to translation in the axial and lateral directions. More specifically, it has been shown that CNNs are not shift equivariant as commonly perceived, meaning that a small translation in the input may cause a noticeable output alteration [222]. Therefore, it is necessary to find a solution if shift variance is an issue in the

beamforming problem as well.

The method proposed in Chapter 4 is only trained with clean simulation data. Nevertheless, considering attenuation in the tissue and inhomogeneities in the speed of sound can enrich the training dataset and improve the method's performance on real *in vivo* test data. Moreover, combining the proposed method with the phase aberration correction techniques may further improve the robustness of the proposed method. As an example, it has been shown that the aberration profile of the tissue can be estimated through CNNs [223]. Therefore, this method can be used as an initial step to compensate the distortion caused by inhomogeneities in the speed of sound of input channel data and then apply the proposed method.

Although the ICA beamformer has shown promising results on PWI and CPWC, the application of this method can be extended to recover images of line-per-line as well as synthetic aperture imaging. Moreover, the sensitivity of this method to input noise warrants further investigation because the low quality of channel data directly affects the performance of adaptive methods. Another avenue for improvement is to incorporate other blind source separation techniques such as non-negative matrix factorization or common spatial pattern in the estimation of the apodization window. Currently, we only use a single window for the entire image. However, our knowledge about the medium and physics of the problem can be incorporated in the form of a Fuzzy expert system with several rules wherein each rule is based on a distinct apodization window.

Both novelties on inverse problem formulation suffer from a high computational cost due to the iterative nature of the algorithms and the massive size of the weighting matrices. Besides speeding up the algorithms, it is possible to improve the linear imaging model by considering more realistic nonlinear models and incorporating the PSF. The methods proposed in Chapters 6 and 7 include several steps, such as optimization, denoising, PSF estimation, etc., and each opens an avenue for the extension. For example, the modular property of ADMM makes the utilization of any optimization algorithm possible. In addition, considering a non-stationary PSF would provide superior results since it is more faithful to actual imaging conditions. Furthermore, we have only used the principal component of the RF channel data, while the harmonic components can also be employed in the reconstruction process. Finally, we have considered the recorded signals from a single plane-wave transmission. At the same time, it is possible to define the weighting matrix for several transmissions in order to improve the resolution further.

References

[1] S. Goudarzi, A. Asif, and H. Rivaz, "Angular apodization estimation using independent component analysis in coherent plane-wave compounding," in *2020 IEEE International Ultrasonics Symposium (IUS)*, 2020, pp. 1–4.

[2] B.M. Asl and A. Mahloojifar, "Eigenspace-based minimum variance beamforming applied to medical ultrasound imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 57, no. 11, pp. 2381–2390, 2010.

[3] J. Camacho, M. Parrilla, and C. Fritsch, "Phase coherence imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 56, no. 5, pp. 958–974, 2009.

[4] A. Besson, M. Zhang, F. Varray, H. Liebgott, D. Friboulet, Y. Wiaux, J.P. Thiran, R.E. Carrillo, and O. Bernard, "A sparse reconstruction framework for fourier-based plane-wave imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 63, no. 12, pp. 2092–2106, 2016.

[5] M. Albulayli and D. Rakhmatov, "Fourier domain depth migration for plane-wave ultrasound imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 65, no. 8, pp. 1321–1333, 2018.

[6] A. Besson, L. Roquette, D. Perdios, M. Simeoni, M. Arditi, P. Hurley, Y. Wiaux, and J.P. Thiran, "A physical model of nonstationary blur in ultrasound imaging," *IEEE Transactions on Computational Imaging*, vol. 5, no. 3, pp. 381–394, 2019.

[7] T. L. Szabo, *Diagnostic Ultrasound Imaging: Inside Out*, Academic Press, second edition edition, 2014.

[8] B. Kimmelman, "Medical diagnostic ultrasound: a retrospective on its 40th anniversary," *American Institute of Ultrasound in Medicine*, 1988.

[9] R. SC. Cobbold, *Foundations of biomedical ultrasound*, Oxford university press, 2006.

[10] "Siemens company website," [Online]. Available: `https://new.siemens.com/global/en/company/about/history/technology/medical-technology/imaging.html`.

[11] "Siemens company website," [Online]. Available: `https://www.siemens-healthineers.com/en-ca/ultrasound/new-era-ultrasound/acuson-redwood`.

[12] "Siemens company website," [Online]. Available: `https://www.siemens-healthineers.com/en-ca/ultrasound/ultrasound-point-of-care/acuson-freestyle-ultrasound-machine`.

[13] G. S. Kino, *Acoustic waves: devices, imaging, and analog signal processingPrentice-Hall Signal Processing Series*, Englewood Cliffs, Prentice-Hall, 1987.

[14] O. Senouf, S. Vedula, G. Zurakhov, A. Bronstein, M. Zibulevsky, O. Michailovich, D. Adam, and D. Blondheim, "High frame-rate cardiac ultrasound imaging with deep learning," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 126–134.

[15] B. Lokesh and Arun K. Thittai, "Diverging beam transmit through limited aperture: A method to reduce ultrasound system complexity and yet obtain better image quality at higher frame rates," *Ultrasonics*, vol. 91, pp. 150 – 160, 2019.

[16] C.H. Frazier and W.D. O'Brien, "Synthetic aperture techniques with a virtual source element," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 45, no. 1, pp. 196–207, 1998.

[17] M. Tanter and M. Fink, "Ultrafast imaging in biomedical ultrasound," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 61, no. 1, pp. 102–119, 2014.

[18] G. Montaldo, M. Tanter, J. Bercoff, N. Benech, and M. Fink, "Coherent plane-wave compounding for very high frame rate ultrasonography and transient elastography," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 56, no. 3, pp. 489–506, 2009.

[19] B. Denarie, T. A. Tangen, I. K. Ekroll, N. Rolim, H. Torp, T. Bjåstad, and L. Lovstakken, "Coherent plane wave compounding for very high frame rate ultrasonography of rapidly moving targets," *IEEE Transactions on Medical Imaging*, vol. 32, no. 7, pp. 1265–1276, 2013.

