TEL AVIV UNIVERSITY

The Iby and Aladar Fleischman Faculty of Engineering

The Zandman-Slaner School of Graduate Studies

Deep Learning Enhanced Ultrasound Beam Shaping for Acoustic Hologram Generation in Non-Continuous Wave Transmission Settings

A thesis submitted toward the degree of

Master of Science in Electrical Engineering

by

Dror Schein

This research was carried out in the Department of Biomedical Engineering

Under the supervision of Dr. Tali Ilovitsh

December 23

TEL AVIV UNIVERSITY

The Iby and Aladar Fleischman Faculty of Engineering

The Zandman-Slaner School of Graduate Studies

Deep Learning Enhanced Ultrasound Beam Shaping for Acoustic Hologram Generation in Non-Continuous Wave Transmission Settings

A thesis submitted toward the degree of

Master of Science in Electrical Engineering

by

Dror Schein

This research was carried out in the Department of Biomedical Engineering

Under the supervision of Dr. Tali Ilovitsh

December 23

Acknowledgments

I would like to thank my supervisor, Dr. Tali Ilovitsh for her support, understanding and confidence during this work. During the research Tali was understanding and supportive of the problems I encountered, while successfully guiding me towards achieving the research goals.

I would also like to thank my parents and my family, who have always encouraged me to pursue my goals and continue to develop myself. This work would not have been possible without their support.

Above all, I wish to express my gratitude to my wife, Shuvael, and dedicate this thesis to her. She has always had an unyielding belief in my abilities and unwavering encouragement throughout the entirety of this thesis. Her love and support have been indispensable to my success.

Abstract

Ultrasound imaging is a safe, cost effective, real time, and reliable method to perform medical imaging deep inside the human body, and as such has gained its widespread usage. Standard ultrasound imaging is performed with transducers that contain an array of elements, using two way focusing. On transmit, a focused wavefront is generated by adding parabolic-shaped delays to each of the transducer elements, such that all the transmitted pulses will constructively interfere at a specific distance and create a focal spot. This method yields enhanced lateral resolution, signal to noise ratio, and penetration depth compared to other ultrasound imaging method.

A major drawback of two way focusing stems from the generation of a single focal spot each transmission, requiring steering the focal spot across the image. This process is time consuming and limits the frame rate. Alternatively, multifoci imaging can be used to increase the frame rate. This can achieved via controlled beam shaping by engineering the transmitted phases to create a desired pattern. Optically-inspired phase retrieval algorithms and standard beam shaping methods successfully generate acoustic holograms for therapeutic applications that involve long burst transmissions. However, a phase engineering technique, designed for single-cycle transmission and capable of achieving spatiotemporal interference of the transmitted pulses is needed for imaging applications.

Towards this goal, we developed a multi-level residual deep convolutional network for calculating the inverse process that will yield the phase map for the creation of a multifoci pattern. The ultrasound deep learning (USDL) method was trained on simulated training pairs of multifoci patterns in the focal plane and their corresponding phase maps in the transducer plane, where propagation between the planes was performed via singe cycle transmission. The USDL method outperformed the standard Gerchberg-Saxton method, when transmitted with

single cycle excitation, in parameters including the number of focal spots that were generated successfully, and their pressure and uniformity. In addition, the USDL method was shown to be flexible in generating patterns with large focal spacing, uneven spacing and nonuniform amplitudes. In simulations, the largest improvement was obtained for 4 foci patterns, where the Gerchberg-Saxton method succeeded in creating 25% of the requested patterns, while the USDL method successfully created 60% of the patterns. These results were confirmed experimentally via hydrophone measurements.

Our findings suggest that deep learning based beam shaping can facilitate the next generation of acoustical holograms for ultrasound imaging applications.

Contents

Abbreviations and symbols

BF ..... Beamforming

B-mode ..... Brightness mode

CW ..... Continuous Wave

DAS ..... Delay and Sum

FC ..... Fully Connected

FOV ..... Field Of View

FWHM ..... Full Width Half max

GS ..... Gerchberg-Saxton

MI ..... Mechanical Index

MSE ..... Mean squared error

λ ..... Wavelength

τ $[sec]$ ..... Time delay

b ..... Bias term

D ..... Aperture width

f ..... Frequency

Fr $\left[\frac{1}{sec}\right]$ ..... Pulse repetetion frequency

FR [Hz] ..... Frame rate

N_{line} ..... Number of lines

T_{frame} $[sec]$ ..... Time per frame

NN ..... Neural Network

ReLU ..... Rectified Linear Unit

ResNet ..... Residual Network

RF ..... Radio frequency

ROI ..... Region Of Interest

SGD ..... Stochastic Gradient Descent

SNR ..... Signal to Noise Ration

TIS ..... Thermal Index

US ..... Ultrasound

USDL ..... Ultrasound Deep Learning

t_{line} $[sec]$ ..... Time for single line

Tr [sec] ..... Pulse repetition interval

v $[\frac{m}{sec^2}]$ ..... Speed of sound

W_n ..... Weights of layer n

W_t^{ij} ..... Weight i in layer j at time t

z [mm] ..... Depth

φ ..... Phase

A ..... Amplitude

List of figures

List of Publications

Journal paper:

- **Dror Schein,** Tal Grutman and Tali Ilovitsh, “Deep learning-based ultrasound beam shaping for spatiotemporal acoustic holograms generation", published in IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol 70, no. 6, 551-561 2023.

1 Introduction

Ultrasound (US) imaging is a widely used medical diagnostics technique enabling a real-time, deep penetrating, cost-effective and reliable method to perform medical imaging. US imaging is being used extensively in practice, with examples such as cancer detection [1], [2], genitourinary imaging, echocardiography [3], obstetrics, point of care ultrasound, and many more. In addition, US has also found usage in the therapeutic setting, with examples such as thermal and mechanical ablations, drug and gene delivery, and brain therapy [4]. The widespread adoption of US stems from its many advantages over standard medical imaging methods. For example. Since US does not utilize dangerous radiation, which is in contrast to other methods such as CT and X-Ray, US is safe for extensive usage. In addition, US systems are getting cheaper and smaller, allowing easy usage and enabling new use cases such as point of care US [5].

Due to the many advantages offered by US, there is extensive research into potential improvements for US systems, both in the technical aspects and in the potential use cases aspect. As more and more improvements are made, novel methods of utilizing US have been enabled and found success [6]. One significant development has been programmable US systems, which offer researchers complete control of the US system, from the transmission stage until the post processing stage. By utilizing these systems, researchers are able to work on new methods which would not have been possible otherwise. For example, these systems allow us to manipulate the phases and frequencies of the waves transmitted separately for each element in the transducer.

In standard US imaging, the most common transmission and receive method is two way focusing. During the transmission phase, all the ultrasound elements are focused to a single spot in order to create constructive interference and generate a focus point. On receive, beamforming

is applied to focus the received waves. Since a single focal spot is generated in each transmit and receive event, a full 2D image requires sweeping the focal spot across a lateral line at single depth [7]. The transmission focusing is performed by utilizing a phased array, such that parabolic shaped delays are added to each of the transducer elements. The necessary delays to be transmitted are computed in a geometric fashion, based on the speed of sound in the medium, the transducer elements location, and the focal spot location. This method is simple to implement, while yielding reasonable resolution, signal to noise (SNR) ratio and penetration depth.

A major drawback of using two way focusing is the time necessary to scan a full image. The time to scan an image is directly correlated to the number of focal spots generated, and since a full 2D image is required the number of focal spots can be large. This problem results in reduced frame rate, which can impede the usage and effectiveness of US. An alternative approach to US transmission is multifoci transmission. In this method, the transmitted wavefront is constructed such that instead of creating a single focal point, a pattern of multiple focal points are generated simultaneously. This reduces the time necessary to acquire images, allowing increased frame rate with minimal resolution costs.

In general, there are two main categories of methods to generate these multifoci patterns. The first category, which includes methods such as SAMI [8], [9] and multiline transmit methods [10], [11], involve super positioning several wavefronts into a single transmitted wavefront that will yield the requested pattern. These methods tend to perform sub-optimally, since some of the generated foci are created only with a fraction of the actual transducer, greatly limiting the resulting resolution and contrast. In addition, these methods need longer transmission bursts, which can increase the thermal index (TIS).

The second category of methods involve acoustical beam shaping in order to construct the necessary wavefront to be transmitted. Examples of methods in this category are the conjugate field method [12], the pseudo inverse method [13], and the Gerchberg-Saxton (GS) algorithm [14]. Of these, the GS method yields the best results in terms of the efficiency and focal spot uniformity [15]. This category of methods has demonstrated superior performance relative to the previous category, and has found usage in cases such as generating a uniform temperature field [16] drug delivery guidance [17], ultrasonic neuro-modulation [15], blood brain barrier opening [18], skull aberration correction [19], and acoustic hologram generation [20]. In addition, The GS algorithm was also previously used for achieving ultrasound super resolution imaging via acoustical structured illumination [21].

While this category of methods has indeed found usage, it has been mostly limited to the therapeutic setting. This is due to an underlying assumption that the transmitted wave is a continuous wave (CW), rather than a single cycle transmission. Although this assumption holds true for the therapeutic setting, in imaging applications each transmission is a single cycle transmission in order to maximize axial resolution. Due to this, these methods do not take into consideration the time domain, and the patterns they generate are sub-optimal in terms of the number of generated focal spots, their shape and their amplitudes. This reduction in performance poses a limitation on the effective possible pattern width, which indirectly also affect the number of focus points that can be generated and the imaged region of interest (ROI). In order words, In order to create patterns that are suitable for imaging applications, it is essential to design acoustic holograms that impose spatiotemporal superposition of single cycle transmission.

The central problem in developing such algorithms stems from the way temporal propagation of acoustic waves is computed. In order to compute the propagation analytically, the methods

discussed above utilize the time independent diffraction theory, which is the core reason for their sub optimality. The other option is to compute the acoustic field numerically, which is the method utilized in Field II [22], [23]. While this method indeed offers accurate results, it is difficult to invert. To this end, we propose a deep learning based method to calculate the phase maps that are required for generating a specific pattern while transmitting in single cycle mode.

Deep learning is a rapidly growing field with extensive applications in medical imaging that outperform standard methods. These models learn a complicated representation of the data and utilize it to perform a specific task. To accomplish this, a series of non-linear modules are combined in a layered structure, such that each module adds to the overall data representation. The resulting model is then trained on a significant amount of data to teach it to perform a specific task. The model learns the relationship between the input it receives and the output it should generate and can then be used during inference time to calculate the result for new data [24].

Deep learning has found usage in the field of medical imaging in general, and in ultrasound in particular. Some examples of its usage in medical imaging include cancer segmentation [25], synthesizing CT images from MRI images[26], and super resolution [27]. In ultrasound, deep learning has been used for classifying breast tumors [28], [29] and liver cancer [30], while [31] utilized it to achieve super resolution in ultrasound images.

Recently, [32], [33] have utilized deep learning to perform acoustic beam shaping. While these methods performed better than GS, they were trained on data that assumed CW transmission, and thus suffer from the same problem discussed above. In order to circumvent this problem, we aim to train the model on data generated via single cycle transmission, allowing the model to learn the inverse relationship between the transducer plane and the imaging plane while taking into consideration the time domain.

This thesis describes the construction and implementation of the USDL model, the simulation results and the validation performed via US experiments with hydrophone scans. Firstly, the main objectives of this research are presented in chapter 2. Then, chapter 3 covers the theoretical background necessary for this work, including general principles of US imaging, an introduction to deep learning, and the motivation for this new method. Chapter 4 will include the methods and material which were utilized in this research. Afterwards, chapter 5 will present the results of the model, including the simulation results and the validation tests conducted via hydrophone measurements. Finally, chapter 6 will discuss the applications and limitations of the proposed model and draw conclusions for further work.

2 Research objectives

This work aims to construct a deep learning based method to perform ultrasound beam shaping for spatiotemporal acoustic hologram generation in ultrasound imaging. By considering the time domain in the training process, we aim to achieve spatiotemporal constructive interference, and thus surpass traditional methods in ultrasound beam shaping. The main objective the research has that is not feasible with GS is incorporating the time domain when computing the necessary phase maps.

3 Theoretical Background

For a better understanding of the thesis and the proposed US imaging method, the main principles of US imaging in general and then of the conventional method used nowadays should be presented. Moreover, an introduction to deep learning will be supplied as well. This chapter

will present the most important aspects of US imaging[7], [34]. The sections concerning deep learning will be primarily based on [24].

3.1 Ultrasound imaging principles

US is a medical imaging modality that employs high-frequency sound waves beyond the range of human auditory perception [35], [36]. As a wave, ultrasound carries energy akin to an electromagnetic wave. The potential of US as a diagnostic imaging modality was first discovered during World War II, when multiple research groups, inspired by sonar and radar technologies, began investigating its diagnostic capabilities. The most commonly used form of US imaging is the pulse-echo technique, which shares fundamental principles with sonar and radar systems. In essence, this technique involves transmitting an acoustic signal and detecting the echoes from the medium being scanned to form an image. Despite its use in diagnosing various medical conditions, the widespread acceptance of pulse-echo US as a diagnostic tool only occurred with the advent of grayscale images, also known as B-mode images.

US imaging utilizes piezoelectric transducers to convert electrical signals to acoustic signals (and vice versa), allowing a non-invasive method of imaging human anatomy. Since the piezoelectric elements can both receive and transmit acoustic signals, the same transducer is employed for both transmission and reception of the ultrasound waves within the medium. Standard ultrasound imaging methods involve transmitting short acoustic pulses at a designated center frequency and focusing them at a desired depth, determined according to the region of interest.[37].

The transmitted US sound wave is propagated through the medium, such that the properties of the medium have a direct influence on the behavior of the wave. Due to the inhomogeneity of

the medium, which is caused by the variety of tissues and organs the human body is composed of, the acoustic waves are partially reflected and scattered, generating echoes that propagate back toward the transducer with the same frequency spectrum.

The received waves are detected via the piezoelectric elements of the transducer, which convert the received waves into an electrical signal. The electric signal is then processed to form a brightness image, known as a B-mode image. In the image generated, the anatomic structure sonicated is quantized into pixels whose brightness corresponds to the strength of the echo received from the regions covered by them.

3.2 Two way focusing

Two way focusing is the most common method of generating B-mode images, as it yields high quality images and is fairly easy to compute and implement. This method includes steering a focused transmitted beam across the field-of-view (FOV), and then focusing the received beam in order to yield an image. Since a typical US transducer consists of many piezoelectric elements, all these elements are focused to a single spot by controlling their phase. Once the echoes are received, the transducer transmits a new beam which is focused to a different focal spot. This process is repeated across the whole FOV, yielding a scanning line. A single image is formed by sweeping the beams along multiple scanning line throughout the FOV such that a typical two-way focusing images are generally composed of 64 to 512 scan lines.

Focus spots are generated by achieving constructive interference of the excitation signals of all transducer elements at a specific location. This process is achieved electronically by controlling the delay of each excitation signal. Different elements transmit their signals at different times, while their emission is delayed relative to the element closest to the focal location. Specifically,

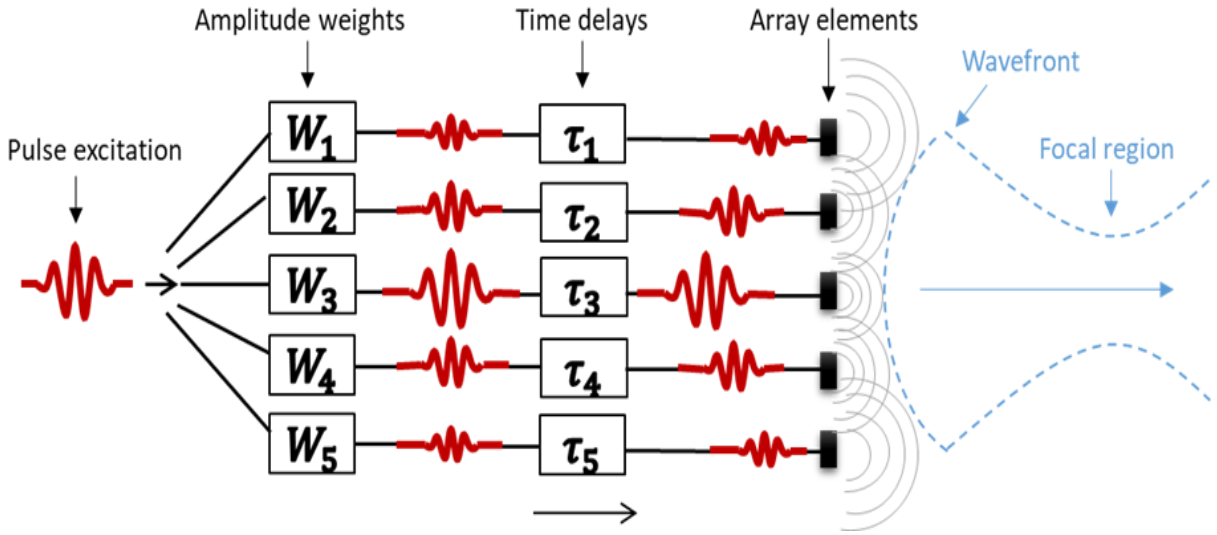

Figure 1. Electronic focusing with a multi-elements array at the transmit stage. Apodization and time delay are applied to each element's excitation pulse, such that constructive interference is generated at the intended focus point.

the transmission time delay of each element is determined by the difference between the time required for its excitation signal to reach the focal location and the time required for the closest element's excitation signal to reach the same location on the transducer (see Figure 1). In order to compute the time delay τ needed for the element i (x_i, y_i, z_i) to be focused to a focal spot f (x_f, y_f, z_f), we can define [38]:

$$\tau_i(x_f, y_f, z_f) = \frac{1}{c}\left(\sqrt{(x_c - x_f)^2 + (y_c - y_f)^2 + (z_c - z_f)^2} - \sqrt{(x_i - x_f)^2 + (y_i - y_f)^2 + (z_i - z_f)^2}\right) \quad (1)$$

Such that c is the speed of sound through the medium, and (x_c, y_c, z_c) is the location of the center of the transducer. In general, we can approximate the speed of sound through water as $1490\ [\frac{m}{s}]$, and through soft tissue in the body as $1540\ [\frac{m}{s}]$. Since this computation is simple and the actual delay is relatively easy to implement electronically, this method of transmission has gained widespread usage.

After the transmission stage, the echo-responses detected from the different scanning directions by the transducer's elements yield an electric signal termed as raw channel data. This raw channel

data is passed into the main stage of the image reconstruction process, which is known as the receive beamforming (BF) stage. The name of the method, two way focusing, stems from the focusing both in transmission and both on receive via BF. The standard method of receive BF is called delay and sum (DAS), which involves integrating the received signals reflected by the transducer with appropriate weighting and time delays, and then summing them in order to yield an RF line (Figure 2). This process enables the deliberate prioritization of echoes that originate from known positions, while signals reflected from other positions are artificially attenuated. DAS improves image quality, both in terms of lateral resolution and in terms of contrast and SNR. In addition, this process allows echoes from different depths to still be in focus. Since DAS can be applied in real time, it has seen widespread adoption in commercial use, and is the standard method of receive BF. We can model the DAS process as follows:

$$RF(t) = \sum_{i=1}^{N} W_i y_i(t - \tau_i) \ , \qquad \tau_i = \frac{\left|\vec{r}_c - \vec{r}_f\right| - \left|\vec{r}_i - \vec{r}_f\right|}{c} \ , \tag{2}$$

Such that *N* is the number of elements in the array, y_i represents the received echo detected by the *i*-th element, τ_i is its time delay required for focusing it at the sampled point of interest and W_i is the apodization weight applied on the received y_i echo. $\vec{r}_i$, $\vec{r}_c$, $\vec{r}_f$ are the coordinate vectors (x, y, z) of the *i*-th element, the element closest to the artificial focus point, and the artificial focus point, respectively, and *c* is the speed of sound in the medium.