[20] S. Goudarzi, A. Asif, and H. Rivaz, "Fast multi-focus ultrasound image recovery using generative adversarial networks," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 1272–1284, 2020.

[21] J. A. Jensen, "Linear description of ultrasound imaging systems," *Notes for the International Summer School on Advanced Ultrasound Imaging, Technical University of Denmark July*, vol. 5, pp. 54, 1999.

[22] M. K. Feldman, S. Katyal, and M. S. Blackwood, "Us artifacts," *RadioGraphics*, vol. 29, no. 4, pp. 1179–1189, 2009, PMID: 19605664.

[23] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," 2014.

[24] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436, 2015.

[25] Y. You, C. Lu, W. Wang, and C. Tang, "Relative cnn-rnn: Learning relative atmospheric visibility from images," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 45–55, Jan 2019.

[26] G. Cheng, J. Han, P. Zhou, and D. Xu, "Learning rotation-invariant and fisher discriminative convolutional neural networks for object detection," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 265–278, Jan 2019.

[27] J. Liang, Q. Hu, C. Dang, and W. Zuo, "Weighted graph embedding-based metric learning for kinship verification," *IEEE Transactions on Image Processing*, vol. 28, no. 3, pp. 1149–1162, March 2019.

[28] M. Arjovsky and L. Bottou, "Towards principled methods for training generative adversarial networks," *arXiv preprint arXiv:1701.04862*, 2017.

[29] D. Berthelot, T. Schumm, and L. Metz, "Began: boundary equilibrium generative adversarial networks," *arXiv preprint arXiv:1703.10717*, 2017.

[30] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training gans," in *Advances in Neural Information Processing Systems*, 2016, pp. 2234–2242.

[31] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.

[32] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein gan," *arXiv preprint arXiv:1701.07875*, 2017.

[33] B. K. Sriperumbudur, K. Fukumizu, A. Gretton, B. Schölkopf, and G. RG Lanckriet, "On integral probability metrics,\phi-divergences and binary classification," *arXiv preprint arXiv:0901.2698*, 2009.

[34] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of wasserstein gans," in *Advances in Neural Information Processing Systems 30*, pp. 5767–5777. Curran Associates, Inc., 2017.

[35] J. Wu, Z. Huang, J. Thoma, D. Acharya, and L. V. Gool, "Wasserstein divergence for gans," in *The European Conference on Computer Vision (ECCV)*, September 2018.

[36] K. Roth, A. Lucchi, S. Nowozin, and T. Hofmann, "Stabilizing training of generative adversarial networks through regularization," in *Advances in Neural Information Processing Systems 30*, pp. 2018–2028. Curran Associates, Inc., 2017.

[37] R. D. Hjelm, A. P. Jacob, T. Che, A. Trischler, K. Cho, and Y. Bengio, "Boundary-seeking generative adversarial networks," *arXiv preprint arXiv:1702.08431*, 2017.

[38] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M. K. Kalra, Y. Zhang, L. Sun, and G. Wang, "Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1348–1357, June 2018.

[39] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014.

[40] H. Shan, Y. Zhang, Q. Yang, U. Kruger, M. K. Kalra, L. Sun, W. Cong, and G. Wang, "3-d convolutional encoder-decoder network for low-dose ct via transfer learning from a 2-d trained network," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1522–1534, June 2018.

[41] G. Yang, S. Yu, H. Dong, G. Slabaugh, P. L. Dragotti, X. Ye, F. Liu, S. Arridge, J. Keegan, Y. Guo, and D. Firmin, "Dagan: Deep de-aliasing generative adversarial networks for fast compressed sensing mri reconstruction," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1310–1321, June 2018.

[42] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014.

[43] D. Nie, R. Trullo, J. Lian, L. Wang, C. Petitjean, S. Ruan, Q. Wang, and D. Shen, "Medical image synthesis with deep convolutional adversarial networks," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 12, pp. 2720–2730, Dec 2018.

[44] M. Mardani, E. Gong, J. Y. Cheng, S. S. Vasanawala, G. Zaharchuk, L. Xing, and J. M. Pauly, "Deep generative adversarial neural networks for compressive sensing mri," *IEEE Transactions on Medical Imaging*, vol. 38, no. 1, pp. 167–179, Jan 2019.

[45] M. Nikoonahad and D. C. Liv, "Medical ultrasound imaging using neural networks," *Electronics Letters*, vol. 26, no. 8, pp. 545–546, April 1990.

[46] A. C. Luchies and B. C. Byram, "Deep neural networks for ultrasound beamforming," *IEEE Transactions on Medical Imaging*, vol. 37, no. 9, pp. 2010–2021, 2018.

[47] Y. H. Yoon, S. Khan, J. Huh, and J. C. Ye, "Efficient b-mode ultrasound image reconstruction from sub-sampled rf data using deep learning," *IEEE Transactions on Medical Imaging*, vol. 38, no. 2, pp. 325–336, 2019.

[48] D. Hyun, L. L. Brickson, K. T. Looby, and J. J. Dahl, "Beamforming and speckle reduction using neural networks," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 66, no. 5, pp. 898–910, 2019.

[49] M. Gasse, F. Millioz, E. Roux, D. Garcia, H. Liebgott, and D. Friboulet, "High-quality plane wave compounding using convolutional neural networks," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 64, no. 10, pp. 1637–1639, Oct 2017.

[50] Z. Zhou, Y. Wang, J. Yu, Y. Guo, W. Guo, and Y. Qi, "High spatial–temporal resolution reconstruction of plane-wave ultrasound images with a multichannel multiscale convolutional neural network," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 65, no. 11, pp. 1983–1996, 2018.

[51] C. Huang, O. T. Chen, G. Wu, C. Chang, and C. Hu, "Ultrasound imaging improved by the context encoder reconstruction generative adversarial network," in *2018 IEEE International Ultrasonics Symposium*, Oct 2018, pp. 1–4.