After enough RF lines are acquired, all of them are combined into a matrix called the RF image. This RF image is passed on to additional stages for further processing, in order to create the final B-mode image. These post processing stages are out of the scope of this thesis, and as such will not be reviewed.

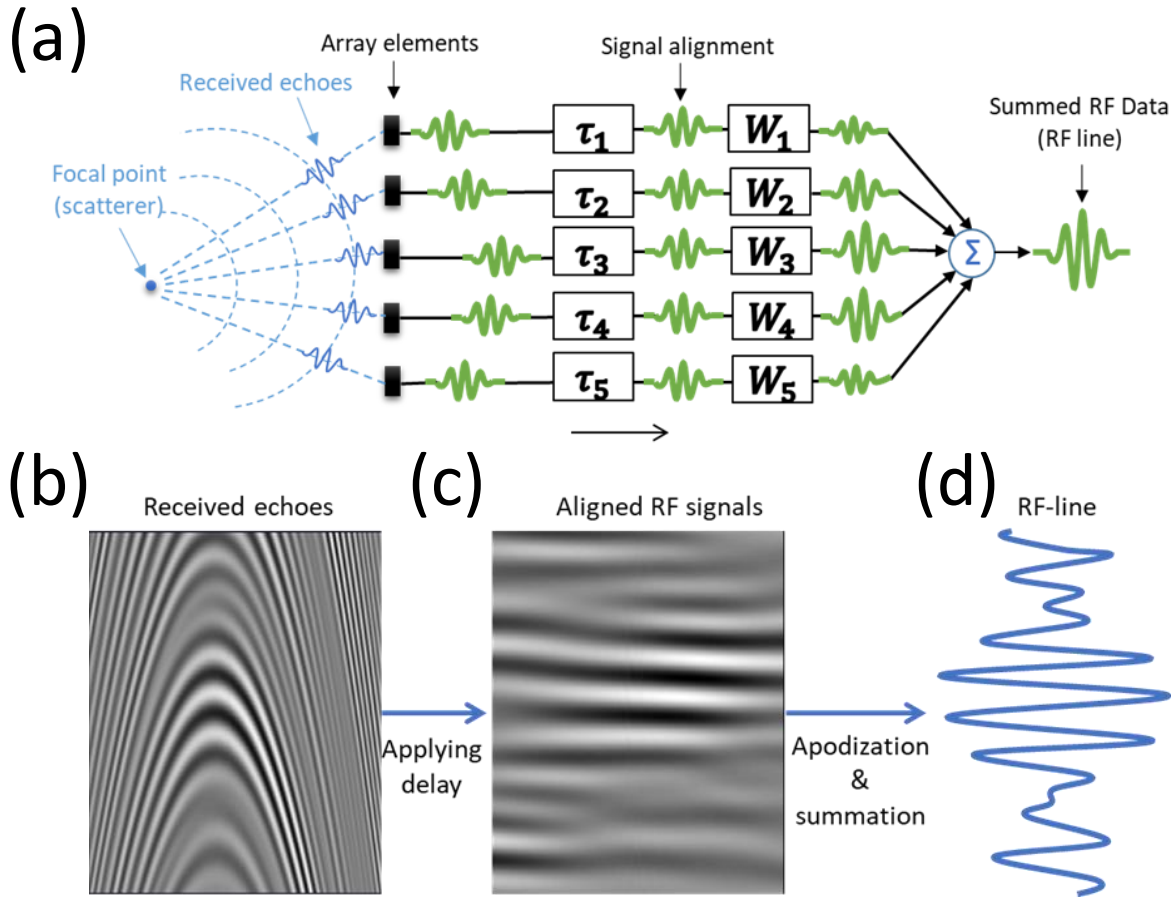

Figure 2. Beamforming illustration. (a) Beamforming principle to produce raw RF data. The received echoes are first weighted and time aligned, and then summed to produce a single RF data line. (b) Gray scale image of the received echoes detected for a single scan line. (c) Time alignment of (b). (d) The summation of all the signals forms a single RF line.

In order to increase axial resolution, The transducer transmits an excitation signal of a single cycle, and then waits for it to return before transmitting a new pulse. This reduces cross transmission interference and allows to isolate echoes from different depths optimally. Thus, the time for each transmission comes out to be a full round trip of the acoustic wave [39]:

$$t_{focus} = \frac{2z}{c} \tag{3}$$

Such that z is the depth of the focus point, and c is the speed of sound. Since all receive BF and post processing can be performed in real time in most practical cases, we can neglect the time

incurred by them. By denoting by z_{max} the maximal distance to which we are focusing, we can get the following expression for the frame rate:

$$f_R = \frac{1}{T_{frame}} = \frac{1}{t_{line} \cdot N_{focus\ points}} = \frac{1}{\frac{2 \cdot z_{max}}{c} \cdot N_{focus\ points}} \quad (4)$$

The implication of this is that the frame rate is mainly dependent on two major factors – the number of focus points used, and the depth at which we are imaging. While the depth being imaged is mainly dependent on the use case, the number of focus points present a tradeoff between image resolution and frame rate. In the acquired image, the image quality is high around the spots focused to, while the farther away we get from them the lower the quality. In order to improve the image quality in additional areas, the US beam is steered across the entire FOV, sacrificing frame rate for resolution.

Another method worth mentioning is the plane wave imaging. Instead of transmitting a focused beam, a planar wave is transmitted leading to a uniform acoustic field in the FOV. In plane wave imaging the focus is done on receive. Since no focusing is applied on transmit, the frame rate is increased significantly, at the cost of reduced image quality. Since the transmitted energy is uniform across the FOV, the receive signal level is significantly reduced. In addition, when two way focusing is used, the focus point is focused twice, essentially squaring the side lobes, yielding improved contrast. Overall, two way focusing is considered the highest quality imaging.

3.3 Multifoci transmission

An alternative to the transmission setup described above is using multifoci patterns. In this setup, BF is applied on transmission in order to create multiple focus points in a single transmission event. The wave transmitted can be controlled similarly to before, by modifying the

delays transmitted and the main issue is how to design the wavefront transmitted in order to yield the desired pattern [40].

The most direct approach to performing transmission BF, is the Multi Line Transmission (MLT) method [11]. In this method, the phases of individual foci are coherently summed in order to achieve multifoci pattern generation. This technique essentially transmits multiple pulses simultaneously in different directions, which leads to two major drawbacks. The first stems from the way the beams are focused to the different focus points. Essentially, only a section of the actual aperture is utilized for each focus point, yielding a smaller effective transducer. This leads to larger foci points, reduced resolution and reduced SNR. The second problem this method suffers from is cross interference between the different beams that are focused to different foci points. The cross interference further reduces resolution and SNR, and introduces artifacts in the received wave. It is possible to improve this method, for example by utilizing the second harmonic signal in order to compensate for the cross interference [41]. Nonetheless, these methods still don't achieve the necessary success in generating multifoci patterns.

In the field of optics, wavefront shaping has been performed and yielded significant success. Examples of usage are Structured Illumination Microscopy [42], [43], hologram generation [44], optical tweezers [45]. In these fields, methods such as GS and the optimal-rotation-angle method [14] have succeeded in designing the necessary wavefront in order to generate the requested pattern. While these methods originate from the optical field, they can also be utilized in the context of US. Examples of adaptations to US are the conjugate field method [12], the pseudo inverse method [13], and GS [14]. Of these, the GS method yields the best results when used in US[15], and as such we will focus on it going forward.

GS is an iterative algorithm for retrieving the phase of a propagating field from a pair of imaging planes related via a propagating function. In the context of US, the angular spectrum method can be utilized as the propagating function [46]. The GS method operates by repeatedly forward propagating and back propagating the designed wavefront between the two planes, until a desired similarity threshold is achieved. After each propagation, the appropriate constraints are applied, such that in the transducer plane the wavefront is set to zero outside of the transducer, and at the imaging plane the requested phase map is superimposed. Denoting the complex harmonic pressure as P(x,z), we can write:

$$P(x,z) = A(x,z) * e^{j\varphi(x,z)} \tag{4}$$

Such that A is the amplitude term, and φ denotes the phase term. P_1 denotes the pressure field at depth Δz from the transducer plane, and P_2 denotes the pressure field at the transducer plane. At the beginning, the amplitude A_1 at the imaging plane is defined and a zero phase φ_1 is imposed. The field is then backpropagated to the transducer plane P_2 using the angular spectrum method. Then, the amplitudes outside of the transducer are set to 0, and it no apodization is required the amplitudes inside the transducer are set to 1. The wavefront is then forward propagated to P_1, and the requested amplitudes are imposed while the calculated phase φ_1 is maintained. This process is repeated until the computed amplitudes A_1 are similar enough to the requested amplitudes, depending on a predefined threshold (Fig. 3a). [14] proved that the error decreases in each iteration, and in practice after a few tens of iterations the algorithm converges.

The size of the ROI at the focal depth is determined by the number and spacing of foci. A larger number of foci allows for the imaging of a larger ROI, but since the transmitted energy is divided equally among all foci the SNR can be reduced if the time-averaged energy is limited. The number of foci is also constrained by the axial length of the transmitted pulse, as temporal and

spatial interference of all transmitted cycles must occur to generate multiple foci simultaneously. In optics, this isn't an issue as all the waves are continuous, but as discussed above in echo-pulse US imaging the transmitted pulse is a single cycle wave. This poses a limitation on the possible extent of the pattern and the number of focus points used. An illustration of why this occurs is in Figure 3.b, such that Δz is defined as:

$$\Delta z = z - z_i \tag{5}$$

Where z is the distance to the imaging plane, and z_i is the distance to the farthest focus point. Temporal and spatial interference occurs when Δz is smaller than the length of the pulse, which

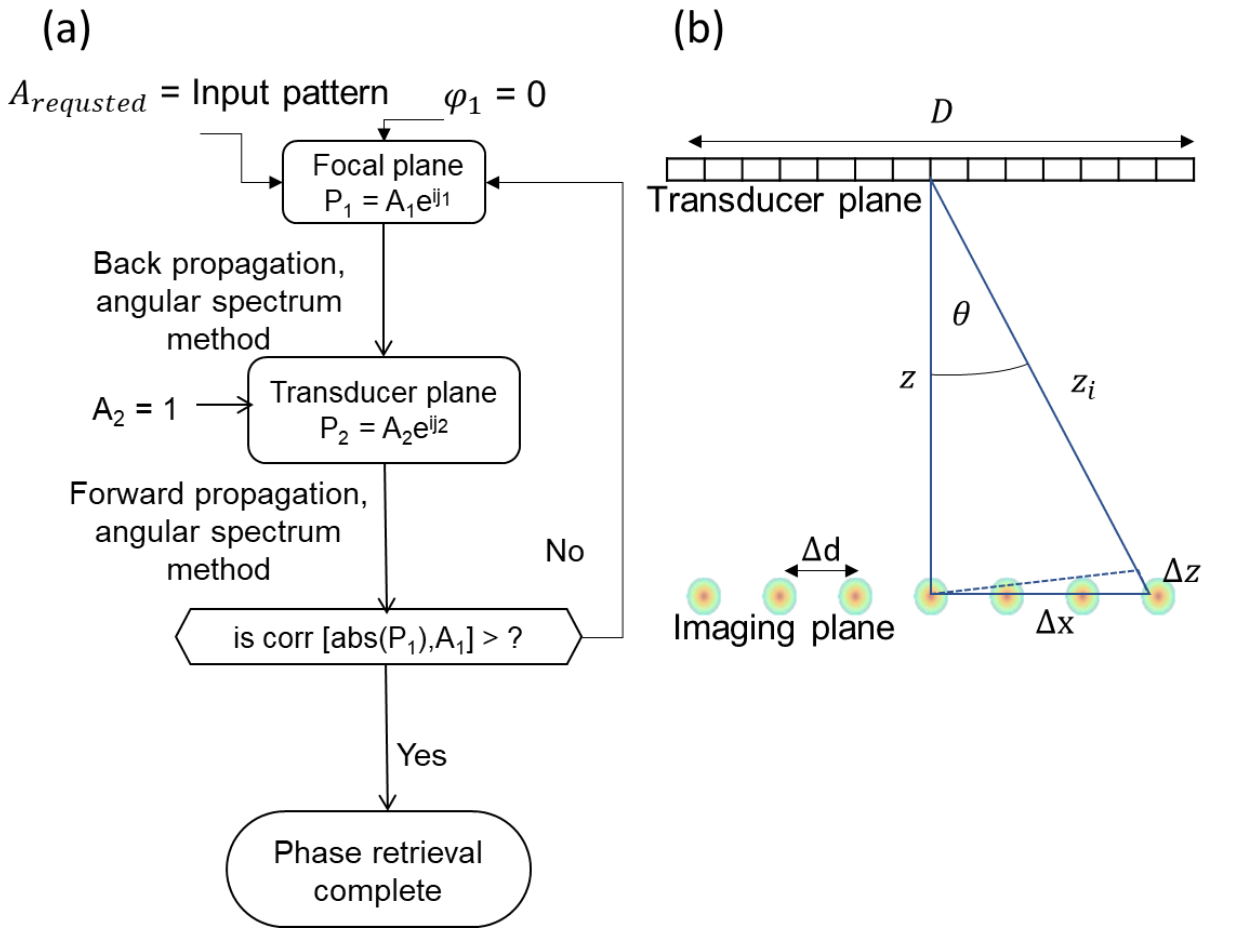

Figure 3. GS flow and constraints. (a) Flow chart of the GS process. The target input is set at the beginning as A_requested, and the initial phase is set to 0. The wave is then backpropagated to the transducer plane, and the amplitudes of the transducer are set to 1 and the rest 0. The iterative process repeats until the correlation between P1 and A1 gets under a predefined threshold. (b) Geometric model for the temporal and spatial interference requirement

for a single cycle transmission is simply the wavelength λ. From the triangle similarity between the big triangle and the little triangle, assuming a small angle approximation we can get:

$$\frac{\Delta z}{\Delta \mathrm{x}} \cong \frac{\Delta x}{z} \tag{6}$$

$$\rightarrow \Delta x \cong \sqrt{z * \Delta z} < \sqrt{z * \lambda} \tag{7}$$

Δx is approximately half the width of the pattern, meaning the maximal possible foci spacing possible is double $\sqrt{\mathrm{z} * \lambda}$. On the other hand, foci spacing is lower bounded by the diffraction limit, which dictates the minimal size possible for each focus point, according to:

$$FWHM = const * \frac{\lambda * z}{D} \tag{8}$$

Such that FWHM is the full width at half the maximum of the focus, z is the distance to the imaging plane, D is the width of the transducer and const depends on the type of aperture and imaging method used. For two way focusing, const = 0.886. In practice, this lower bound is achieved only when a single focus point is generated. When transmitting in a multifoci setup, the effective transducer size used per focus is smaller than the actual transducer used. This effect occurs due to partial usage of the elements to generate each foci point, where the actual fraction of the aperture used depends on the transmission method. For example, in the MLT method the aperture is effectively split between the different foci, yielding a significantly smaller effective aperture. GS on the other hand doesn't use the simple solution of directly splitting the transducer, allowing for smaller foci to be formed. It is important to note that the size of the generated foci isn't necessarily uniform and can vary in width and intensity.

In total, by combining (7) and (8), we reach a constraint on the possible number of foci and spacing between them. For example, for an imaging depth of z = 40 mm, a center frequency of

3.6 MHz, in water (v =1490 $\frac{m}{s}$), we get that $\lambda = 0.41\,mm$. This means $\Delta x < 4\,mm$, and the maximal possible width is $2 * \Delta x = 8\,mm$. Assuming an 128 element transducer, with a pitch of 0.218 mm (these numbers correspond to the transducer which the research was performed on) $D = 27.9$ and the minimal FWHM per foci is $FWHM = 0.886 * \frac{0.41*40}{27.9} = 0.52\,mm$. Assuming Δd, meaning the spacing between foci points, equals 1 mm, we get that we can fit it in approximately 8 foci points. If the spacing is increased to 1.5 mm, the number of possible foci points generated is reduced to 5. As described above, the actual size of the generated foci will be larger than the theoretical FWHM, meaning we need to take larger values for the actual size used. When transmitting in a single cycle mode, the problem is further exacerbated since there is no constructive interference between different cycles of transmission. This leads to a reduced ability to generate foci spots, both in terms of intensity and in terms of possible width of patterns.

3.4 Deep learning

In recent years, the field of deep learning has grown rapidly and has found success across many use cases. In many cases, deep learning has succeeded in surpassing traditional methods. There are many such examples, but a handful of them are text to speech applications [47], text to image synthesis [48], and image processing [49]. In specific fields, such as natural language processing, deep learning has greatly surpassed the previous achievements by traditional models, pushing the boundaries of what was believed possible [50]. In the context of medical imaging, deep learning has also been shown to achieve success, and examples of this include cancer segmentation [25], synthesizing CT images from MRI images[26], and super resolution [27]. For classification tasks in US, deep learning has been applied to identifying breast tumors [28], [29], liver cancer [30], and congenital heart disease [51]. Deep learning has also been used for super resolution in general

ultrasound imaging [31], and the DeepULM model has been applied to super resolution in the context of US localization microscopy. Other use cases in the context of US have been receive beamforming, detection and imaging of fetuses, suppression of clutter, and automatic segmentation of ultrasound images [52], [53]. These use cases demonstrate the versatility of deep learning in the field of medical imaging in general, and in US in particular.

Deep learning is a method which aims to learn Neural Networks (NN) which contain complex representations of data and use them to perform specific tasks. To achieve this, the NN's are constructed by assembling multiple layers of non-linear modules, such that each layer contributes to the overall representation. Once the NN is built, it is trained on large datasets until it achieves sufficient performance and can then be used in order to perform inference on new data.

The basic building block from which NN are constructed are blocks of a linear operation, followed by a nonlinear function. Assume $x \in R^d$ be the input to model, we get that the basic building block can be expressed as:

$$y = \psi(Wx + b) \tag{9}$$

Such that ψ is the nonlinear function, W is the linear operation and b is the bias. Since in the context of deep learning we stack multiple of these building blocks, we get that the output of the whole model can be described as:

$$y = \psi_n(\dots(\psi_2(\psi_1(W_1 + b_1)W_2 + b_2)W_n + b_n) \tag{10}$$

Usually, the final few layers in the NN are task dependent and are designed in order to get the output of the model in a way that fits the required task (Figure 4.a). For example, in the context of a binary classification task a common choice for the final layer is the binary cross entropy function, which yields probabilities for the two possible classes. The final layers of the model are called the head of the model, while the rest of the model is termed the backbone of the model.