[52] F. Dietrichson, E. Smistad, A. Ostvik, and L. Lovstakken, "Ultrasound speckle reduction using generative adversial networks," in *2018 IEEE International Ultrasonics Symposium*, Oct 2018, pp. 1–4.

[53] X. Zhang, J. Li, Q. He, H. Zhang, and J. Luo, "High-quality reconstruction of plane-wave imaging using generative adversarial network," in *2018 IEEE International Ultrasonics Symposium*, Oct 2018, pp. 1–4.

[54] N. Bottenus, "Recovery of the complete data set from focused transmit beams," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 65, no. 1, pp. 30–38, Jan 2018.

[55] N. Bottenus, "Comparison of virtual source synthetic aperture beamforming with an element-based model," *The Journal of the Acoustical Society of America*, vol. 143, no. 5, pp. 2801–2812, 2018.

[56] R. Ali, J. J. Dahl, and N. Bottenus, "Regularized inversion method for frequency-domain recovery of the full synthetic aperture dataset from focused transmissions," in *2018 IEEE International Ultrasonics Symposium*, Oct 2018, pp. 1–9.

[57] A. Ilovitsh, T. Ilovitsh, J. Foiret, D. N. Stephens, and K. W. Ferrara, "Simultaneous axial multifocal imaging using a single acoustical transmission: A practical implementation," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 66, no. 2, pp. 273–284, Feb 2019.

[58] S. Goudarzi, A. Asif, and H. Rivaz, "Multi-focus ultrasound imaging using generative adversarial networks," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, April 2019, pp. 1118–1121.

[59] Balázs Csanád Csáji, "Approximation with artificial neural networks," *Faculty of Sciences, Etvs Lornd University, Hungary*, vol. 24, pp. 48, 2001.

[60] L. Hongzhou and J. Stefanie, "Resnet with one-neuron hidden layers is a universal approximator," in *Advances in Neural Information Processing Systems 31*, pp. 6172–6181. Curran Associates, Inc., 2018.

[61] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, "Context encoders: Feature learning by inpainting," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2536–2544.

[62] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016, pp. 770–778.

[63] J. A. Jensen, "Field: A program for simulating ultrasound systems," in *10th 10th Nordic-Baltic Conference on Biomedical Engineering*, 1996, vol. 4, pp. 351–353.

[64] J. A. Jensen and N. B. Svendsen, "Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 39, no. 2, pp. 262–267, 1992.

[65] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *CoRR*, vol. abs/1412.6980, 2014.

[66] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295–307, Feb 2016.

[67] I. Goodfellow, "Nips 2016 tutorial: Generative adversarial networks," *arXiv preprint arXiv:1701.00160*, 2016.

[68] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A.W.M. van der Laak, Bram van Ginneken, and Clara I. Sánchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60 – 88, 2017.

[69] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional neural networks for medical image analysis: Full training or fine tuning?," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1299–1312, May 2016.

[70] H. Liebgott, A. Rodriguez-Molares, F. Cervenansky, J. A. Jensen, and O. Bernard, "Plane-wave imaging challenge in medical ultrasound," in *2016 IEEE International Ultrasonics Symposium*, 2016, pp. 1–4.

[71] S. Goudarzi, A. Asif, and H. Rivaz, "Ultrasound beamforming using mobilenetv2," in *2020 IEEE International Ultrasonics Symposium*, 2020, pp. 1–4.

[72] S. Goudarzi, A. Asif, and H. Rivaz, "Pruning mobilenetv2 for efficient implementation of minimum variance beamforming," in *Simplifying Medical Ultrasound*. 2021, pp. 211–219, Springer International Publishing.

[73] J. Synnevag, A. Austeng, and S. Holm, "Benefits of minimum-variance beamforming in medical ultrasound imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 56, no. 9, pp. 1868–1879, 2009.

[74] J. Chen, J. Chen, R. Zhuang, and H. Min, "Multi-operator minimum variance adaptive beamforming algorithms accelerated with gpu," *IEEE Transactions on Medical Imaging*, pp. 1–1, 2020.

[75] Avinash S. Vaidya and M.B. Srinivas, "A low-complexity and robust minimum variance beamformer for ultrasound imaging systems using beamspace dominant mode rejection," *Ultrasonics*, vol. 101, pp. 105979, 2020.

[76] S. Afrakhteh and H. Behnam, "Low-complexity adaptive minimum variance ultrasound beamformer based on diagonalization," *Biomedical Signal Processing and Control*, vol. 62, pp. 102110, 2020.

[77] S. Khan, J. Huh, and J. C. Ye, "Adaptive and compressive beamforming using deep learning for medical ultrasound," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, no. 8, pp. 1558–1572, 2020.

[78] B. Luijten, R. Cohen, F. J. de Bruijn, H. A. W. Schmeitz, M. Mischi, Y. C. Eldar, and R. J. G. van Sloun, "Adaptive ultrasound beamforming using deep learning," *IEEE Transactions on Medical Imaging*, vol. 39, no. 12, pp. 3967–3978, 2020.

[79] A. C. Luchies and B. C. Byram, "Assessing the robustness of frequency-domain ultrasound beamforming using deep neural networks," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, no. 11, pp. 2321–2335, 2020.

[80] M. A. L. Bell, J. Huang, D. Hyun, Y. C. Eldar, R. J. G. van Sloun, and M. Mischi, "Challenge on ultrasound beamforming with deep learning (cubdl)," in *2020 IEEE International Ultrasonics Symposium*, 2020, pp. 1–1.

[81] "Challenge on ultrasound beamforming with deep learning (cubdl)," IEEE Dataport, 2019. [Online]. Available: `http://dx.doi.org/10.21227/f0hn-8f92`.

[82] D. Hyun, A. Wiacek, S. Goudarzi, S. Rothlübbers, A. Asif, K. Eickel, Y.C. Eldar, J. Huang, M. Mischi, H. Rivaz, D. Sinden, R.J.G. van Sloun, H. Strohm, and M.A. Lediju Bell, "Deep learning for ultrasound image formation: Cubdl evaluation framework and open datasets," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 68, no. 12, pp. 3466–3483, 2021.