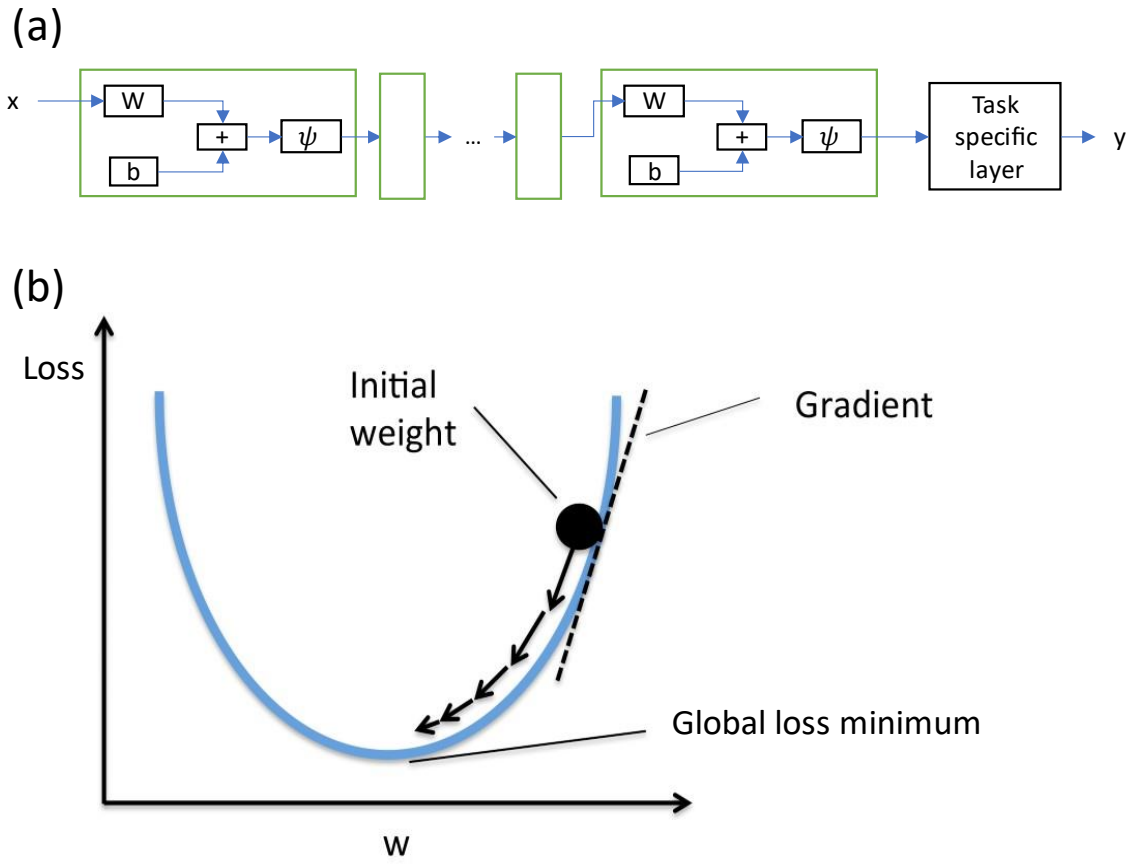

Figure 4. Illustration of a NN model and the gradient descent algorithm. (a) Illustration of NN model. The input x is passed through a series of layers which consist of some multiplication by a matrix W, adding a bias term b and applying a nonlinear function ψ. At the end of these layers there is some task specific layer to generate the output y. (b) Illustration of the learning process in Gradient Descent [78]

There are many different options for constructing the linear operations described above. The most basic example of a linear operation is the Fully Connected (FC) layer, which simply consists of a dense matrix W which is multiplied by the input to the layer. This means each element in the input has an influence to each element in the output of the layer. The other most common linear layer used is the convolutional layer, which applies one or more filters to the input x. The implication of applying a filter to the input is that each element in the output is the result of applying said filter around some point in the input. In other words, the convolutional layer preserves spatial information, in contrast to the FC layer which mixes all the elements of the previous layers. Due to this, usually convolutional layers are applied in the backbone of the NN,

while FC layers are reserved for the head of the NN where their connectivity enhances the ability to correctly compute the value for each element in the output. An important note is that many times the input is expanded to an additional dimension, such that a vector of size N turns into a matrix of size NxC, and an image of dimensions HxW turns into a tensor of size HxWxC, such that the final dimension is termed the channel dimension and we have C channels. Each filter applied in the convolutional layer corresponds to a channel in the output, allowing control of the output size.

The nonlinear functions consist of many different options, and a major group of them are the activation functions. These functions aim to "break" the linearity of the model, allowing it to learn more complex representations of the data. Common choices are the ReLU function [54], the Leaky ReLU function [55], and the tanh function. The tanh function is the classic hyperbolic function, and the ReLU function is described as:

$$y = \begin{cases} x, x > 0 \\ 0, x \leq 0 \end{cases} \tag{11}$$

Leaky ReLU is similar to ReLU, but for values under 0 we get $y = \alpha x$ for some small α. Of these, the most used functions are the ReLU and Leaky ReLU functions, as they have been found to successful in many cases.

The NN undergoes training by minimizing a task-specific loss function. Initially, the inputs are fed into the model, and subsequently, the loss function evaluates certain metrics based on the resulting outputs. If the losses are computed with respect to some ground truth data, the training process is called supervised learning, while if the loss is computed without ground truth data the learning process is termed unsupervised. The selection of an appropriate loss function plays a crucial role in the training process, and variations in the choice of loss function can greatly impact

the learning process. For example, comparing the outputs of the model to the ground truth via the mean squared error loss will motivate the model to minimize the outliers since their loss is squared, while the L1 loss will incentivize the NN to perform well across all cases while neglecting these outliers.

The actual minimization of the loss is performed mainly based on variations of gradient descent. Gradient descent is an iterative optimization algorithm which works by updating the parameters of the model based on the gradient of the loss with respect to the weights of the NN. Each time the loss is computed, the gradient is computed for each parameter in the model, which is then updated proportionally to the negative gradient, scaled by a learning rate. Let W_t^{ij} be the weight j in the layer i at iteration t, $\Delta g(W_t^{ij})$ the gradient with respect to W_t^{ij}, and α the learning rate, the update rule can defined as:

$$W_t^{ij} = W_{t-1}^{ij} - \alpha * \Delta g(W_t^{ij}) \tag{12}$$

This process brings each iteration of the model closer to convergence to some global or local minimum. This process is repeated until some predefined stopping criteria is achieved, and then the model is said to be trained. An illustration of this process can be seen in Figure 4.b. It is important to note that the model doesn't necessarily converge to the global minimum, but rather can get stuck on a local minimum.

In order to compute the gradient for each parameter, the loss is backpropagated to each layer by using the chain rule. Denote by z^l the output of the model at layer l, and the loss function as L(x, W) such that x is the input and W are the model parameters, we can write:

$$\frac{\partial L(x,W)}{\partial z^{l-1}} = \frac{\partial L(x,W)}{\partial z^{l}} * \frac{\partial z^{l}}{\partial z^{l-1}} \tag{13}$$

$$\rightarrow \frac{\partial L(x,W)}{\partial W^l} = \frac{\partial L(x,W)}{\partial z^l} * \frac{\partial z^l}{\partial W^l}$$

Thus, we can iteratively compute the gradient with respect to each of the weights in the model, by starting from the end and propagating the gradients across the entire NN.

In order to measure the performance of the NN on new data, the datasets are split into 3 different sets – training, validation, and test. The NN is trained on the training dataset, and the performance of the current iteration is measured against the validation set. After the training is completed, the NN is applied to the test dataset, in order to see how it performs on new data. Usually, the NN goes over the training dataset many times, such that each time is called an epoch, and the training is stopped once a certain number of epochs pass or the loss on the validation set is small enough.

Naturally, we expect the model to perform on the training set better than on the test set. The difference in performance is termed the generalization error, and it quantifies how the model will perform on new data that it hasn't seen yet. In order for this metric to be accurate, it is very important to have a strict separation between the test and training dataset. In some cases, the NN will overfit the training data, meaning the generalization error will be very large. This implies that the model became too specialized in learning the training data, essentially memorizing the specific examples it saw and not learning the underlying patterns.

3.5 Deep Learning Improvements

Over the years, many optimizations and changes have been proposed to the standard method of learning. Covering all of them is beyond the scope of this thesis, and as such we will only discuss the ones utilized in this research.

The first optimization we will discuss aims to solve the vanishing gradient problem. The vanishing gradient is a problem that greatly impairs the ability of models to learn, particularly when training very deep models. By observing the backpropagation described in (13), we can see that the gradients are multiplied at each layer. If the computed gradients are smaller than 1, the result of the multiplication will decrease at an exponential rate. In other words, for very deep models that gradients with respect to the weights at the beginning of the model will vanish to 0.

Residual Neural Networks (ResNet) were introduced to address the vanishing gradient problem. ResNets utilized skip connections, also known as residual connections, in order to allow the gradients to essentially skip entire layers. This is performed by adding a path to each block in the backbone, such that the output of the block equals the output of the standard layers plus the input itself. The result is a shortcut for the gradient to pass through without vanishing, allowing the early layers of the NN to learn effectively [56].

A further improvement based on this idea can be achieved via multi-level ResNets. This type of model essentially constructs a large NN model from several smaller ResNets that act on different scales, adding skip connections between the different ResNets. This allows each ResNet to operate on different scales of the model, making the NN more flexible. The idea of this method is that different properties of the input can be captured when observed at different scale. For example, for face recognition it can be valuable to observe both the general look on a large scale, and the specific eye color on a smaller scale [57].

This type of model can be combined with an encoder decoder architecture, or an autoencoder architecture. These architectures consist of an encoder section, which aims to learn an efficient representation of the input data, while the decoder aims to decode the compressed representation into the original data in the case of an autoencoder, or to some other data space in the case of an

encoder-decoder architecture. In both cases, the encoder will usually consist of a contractive path which aims to reduce the dimensionality of the input to the requested latent space, while the decoder section will expand the latent space [58].

Another important optimization relates to the way gradient descent is performed. Over the years, many different optimizations have been offered to the basic gradient descent algorithm:

1. Stochastic Gradient Descent (SGD), or batch SGD, aims to introduce randomness into the training process. Instead of using the full dataset, these methods use random batches, or even single samples from the dataset at each iteration of the training process. This adds randomness to the training process which helps avoid local minima points, and improve convergence speed. In addition, in most cases using the full dataset would add significant complexity, since while batches can be computed independently, utilizing the full dataset would demand splitting it up due to hardware constraints, creating the need for managing the gradients on each part.
2. Normalization in general, and batch normalization in particular is a method which aims to normalize the inputs in order to improve convergence. Before training, if possible, the inputs and the outputs to the model should be normalized across the full dataset. It is important to note that this isn't always possible, since if the dataset relates to physical properties the normalization can ruin the result. In addition, the same normalization must be applied to new data that is received in order to achieve good generalization results. Batch normalization is a technique that normalizes the intermediate outputs of the model, by normalizing the values with respect to the batch currently computed. This improves training by reducing the internal covariant shift across the model by maintaining consistent

input distributions to each layer. Both processes reduce the dependence on the initialization of the model and increase the stability of the training process [59].

3. Regularization, also known as weight decay, aims to prevent overfitting thus improving generalization. When the model weights are large, the NN can become heavily reliant on specific features, while being very prone to noise in those said features leading to overfitting. In order to prevent this, weight decay adds a penalty term to the loss function, by adding to it the norm of the weights of the NN, multiplied by some constant. This encourages the model to use smaller weights, which reduces the dependency on a specific part of the input.
4. Momentum is a technique that aims to avoid local minima and help navigate "flat" regions in the loss function. By adding to the gradient an accumulation of the previous gradients multiplied by a constant, local minimums can be escaped, and the model can continue learning even when the gradients are very small. In order to implement this, we can update the update rule used to:

$$v_i = \mu v_{i-1} + \alpha \Delta g$$
$$a_i = a_{i-1} - v_i \quad (14)$$

Where here we used a_i to denote some parameter at step i, v_i the velocity of the gradient with respect to a_i, μ the momentum constant, α is the learning rate, and Δg is the gradient with respect to a_i. This smooths the gradient descent, and speeds up convergence [60]

5. Learning rate scheduling is a technique which aims to improve learning by varying the learning rate throughout the training process. Using a fixed learning rate can lead to several problems, which can impair learning. If the learning rate is too small, the model can get stuck on local minima, similarly to the problem which momentum aims to solve. On the

other hand, if the learning rate is too large, the model will stay in the vicinity of some minimum but won't actually mange to reach it since the jumps in each step will be very large. Even if the learning rate is carefully chosen and can circumvent these problems, the training itself can take a long time and optimizing the learning rate itself is a tiresome process in of itself.

In order to solve this, learning rate scheduling updates the learning rate while learning according to some predefined function. This allows the model to start learning with a high learning rate, converge close to a minimum, and then "close in" on the minimum with a smaller learning rate. In the initial stage, the learning process will be quick, while the phase with the lower learning rate will allow the model to avoid jumping over minimums.

In the context of convolutional layers, there are 4 points worth mentioning.

Firstly, it is usually better to use smaller kernels for each layer, preferring to use more channels and more layers with smaller kernels for each rather than large kernels. This has been shown to perform better and have more expressivity, while demanding less parameters.

The second point is the usage of stride when computing the convolution. Stride essentially is the step size used when sliding the convolutional kernel across the input, such that instead of sliding it one element at a time the convolutional window jumps in units of stride. This allows the model to reduce the dimensionality of the input by a factor of the stride.

An additional point of importance is the usage of point-to-point convolutions. These convolutions are essentially standard convolutions, but they have a kernel size of 1. This type of convolution is useful for manipulating the channel dimension, as it doesn't change the spatial dimensions. By choosing the number of kernels applied we can choose the number of channels in the output.

The final point which we will discuss is the transposed convolutional layer, which can be thought of as the inverse of the convolution layer. This convolution operation increases the spatial dimension instead of decreasing it, allowing the model to grow and shrink the input as needed.

3.6 Proposed Method

Recently, deep learning has been employed in the context of acoustic hologram generation [61], [62]. These models receive as input the requested acoustic field and compute the necessary phases needed in order to generate the required acoustic field. While each of these models achieve good results, they both assume CW transmission, and as such they can't be transferred directly to single cycle transmission. During their training process, the outputs of the models are forward propagated to the imaging plane, allowing the loss to be computed directly in the image plane.

On the other hand, forward propagating the outputs of the model when using single cycle transmission is complex and would require essentially implementing Field II in the library used to train the model in order to allow backpropagation of the gradients. Separately, [63] encountered a similar problem when training a model for optical holography and solved it by creating a large dataset of real data and computing the losses in the transducer plane.

In our research, we aim to similarly train a NN (USDL) to solve the acoustical hologram generation problem in single cycle transmission. This will be done by generating a large dataset via Field II, such that the outputs of the simulation are the inputs for the model, while the inputs to the simulation are the ground truth for the model. Essentially, we want to train USDL to the inverse process of the forward propagation of single cycle US transmission (Figure 5) [64].

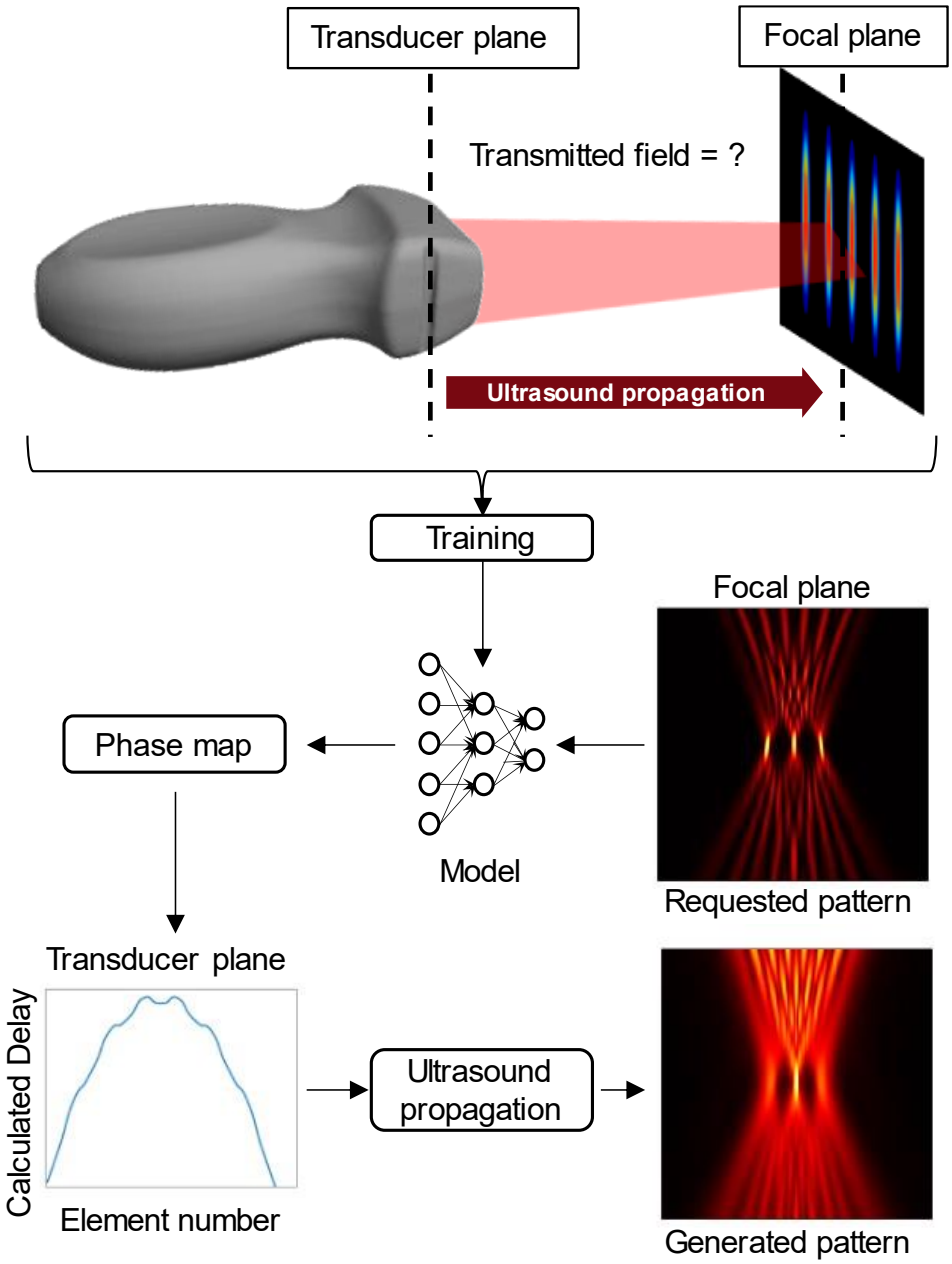

Figure 5. Illustration of the proposed method. The model computes the delays needed in order to create a pattern in the focal plane, via a deep learning neural network. Training pairs of multifoci patterns in the focal plane and their inverse propagated phase maps in the transducer plane were calculated by the GS phase retrieval algorithm. These phase maps were then propagated using a single cycle excitation. The model learns the inverse process of acoustical propagation with single cycle excitation using the training pairs.

4 Materials and methods

In this chapter, we will discuss the methods employed in this research. Firstly, the selected NN architecture will be presented. In addition, the process and considerations involved in this

selection will be explained. Next, we will outline the method devised for constructing and preparing the data set for training the model, and the training process utilized. Lastly, we will describe the approach taken to evaluate the performance of the model and conduct necessary testing. This will include discussing the simulations used for performance evaluations, and the physical experimental setup used to demonstrate the success of the model.

4.1 Design of the neural network

The expected input to the model is a vector x of size 512, which represents the requested acoustic field. The output of the model is a vector y of size 128, such that each element in it relates to the phase of an element in the transducer. The amplitudes were omitted as we found they didn't have much effect on the resulting pressure field. Both vectors are in units of the transducer pitch in order to maintain consistency.

The USDL model is loosely based on [65], and was implemented in Pytorch [66]. At the beginning of the model there are expansion blocks, which aim to increase the number of channels in the input to 512 without affecting the spatial dimension. The actual expansion is performed via point-to-point convolutions, which gradually increase the number of channels. This means that after the first expansion blocks the size of the input is 512x512. This step essentially allows the model to view different channels as having different meanings for each channel w.r.t to each element in the original vector, similarly to the meaning of the different channels in RGB images.

The main component of the model is a series of multi-scale residual nets that operate in a chain. The residual nets are split into two paths, down sampling blocks and up sampling blocks, similar to the autoencoder architecture. The down sampling blocks reduce the size of the input by a factor of half at each block. This allows for a progressive reduction in dimensionality as the data flows

through the network, meaning the model can choose what spatial information it wants to retain or compress. The dimensions of the input at the bottleneck of the model, i.e., where the spatial dimension is the smallest, is 1x512. Afterwards, the up sampling blocks are responsible for increasing the size of the input back to its original size. This is performed by mirroring the down sampling blocks, such that for every down sampling block there is an up sampling block that up samples by a factor of half. After this stage the input size returns to 512x512.

Between each down sampling block and its corresponding up sampling block there is a skip connection, which is implemented by passing the input through a residual block and then adding the result to the input of the up sampling block. This gives the model flexibility to handle information from different scales together and retain information which might have been lost from the down sampling process.