[83] S. Rothlübbers, H. Strohm, K. Eickel, J. Jenne, V. Kuhlen, D. Sinden, and M. Günther, "Improving image quality of single plane wave ultrasound via deep learning based channel compounding," in *2020 IEEE International Ultrasonics Symposium*, 2020, pp. 1–4.

[84] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520.

[85] A. Rodriguez-Molares, O. M. H. Rindal, O. Bernard, A. Nair, M. A. Lediju Bell, H. Liebgott, A. Austeng, and L. Lovstakken, "The ultrasound toolbox," in *2017 IEEE International Ultrasonics Symposium*, 2017, pp. 1–4.

[86] D. Blalock, J. J. G. Ortiz, J. Frankle, and J. Guttag, "What is the state of neural network pruning?," *arXiv preprint arXiv:2003.03033*, 2020.

[87] S. Han, J. Pool, J. Tran, and W. Dally, "Learning both weights and connections for efficient neural network," in *Advances in Neural Information Processing Systems*. 2015, vol. 28, Curran Associates, Inc.

[88] S. Goudarzi and H. Rivaz, "Deep reconstruction of high-quality ultrasound images from raw plane-wave data: A simulation and in vivo study," *Ultrasonics*, vol. 125, pp. 106778, 2022.

[89] M. A. L. Bell, J. Huang, D. Hyun, Y. C. Eldar, R. van Sloun, and M. Mischi, "Challenge on ultrasound beamforming with deep learning (cubdl)," in *2020 IEEE International Ultrasonics Symposium*, 2020, pp. 1–5.

[90] Y. Qi, Y. Guo, and Y. Wang, "Image quality enhancement using a deep neural network for plane wave medical ultrasound imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, in press, 2021.

[91] S. K. Zhou, H. Greenspan, C. Davatzikos, J. S. Duncan, B. V. Ginneken, A. Madabhushi, J. L. Prince, D. Rueckert, and R. M. Summers, "A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises," *Proceedings of the IEEE*, vol. 109, no. 5, pp. 820–838, 2021.

[92] R. J. G. van Sloun, R. Cohen, and Y. C. Eldar, "Deep learning in ultrasound imaging," *Proceedings of the IEEE*, vol. 108, no. 1, pp. 11–29, 2020.

[93] W. Wang, X. G. Xia, C. He, Z. Ren, J. Lu, T. Wang, and B. Lei, "An end-to-end deep network for reconstructing ct images directly from sparse sinograms," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 1548–1560, 2020.

[94] G. Oh, B. Sim, H. Chung, L. Sunwoo, and J. C. Ye, "Unpaired deep learning for accelerated mri using optimal transport driven cyclegan," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 1285–1296, 2020.

[95] J. Quionero-Candela, M. Sugiyama, A. Schwaighofer, and N. D. Lawrence, *Dataset Shift in Machine Learning*, The MIT Press, 2009.

[96] A. A. Nair, K. N. Washington, T. D. Tran, A. Reiter, and M. A. Lediju Bell, "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*, vol. 67, no. 12, pp. 2493–2509, 2020.

[97] E. Mor and A. Bar-Hillel, "A unified deep network for beamforming and speckle reduction in plane wave imaging: A simulation study," *Ultrasonics*, vol. 103, pp. 106069, 2020.

[98] J. Lu, F. Millioz, D. Garcia, S. Salles, W. Liu, and D. Friboulet, "Reconstruction for diverging-wave imaging using deep convolutional neural networks," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, no. 12, pp. 2481–2492, 2020.

[99] J. Youn, M. L. Ommen, M. B. Stuart, E. V. Thomsen, N. B. Larsen, and J. A. Jensen, "Detection and localization of ultrasound scatterers using convolutional neural networks," *IEEE Transactions on Medical Imaging*, vol. 39, no. 12, pp. 3855–3867, 2020.

[100] J. A. Jensen, J. Mathorne, T. Gravesen, and B. Stage, "Deconvolution of in-vivo ultrasound b-mode images," *Ultrasonic Imaging*, vol. 15, no. 2, pp. 122–133, 1993.

[101] Y. Han and J. C. Ye, "Framing u-net via deep convolutional framelets: Application to sparse-view ct," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1418–1429, 2018.

[102] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 2015, pp. 234–241.

[103] J. C. Ye, Y. Han, and E. Cha, "Deep convolutional framelets: A general deep learning framework for inverse problems," *SIAM Journal on Imaging Sciences*, vol. 11, no. 2, pp. 991–1048, 2018.

[104] J. A. Jensen, "Field: A program for simulating ultrasound systems," in *10th Nordic-Baltic Conference on Biomedical Imaging*, 1996, vol. 4, pp. 351–353.

[105] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and F. F. Li, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211–252, 2015.

[106] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba, "Places: A 10 million image database for scene recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 6, pp. 1452–1464, 2018.

[107] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," *arXiv preprint arXiv:1711.05101*, 2017.

[108] M. A. Lediju, G. E. Trahey, B. C. Byram, and J. J. Dahl, "Short-lag spatial coherence of backscattered echoes: imaging characteristics," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 58, no. 7, pp. 1377–1388, 2011.

[109] S.W. Smith, H. Lopez, and W.J. Bodine, "Frequency independent ultrasound contrast-detail analysis," *Ultrasound in Medicine and Biology*, vol. 11, no. 3, pp. 467 – 477, 1985.

[110] A. Rodriguez-Molares, O. M. H. Rindal, J. D'hooge, S. E. Måsøy, A. Austeng, M. A. Lediju Bell, and H. Torp, "The generalized contrast-to-noise ratio: A formal definition for lesion detectability," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, no. 4, pp. 745–759, 2020.