After this section, the input is fed into another expansion block, yielding a size 512x1024. However, unlike the rest of the up sampling blocks, this additional block does not have a direct residual connection to the down sampling blocks.

After this block the resulting output is fed into a series of expansions blocks, which are now used as compression blocks instead. In this case, their goal is to reduce the number of channels to 1, while maintaining the spatial dimension. After this step the output is a vector of size 1024x1. At this point we reach the regression head utilized in the model, which is simply a series of FC and ReLU layers bringing the size of the vector to 128x1, matching the required output size.

An illustration of the full model can be found in Figure 6.a.

The down sampling blocks consist of three sets of convolutional layers, and a residual connection (Figure 6.c). The first convolutional layer utilized a stride of 2 in order to decrease the size of the model by a factor of 2, while the other two convolutional layers preserve the spatial

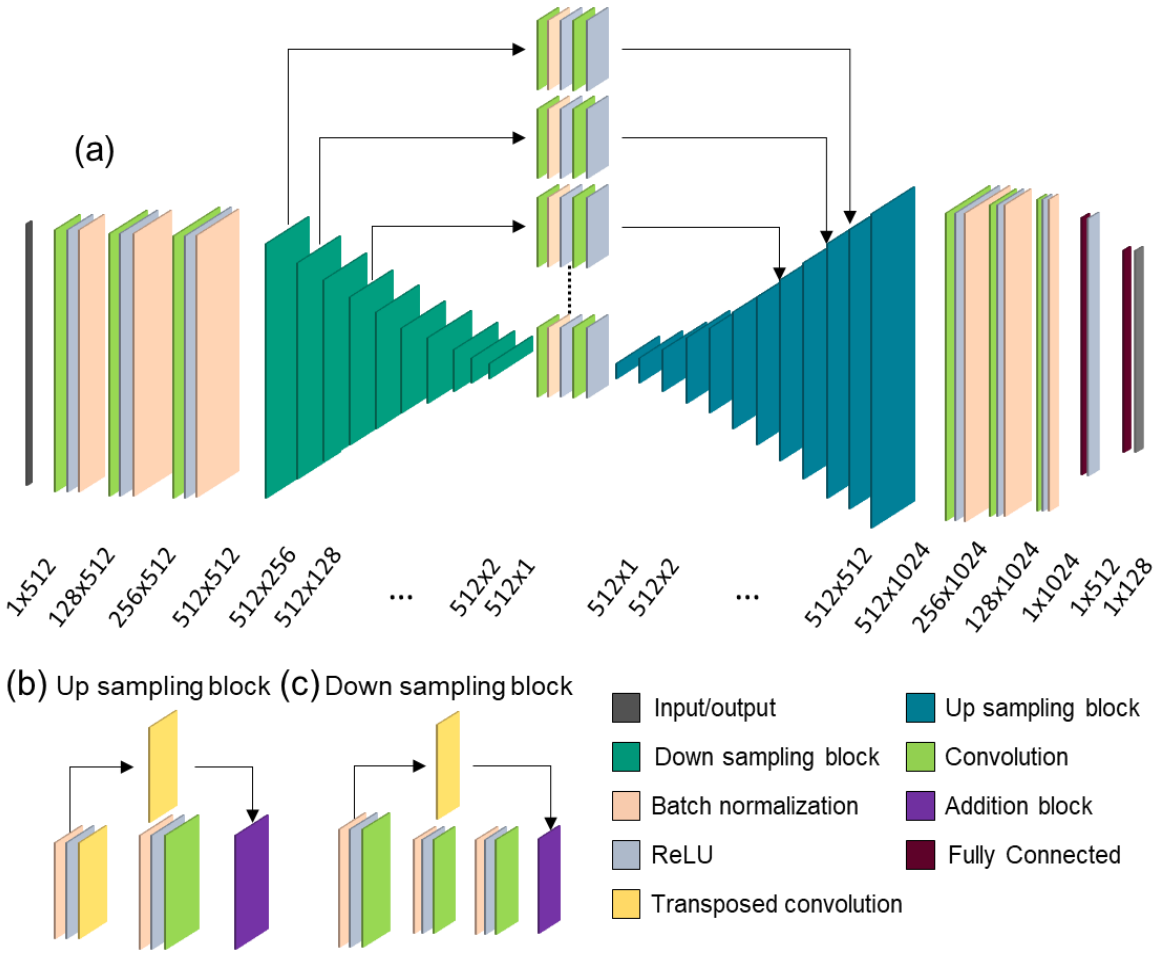

Figure 6. Network architecture. (a) structure of the whole network (b) up sampling block (c) down sampling block

dimensions of the input. Before each one of the convolutional layers, a layer of batch normalization and ReLU were added. The residual connection includes a simple convolution to maintain dimensionality with the regular computation path, while adding further flexibility to the model.

The up sampling blocks consist of one transposed convolutional layer, a regular convolutional layer, and a residual connection (Figure 6.b). The transposed convolution layer is responsible for actually up sampling the input, with a stride of 2. The regular convolution maintains spatial dimensions, and simply adds to the computational process. Similarly, to the down sampling blocks, batch normalization and ReLU have been applied before the convolutional layers, and the skip connection includes a transposed convolution in order to maintain dimensionality.

Residual blocks consist of a convolutional layer, ReLU and batch normalization, followed by another convolutional layer and ReLU. The expansion blocks simply consist of a point-to-point convolution, ReLU, and a batch normalization layer.

We note that several additional options were tested before reaching this final architecture, the major ones were:

1. Dropout – Dropout is a standard method utilized in deep learning, which aims to improve generalization performance of models. This is done by setting part of the inputs to each layer to 0 with some predefined probability. Removing parts of the input forces the model to understand the underlying patterns in the data, rather than focus on specific details in the input vector. While this is considered standard in most models and reduces overfit, in our case we found it degraded performance. The reduction in performance was consistent when moving the location of the dropout and changing the probability of the dropout, leading to the removal of it altogether. We believe the reason dropout didn't improve performance is that we employed a very large dataset with enough variance such that the model was forced to learn a good fit. Since the training dataset consisted of enough different examples, the model couldn't capture the full dataset without learning the underlying patterns.
2. Fully convolutional networks – A fully convolutional network is a type of NN architecture which does not consist of any FC layers at all, instead utilizing solely convolutional layers. The architecture was envisioned as a fully convolutional network at the beginning of the development, but the regression head was replaced with FC layers once we realized the convolutional layers were failing.
3. U-NET – an implementation based directly on [65] was also attempted but didn't reach satisfactory results.

4.2 Dataset construction and training

In order to train the model, an appropriate dataset needed to be constructed, and then a training methodology was chosen.

The dataset constructed was simulated in Field II, to get data that corresponds to a single cycle US transmission. Each example in the dataset is the result of feeding a phase map into Field II, and then creating a training pair of the resulting pressure field and corresponding phase map. $1x10^6$ random examples were generated, such that each one consists of up to 10 foci spots. Of these, 75% were assigned to the training dataset, 15% were assigned to the validation dataset, and the last 10% were assigned to the test dataset. This separation was random and was preserved throughout the training and testing process.

For the examples in the dataset to be meaningful, the phase maps used were created with GS before feeding them into Field II. For each example, a pattern with a random number of focus points was generated, with a maximum of 10 focus points per pattern. This pattern was then fed into GS to generate a phase map for the requested pattern. This step is important, as it helps the model learn data that is similar to the actual data it will see during inference, where it will be asked to generate foci points and not random amplitudes.

Field II was configured to mimic the parameters of P6-3 (Philips, Bothell, WA, USA), the transducer that was also used later in the experiments. This transducer has 128 elements, with a pitch of 0.218 mm. The transmitted waves were propagated with single cycle excitation, enveloped with a Tukey window, and a central frequency of 4.46 MHz which is the default center frequency of the P6-3 transducer. This translated to a total pulse duration of 0.22 μs and the generated acoustic waves were forward propagated to a depth of 40 mm.

In order to make the training data as similar as possible to the inference data, the following two steps were taken:

1. Normalize the acoustic wave generated in Field II, such that the maximal intensity will be 1. During inference the input to the model is a set of focus points, such that the intensity of each point is a fraction of the transmission energy. This means that the input will be with values ranging from 0 to 1.
2. Extract the foci points from the normalized pressure line. This was performed by using the find peaks function found in Scipy python library [67]. The cutoff value for the peaks was 0.5, meaning that we are defining a foci point as having half the maximal energy in the acoustical wave at depth 40 mm. Similarly to the previous step, this step aims to feed the model with focus points as close as possible to what it will see during inference.

The result is a vector with a set of what we defined as foci points, with normalized energy, which corresponds to what we expect to receive during inference time.

In addition to the two previous steps described above, one more final preprocessing step was taken in order to improve performance. The inputs generated by the previous steps are sparse vectors, which impedes the effectiveness of models relying on convolutional layers. This problem stems from the way the convolution is applied – the filter is moved across the input, and each time it sees a small area in the input. If the input in said area doesn't contain any information, the filter won't yield any information also. Deep models can alleviate this problem somewhat, since as we go deeper into the model, the input to each layer is a compounded function of a larger section of the original input. Nonetheless, in order to reduce this problem, the input to the model was convolved with the PSF of the transducer essentially spreading the input foci points across a larger section of the input. The PSF was computed for the P6-3 transducer, such that it came

out to be a sinc function with FWHM of 0.52 mm. This allowed the earlier layers to encounter more information, yielding a better learning process and better performance.

The dataset itself consists of pairs of the output from the steps described above and the phase maps that originally generated these outputs. The phase maps were considered the ground truth which the outputs of the model were compared against, while the processed pressure lines were used as the input to the model.