[111] O. Huang, J. Long, W. Long, G. Pinton, G. E. Trahey, and M. L. Palmeri, "Ultranet: Deep learning tools for modeling acoustic wall clutter," in *2020 IEEE International Ultrasonics Symposium*, 2020, pp. 1–4.

[112] G. F. Pinton, G. E. Trahey, and J. J. Dahl, "Sources of image degradation in fundamental and harmonic ultrasound imaging using nonlinear, full-wave simulations," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 58, no. 4, pp. 754–765, 2011.

[113] J. J. Dahl and N. M. Sheth, "Reverberation clutter from subcutaneous tissue layers: Simulation and in vivo demonstrations," *Ultrasound in Medicine and Biology*, vol. 40, no. 4, pp. 714 – 726, 2014.

[114] Z. Allen-Zhu, Y. Li, and Z. Song, "A convergence theory for deep learning via over-parameterization," in *Proceedings of the 36th International Conference on Machine Learning*. 09–15 Jun 2019, vol. 97, pp. 242–252, PMLR.

[115] M. Kumar, N. Houlsby, N. Kalchbrenner, and E. D. Cubuk, "On the surprising tradeoff between imagenet accuracy and perceptual similarity," *arXiv preprint arXiv:2203.04946*, 2022.

[116] X. Huang and S. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[117] S. Goudarzi, A. Asif, and H. Rivaz, "Plane-wave ultrasound beamforming through independent component analysis," *Computer Methods and Programs in Biomedicine*, vol. 203, pp. 106036, 2021.

[118] N. Q. Nguyen and R. W. Prager, "Minimum variance beamformers for coherent plane-wave compounding," in *Medical Imaging 2017: Ultrasonic Imaging and Tomography*. International Society for Optics and Photonics, 2017, vol. 10139, pp. 254 – 263, SPIE.

[119] O. M. H. Rindal and A. Austeng, "Double adaptive plane-wave imaging," in *2016 IEEE International Ultrasonics Symposium*, Sep. 2016, pp. 1–4.

[120] A. M. Deylami, J. A. Jensen, and B. M. Asl, "An improved minimum variance beamforming applied to plane-wave imaging in medical ultrasound," in *2016 IEEE International Ultrasonics Symposium*, 2016, pp. 1–4.

[121] F. Varray, M. Azizian Kalkhoran, and D. Vray, "Adaptive minimum variance coupled with sign and phase coherence factors in iq domain for plane wave beamforming," in *2016 IEEE International Ultrasonics Symposium*, 2016, pp. 1–4.

[122] N. Q. Nguyen and R. W. Prager, "A spatial coherence approach to minimum variance beamforming for plane-wave compounding," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 65, no. 4, pp. 522–534, April 2018.

[123] C. Wang, X. Peng, D. Liang, and H. Zheng, "Plane-wave ultrasound imaging based on compressive sensing with low memory occupation," in *2015 IEEE International Ultrasonics Symposium*, 2015, pp. 1–4.

[124] D. Guillaume, R. Jean-luc, Z. Bo, and L. Andrew F, "Time domain compressive beam forming of ultrasound signals," *The Journal of the Acoustical Society of America*, vol. 137, no. 5, pp. 2773–2784, 2015.

[125] A. Besson, D. Perdios, F. Martinez, Z. Chen, R.E. Carrillo, M. Arditi, Y. Wiaux, and J.P. Thiran, "Ultrafast ultrasound imaging as an inverse problem: Matrix-free sparse image reconstruction," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 65, no. 3, pp. 339–355, 2018.

[126] K. Dei, J. Tierney, and B. Byram, "Aperture domain model image reconstruction (ADMIRE) with plane wave synthesis," in *Medical Imaging 2017: Ultrasonic Imaging and Tomography*. International Society for Optics and Photonics, 2017, vol. 10139, pp. 244 – 253, SPIE.

[127] K. Dei, J. E. Tierney, and B. C. Byram, "Model-based beamforming with plane wave synthesis in medical ultrasound," *Journal of Medical Imaging*, vol. 5, no. 2, pp. 1 – 12, 2018.

[128] J. Y. Lu, "2d and 3d high frame rate imaging with limited diffraction beams," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 44, no. 4, pp. 839–856, 1997.

[129] D. Garcia, L. L. Tarnec, S. Muth, E. Montagnon, J. Porée, and G. Cloutier, "Stolt's f-k migration for plane wave ultrasound imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 60, no. 9, pp. 1853–1867, 2013.

[130] O. Bernard, M. Zhang, F. Varray, P. Gueth, J.P. Thiran, H. Liebgott, and D. Friboulet, "Ultrasound fourier slice imaging: a novel approach for ultrafast imaging technique," in *2014 IEEE International Ultrasonics Symposium*, 2014, pp. 129–132.

[131] T. Szasz, A. Basarab, and D. Kouamé, "Beamforming through regularized inverse problems in ultrasound medical imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 63, no. 12, pp. 2031–2044, 2016.

[132] E. Ozkan, V. Vishnevsky, and O. Goksel, "Inverse problem of ultrasound beamforming with sparsity constraints and regularization," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 65, no. 3, pp. 356–365, 2018.

[133] T. Chernyakova, D. Cohen, M. Shoham, and Y. C. Eldar, "imap beamforming for high-quality high frame rate imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 66, no. 12, pp. 1830–1844, 2019.

[134] R. R. Wildeboer, F. Sammali, R. J. G. Van Sloun, Y. Huang, P. Chen, M. Bruce, C. Rabotti, S. Shulepov, G. Salomon, B. C. Schoot, H. Wijkstra, and M. Mischi, "Blind source separation for clutter and noise suppression in ultrasound imaging: Review for different applications," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, pp. 1–1, 2020.

[135] A. C. H. Yu and L. Lovstakken, "Eigen-based clutter filter design for ultrasound color flow imaging: a review," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 57, no. 5, pp. 1096–1111, 2010.

[136] J. E. Tierney, D. M. Wilkes, and B. C. Byram, "Independent component analysis-based tissue clutter filtering for plane wave perfusion ultrasound imaging," in *Medical Imaging 2019: Ultrasonic Imaging and Tomography*. International Society for Optics and Photonics, 2019, vol. 10955, p. 1095503.

[137] C. M. Gallippi, K. R. Nightingale, and G. E. Trahey, "Bss-based filtering of physiological and arfi-induced tissue and blood motion," *Ultrasound in medicine & biology*, vol. 29, no. 11, pp. 1583–1592, 2003.

[138] P. Li, X. Yang, D. Zhang, and Z. Bian, "Adaptive clutter filtering based on sparse component analysis in ultrasound color flow imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 55, no. 7, pp. 1582–1596, 2008.

[139] C. M. Gallippi and G. E. Trahey, "Adaptive clutter filtering via blind source separation for two-dimensional ultrasonic blood velocity measurement," *Ultrasonic imaging*, vol. 24, no. 4, pp. 193–214, 2002.

[140] K. Dei, S. Schlunk, and B. Byram, "Computationally efficient implementation of aperture domain model image reconstruction," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 66, no. 10, pp. 1546–1559, 2019.

[141] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural Networks*, vol. 13, no. 4, pp. 411 – 430, 2000.

[142] A. Hyvärinen, "New approximations of differential entropy for independent component analysis and projection pursuit," in *Advances in neural information processing systems*, 1998, pp. 273–279.

[143] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp. 626–634, May 1999.

[144] J. H. Friedman and J. W. Tukey, "A projection pursuit algorithm for exploratory data analysis," *IEEE Transactions on Computers*, vol. C-23, no. 9, pp. 881–890, Sep. 1974.

[145] P.C. Li and M.L. Li, "Adaptive imaging using the generalized coherence factor," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 50, no. 2, pp. 128–141, 2003.

[146] O. M. H. Rindal, A. Austeng, A. Fatemi, and A. Rodriguez-Molares, "The effect of dynamic range alterations in the estimation of contrast," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 66, no. 7, pp. 1198–1208, July 2019.

[147] R. F. Wagner, "Statistics of speckle in ultrasound b-scans," *IEEE Trans. Sonics & Ultrason.*, vol. 30, no. 3, pp. 156–163, 1983.

[148] E. Bingham and A. Hyvärinen, "A fast fixed-point algorithm for independent component analysis of complex valued signals," *International journal of neural systems*, vol. 10, no. 01, pp. 1–8, 2000.

[149] S. Goudarzi, A. Basarab, and H. Rivaz, "Inverse problem of ultrasound beamforming with denoising-based regularized solutions," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, pp. 1–1, 2022.

[150] N. Ozmen, R. Dapp, M. Zapf, H. Gemmeke, N. V. Ruiter, and K. W. A. van Dongen, "Comparing different ultrasound imaging methods for breast cancer detection," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 62, no. 4, pp. 637–646, 2015.

[151] K. W. Hollman, K. W. Rigby, and M. O'Donnell, "Coherence factor of speckle from a multi-row probe," in *1999 IEEE Ultrasonics Symposium*, 1999, vol. 2, pp. 1257–1260 vol.2.

[152] G. Matrone, A. S. Savoia, G. Caliano, and G. Magenes, "The delay multiply and sum beamforming algorithm in ultrasound b-mode medical imaging," *IEEE Transactions on Medical Imaging*, vol. 34, no. 4, pp. 940–949, 2015.

[153] N. Wagner, Y.C. Eldar, and Z. Friedman, "Compressed beamforming in ultrasound imaging," *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4643–4657, 2012.

[154] S. Goudarzi, A. Basarab, and H. Rivaz, "A unifying approach to inverse problems of ultrasound beamforming and deconvolution," *arXiv preprint arXiv:2112.14294*, 2021.

[155] A. Achim, A. Basarab, G. Tzagkarakis, P. Tsakalides, and D. Kouamé, "Reconstruction of ultrasound rf echoes modeled as stable random variables," *IEEE Transactions on Computational Imaging*, vol. 1, no. 2, pp. 86–95, 2015.

[156] S. H. Chan, X. Wang, and O. A. Elgendy, "Plug-and-play admm for image restoration: Fixed-point convergence and applications," *IEEE Transactions on Computational Imaging*, vol. 3, no. 1, pp. 84–98, 2017.

[157] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *2013 IEEE Global Conference on Signal and Information Processing*, 2013, pp. 945–948.

[158] S. Sreehari, S. V. Venkatakrishnan, B. Wohlberg, G. T. Buzzard, L. F. Drummy, J. P. Simmons, and C. A. Bouman, "Plug-and-play priors for bright field electron tomography and sparse interpolation," *IEEE Transactions on Computational Imaging*, vol. 2, no. 4, pp. 408–423, 2016.

[159] Y. Romano, M. Elad, and P. Milanfar, "The little engine that could: Regularization by denoising (red)," *SIAM Journal on Imaging Sciences*, vol. 10, no. 4, pp. 1804–1844, 2017.

[160] K. Zhang, Y. Li, W. Zuo, L. Zhang, L. Van Gool, and R. Timofte, "Plug-and-play image restoration with deep denoiser prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2021.

[161] J. He, . Yang, Y. Wang, D. Zeng, Z. Bian, H. Zhang, J. Sun, Z. Xu, and J. Ma, "Optimizing a parameterized plug-and-play admm for iterative low-dose ct reconstruction," *IEEE Transactions on Medical Imaging*, vol. 38, no. 2, pp. 371–382, 2019.

[162] S. Hurault, A. Leclaire, and N. Papadakis, "Gradient step denoiser for convergent plug-and-play," *arXiv preprint arXiv:2110.03220*, 2021.

[163] R. Cohen, Y. Blau, D. Freedman, and E. Rivlin, "It has potential: Gradient-driven denoisers for convergent solutions to inverse problems," in *Advances in Neural Information Processing Systems*. 2021, vol. 34, pp. 18152–18164, Curran Associates, Inc.

[164] S. Hurault, A. Leclaire, and N. Papadakis, "Proximal denoiser for convergent plug-and-play optimization with nonconvex regularization," *arXiv preprint arXiv:2201.13256*, 2022.

[165] E. T. Reehorst and P. Schniter, "Regularization by denoising: Clarifications and new interpretations," *IEEE Transactions on Computational Imaging*, vol. 5, no. 1, pp. 52–67, 2019.

[166] M. Mani, V. A. Magnotta, and M. Jacob, "qmodel: A plug-and-play model-based reconstruction for highly accelerated multi-shot diffusion mri using learned priors," *Magnetic Resonance in Medicine*, vol. 86, no. 2, pp. 835–851, 2021.

[167] E. M. Eksioglu and A. K. Tanc, "Denoising amp for mri reconstruction: Bm3d-amp-mri," *SIAM Journal on Imaging Sciences*, vol. 11, no. 3, pp. 2090–2109, 2018.

[168] H. Gupta, K. H. Jin, H. Q. Nguyen, M. T. McCann, and M. Unser, "Cnn-based projected gradient descent for consistent ct image reconstruction," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1440–1453, 2018.

[169] Z. Chen, A. Basarab, and D. Kouamé, "Compressive deconvolution in medical ultrasound imaging," *IEEE Transactions on Medical Imaging*, vol. 35, no. 3, pp. 728–737, 2016.

[170] M. Hourani, A. Basarab, D. Kouamé, and J.Y. Tourneret, "Ultrasound image deconvolution using fundamental and harmonic images," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 68, no. 4, pp. 993–1006, 2021.

[171] O. V. Michailovich and A. Tannenbaum, "Despeckling of medical ultrasound images," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 53, no. 1, pp. 64–78, 2006.

[172] P. C. Tay, C. D. Garson, S. T. Acton, and J. A. Hossack, "Ultrasound despeckling for contrast enhancement," *IEEE Transactions on Image Processing*, vol. 19, no. 7, pp. 1847–1860, 2010.

[173] H. Liebgott, R. Prost, and D. Friboulet, "Pre-beamformed rf signal reconstruction in medical ultrasound using compressive sensing," *Ultrasonics*, vol. 53, no. 2, pp. 525–533, 2013.

[174] O. Lorintiu, H. Liebgott, M. Alessandrini, O. Bernard, and D. Friboulet, "Compressed sensing reconstruction of 3d ultrasound data using dictionary learning and line-wise subsampling," *IEEE Transactions on Medical Imaging*, vol. 34, no. 12, pp. 2467–2477, 2015.

[175] G. David, J. Robert, B. Zhang, and A. F. Laine, "Time domain compressive beam forming of ultrasound signals," *The Journal of the Acoustical Society of America*, vol. 137, no. 5, pp. 2773–2784, 2015.

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

[177] T. Chernyakova and Y.C. Eldar, "Fourier-domain beamforming: the path to compressed ultrasound imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 61, no. 8, pp. 1252–1267, 2014.

[178] T. Szasz, A. Basarab, and D. Kouamé, "L1-norm regularized beamforming in ultrasound imaging," in *2016 IEEE International Ultrasonics Symposium*, 2016, pp. 1–3.

[179] J. Dahl and D. Hyun, "Neural networks as an applied tool for ultrasound beamforming and image reconstruction," *The Journal of the Acoustical Society of America*, vol. 148, no. 4, pp. 2446–2446, 2020.

[180] L. L. Brickson, D. Hyun, M. Jakovljevic, and J. J. Dahl, "Reverberation noise suppression in ultrasound channel signals using a 3d fully convolutional neural network," *IEEE Transactions on Medical Imaging*, vol. 40, no. 4, pp. 1184–1195, 2021.

[181] Z. Zhou, Y. Guo, and Y. Wang, "Ultrasound deep beamforming using a multiconstrained hybrid generative adversarial network," *Medical Image Analysis*, vol. 71, pp. 102086, 2021.

[182] J. Zhang, Q. He, Y. Xiao, H. Zheng, C. Wang, and J. Luo, "Ultrasound image reconstruction from plane wave radio-frequency data by self-supervised deep neural network," *Medical Image Analysis*, vol. 70, pp. 102018, 2021.

[183] J. Lu, F. Millioz, D. Garcia, S. Salles, D. Ye, and D. Friboulet, "Complex convolutional neural networks for ultrafast ultrasound imaging reconstruction from in-phase/quadrature signal," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 69, no. 2, pp. 592–603, 2022.

[184] A. Mamistvalov, A. Amar, N. Kessler, and Y. C. Eldar, "Deep-learning based adaptive ultrasound imaging from sub-nyquist channel data," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 69, no. 5, pp. 1638–1648, 2022.

[185] A. Mamistvalov and Y. C. Eldar, "Deep unfolded recovery of sub-nyquist sampled ultrasound images," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 68, no. 12, pp. 3484–3496, 2021.

[186] R. F. Wagner, M. F. Insana, and D. G. Brown, "Statistical properties of radio-frequency and envelope-detected signals with applications to medical ultrasound," *JOSA A*, vol. 4, no. 5, pp. 910–922, 1987.

[187] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen, *Variational Methods in Imaging*, Applied Mathematical Sciences. Springer New York, 1 edition, 2008.

[188] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*, Springer Texts in Statistics. Springer New York, NY, 2 edition, 2021.

[189] P. L. Combettes and J. C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-point algorithms for inverse problems in science and engineering*, pp. 185–212. Springer, 2011.

[190] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[191] C. A. Bouman, "Model based imaging," 2013 [Online], Available: https://engineering.purdue.edu/ bouman/publications/pdf/MBIPbook.pdf.

[192] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer Series in Operations Research and Financial Engineering. Springer New York, 2 edition, 2006.

[193] P. Coupe, P. Hellier, C. Kervrann, and C. Barillot, "Nonlocal means-based speckle filtering for ultrasound images," *IEEE Transactions on Image Processing*, vol. 18, no. 10, pp. 2221–2229, 2009.

[194] J. Immerkær, "Fast noise variance estimation," *Computer Vision and Image Understanding*, vol. 64, no. 2, pp. 300–302, 1996.

[195] E. E. Hundt and E. A. Trautenberg, "Digital processing of ultrasonic data by deconvolution," *IEEE Transactions on Sonics and Ultrasonics*, vol. 27, no. 5, pp. 249–252, 1980.

[196] J.A. Jensen, J. Mathorne, T. Gravesen, and B. Stage, "Deconvolution of in-vivo ultrasound b-mode images," *Ultrasonic Imaging*, vol. 15, no. 2, pp. 122–133, 1993.

[197] O. Michailovich and A. Tannenbaum, "Blind deconvolution of medical ultrasound images: A parametric inverse filtering approach," *IEEE Transactions on Image Processing*, vol. 16, no. 12, pp. 3005–3019, 2007.

[198] C. Chen, M.K. Ng, and X.L. Zhao, "Alternating direction method of multipliers for nonlinear image restoration problems," *IEEE Transactions on Image Processing*, vol. 24, no. 1, pp. 33–43, 2015.

[199] A. Agarwal, J. Reeg, A.S. Podkowa, and M.L. Oelze, "Improving spatial resolution using incoherent subtraction of receive beams having different apodizations," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 66, no. 1, pp. 5–17, 2019.

[200] J.F. Synnevag, A. Austeng, and S. Holm, "Adaptive beamforming applied to medical ultrasound imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 54, no. 8, pp. 1606–1613, 2007.

[201] O.M.H. Rindal, J.P. Asen, S. Holm, and A. Austeng, "Understanding contrast improvements from capon beamforming," in *2014 IEEE International Ultrasonics Symposium*, 2014, pp. 1694–1697.

[202] C.-I.C. Nilsen and I. Hafizovic, "Beamspace adaptive beamforming for ultrasound imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 56, no. 10, pp. 2187–2197, 2009.

[203] A.C. Jensen and A. Austeng, "An approach to multibeam covariance matrices for adaptive beamforming in ultrasonography," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 59, no. 6, pp. 1139–1148, 2012.

[204] K. Kim, S. Park, J. Kim, S.B. Park, and M. Bae, "A fast minimum variance beamforming method using principal component analysis," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 61, no. 6, pp. 930–945, 2014.

[205] M. Bae, S.B. Park, and S.J. Kwon, "Fast minimum variance beamforming based on legendre polynomials," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 63, no. 9, pp. 1422–1431, 2016.

[206] R. Mallart and M. Fink, "Adaptive focusing in scattering media through sound-speed inhomogeneities: The van cittert zernike approach and focusing criterion," *The Journal of the Acoustical Society of America*, vol. 96, no. 6, pp. 3721–3732, 1994.

[207] K. Nagai, "A new synthetic-aperture focusing method for ultrasonic b-scan imaging by the fourier transform," *IEEE Transactions on Sonics and Ultrasonics*, vol. 32, no. 4, pp. 531–536, 1985.

[208] L.J. Busse, "Three-dimensional imaging using a frequency-domain synthetic aperture focusing technique," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 39, no. 2, pp. 174–179, 1992.

[209] P. Kruizinga, F. Mastik, N. de Jong, A.F.W. van der Steen, and G. van Soest, "Plane-wave ultrasound beamforming using a nonuniform fast fourier transform," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 59, no. 12, pp. 2684–2691, 2012.

[210] T. Chernyakova, R. Cohen, R. Mulayoff, Y. Sde-Chen, C. Fraschini, J. Bercoff, and Y.C. Eldar, "Fourier-domain beamforming and structure-based reconstruction for plane-wave imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 65, no. 10, pp. 1810–1821, 2018.

[211] T. Taxt and J. Strand, "Two-dimensional noise-robust blind deconvolution of ultrasound images," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 48, no. 4, pp. 861–866, 2001.

[212] C. Yu, C. Zhang, and L. Xie, "A blind deconvolution approach to ultrasound imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 59, no. 2, pp. 271–280, 2012.

[213] L. Roquette, M. Simeoni, P. Hurley, and A. Besson, "On an analytical, spatially-varying, point-spread-function," in *2017 IEEE International Ultrasonics Symposium*, 2017, pp. 1–4.

[214] T. Taxt, "Restoration of medical ultrasound images using two-dimensional homomorphic deconvolution," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 42, no. 4, pp. 543–554, 1995.

[215] J. Eckstein and W. Yao, "Understanding the convergence of the alternating direction method of multipliers: Theoretical and computational perspectives," *Pacific Journal of Optimization*, vol. 11, no. 4, pp. 619–644, 2015.

[216] N. Bottenus, B. C. Byram, and D. Hyun, "Histogram matching for visual ultrasound image comparison," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 68, no. 5, pp. 1487–1495, 2021.

[217] D. Hyun, G. B. Kim, N. Bottenus, and J. J. Dahl, "Ultrasound lesion detectability as a distance between probability measures," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 69, no. 2, pp. 732–743, 2022.

[218] O.V. Michailovich and D. Adam, "A novel approach to the 2-d blind deconvolution problem in medical ultrasound," *IEEE Transactions on Medical Imaging*, vol. 24, no. 1, pp. 86–104, 2005.

[219] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[220] M. Amiri, R. Brooks, and H. Rivaz, "Fine-tuning u-net for ultrasound image segmentation: Different layers, different outcomes," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, no. 12, pp. 2510–2518, 2020.

[221] A. Salari and B. M. Asl, "User parameter-free minimum variance beamformer in medical ultrasound imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 68, no. 7, pp. 2397–2406, 2021.

[222] M. Sharifzadeh, H. Benali, and H. Rivaz, "Investigating shift variance of convolutional neural networks in ultrasound image segmentation," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 69, no. 5, pp. 1703–1713, 2022.

[223] M. Sharifzadeh, H. Benali, and H. Rivaz, "Phase aberration correction: A convolutional neural network approach," *IEEE Access*, vol. 8, pp. 162252–162260, 2020.