During the training process, Huber Loss [68] was applied. In general, the two most common loss functions utilized in the context of regression are the MSE loss and the L1 loss. Since MSE loss squares the difference between the two data points, outliers tend to be greatly emphasized during the training process, instead of focusing on optimizing for most of the inputs. On the other hand, L1 loss isn't smooth around 0. Huber loss combines these two, by using MSE loss for small values, and L1 loss for larger values according to the following definition:

$$Huber(x, y) = \begin{cases} 0.5 * (x - y)^2, & if\ |x - y| < \delta \\ \delta * (|x - y| - 0.5 * \delta) & o.w \end{cases} \quad (11)$$

Such that δ is typicallyl set to 1. This is essentially a continuous and thus differentiable combination of L1 loss and MSE loss, avoiding the issues with each one of them.

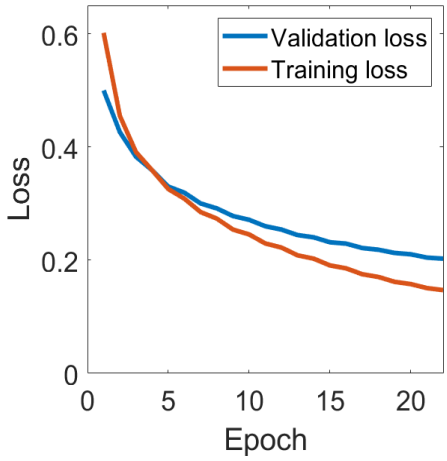

Figure 7. The USDL training vs validation loss

The optimizer used was SGD with weight decay and momentum. The hyperparameters used to train the model were found via grid search, such that the optimal values were found to be a learning rate of 0.01, momentum of 0.9, and weight decay of $1x10^{-7}$. The model weights were initialized with orthogonal weights, which were found to improve convergence [69]. During the training process, the learning rate was annealed every 2 epochs by a factor of 2.

The model was trained until a stopping criterion was met, such that if an improvement of less than 0.0001 was achieved the training was stopped. Typically, this occurred at approximately 20 epochs. A plot of the training and validation loss can be viewed in Figure 7. Since the training loss is lower than the validation loss by only a small margin, we can conclude that there wasn't overfitting. We note that there is an oddity in the graph, namely that the validation loss is smaller than the training loss at the beginning of the training process. We attribute this to the large dataset used, which means that in the first few epochs when the model isn't trained yet it doesn't perform well on the training data, but it already managed to learn enough to have partial success on the validation loss.

The USDL model and training code were implemented in PyTorch [66], while the code interacting with Field ii was implemented in MATLAB (version 2018a, MathWorks, Natick, MA, USA). All of the coding was performed on a desktop computer with an Intel core i9-11900K @ 3.50GHz with 32 GB of RAM, and an Nvidia GeForce RTX 3080 GPU.

4.3 Simulation and Ultrasound setup

The setup used across the whole research was identical and was based on the phased array P6-3 transducer (Fig. 8a). During simulations, Field II was used, such that all the transmitted waves

were propagated with single cycle excitation, enveloped with a Tukey window. For a single focus, the one-way lateral resolution was 0.57 mm and the axial resolution was 0.17 mm.

In order to validate our method, a programmable ultrasound system (Vantage 256, Verasonics Inc., Kirkland, WA, USA) was utilized (Fig. 8b). The computed phases were transmitted via the standard transmission function, such that the phases were fed into the elements of the US system as delays in time.

The acoustic wave transmitted from the P6-3 transducer were measured via a needle hydrophone, with an aperture width of 0.2 mm. The hydrophone probe was positioned perpendicularly to the emitted field inside a degassed water tank and mounted on a three-dimensional positioning system (Newport motion controller ESP 300), (Figure 9). The needle was moved across the measured plane, and for each point the pressure signals received by the hydrophone were recorded by a digital oscilloscope (MDO3024, Tektronix, OR, USA), and then post processed into a normalized pressure map [70].

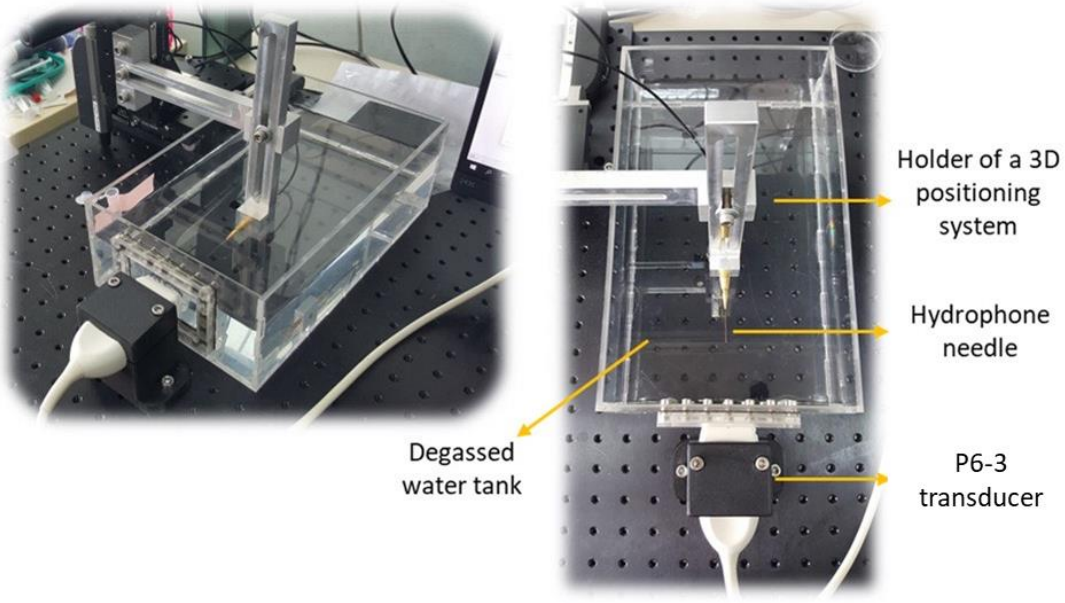

Figure 8. Hydrophone set-up used for the measurement of the acoustic fields generated by the P6-3 transducer.

Pulses were transmitted from the Verasonics system with a voltage of 10V. This translated to a peak negative pressure of 0.82 MPa, measured in-situ using the hydrophone. Since the mechanical index (MI) is defined as the square root of the peak negative pressure, we get a MI of 0.4. For multifoci transmissions, the peak negative pressure decreases proportionally to the number of generated foci. As such, the MI also decreased proportionally to the number of generated foci.

In order to evaluate the quality of the generated acoustic wavefront we defined the success rate of a transmission w.r.t to the requested pattern. A focus point was considered successful if a peak was detected within one FWHM of the requested location, with a normalized intensity of at least half the requested value. The peaks were detected via the find peaks functions in MATLAB, with a cutoff value of half the requested foci intensity. A transmission was considered successful if all the foci in it were generated successfully, and only them. If more foci points were generated, or they were shifted, the transmission was still considered a failure.

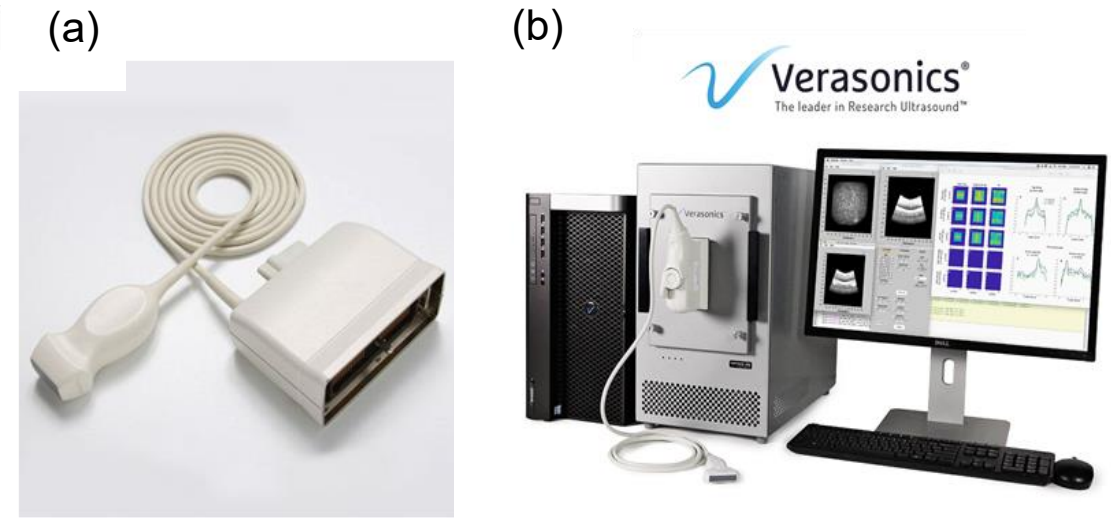

Figure 9. Hardware used in ultrasound experiments. (a) P6-3 phased array and (b) Verasonics programmable research ultrasound system.

5 Results

5.1 Pattern Generation Comparison

Firstly, the USDL network was validated by creating a striped pattern of five focal points, which resembled the patterns utilized in previous research [21]. A lateral pattern consisting of five focal points with uniform amplitudes was generated, with varying distances of 10, 15, and 20 pitches, corresponding to 2.2, 3.3 and 4.4 mm. Naturally, when transmitting GS in a CW method the best results were achieved and they were presented for comparison. For single cycle transmission, GS one cycle and the USDL method were compared (Figure 10). It is important to note that the fields displayed are normalized pressure fields, and USDL and GS had similar pressure levels. These pressure levels correspond approximately to the pressure level in a single focus point, spread across several foci.

By using the previous definition of foci success rate, we can compare the different generated acoustic fields. For the smallest foci spacing of ten pitches between each focus point, GS 1-cycle created five non-uniform foci points, while USDL succeeded in creating all the requested foci points. When the spacing was increased to 15 pitches, GS 1-cycle only managed to create 3 foci points, and when it was increased to 20 pitches GS 1-cycle only managed to generate a single foci point. In contrast, USDL succeeded in creating all foci points in all three cases.

For the case of 10 pitches separation between the different foci, USDL and GS 1-cycle reach approximately the same FWHM. It isn't possible to compare their FWHM for the case of 15 and 20 pitches, since not all foci were generated by GS. This is especially true in the case of 20 pitches of spacing between each focus point, as GS 1-cycle was successful in creating only a

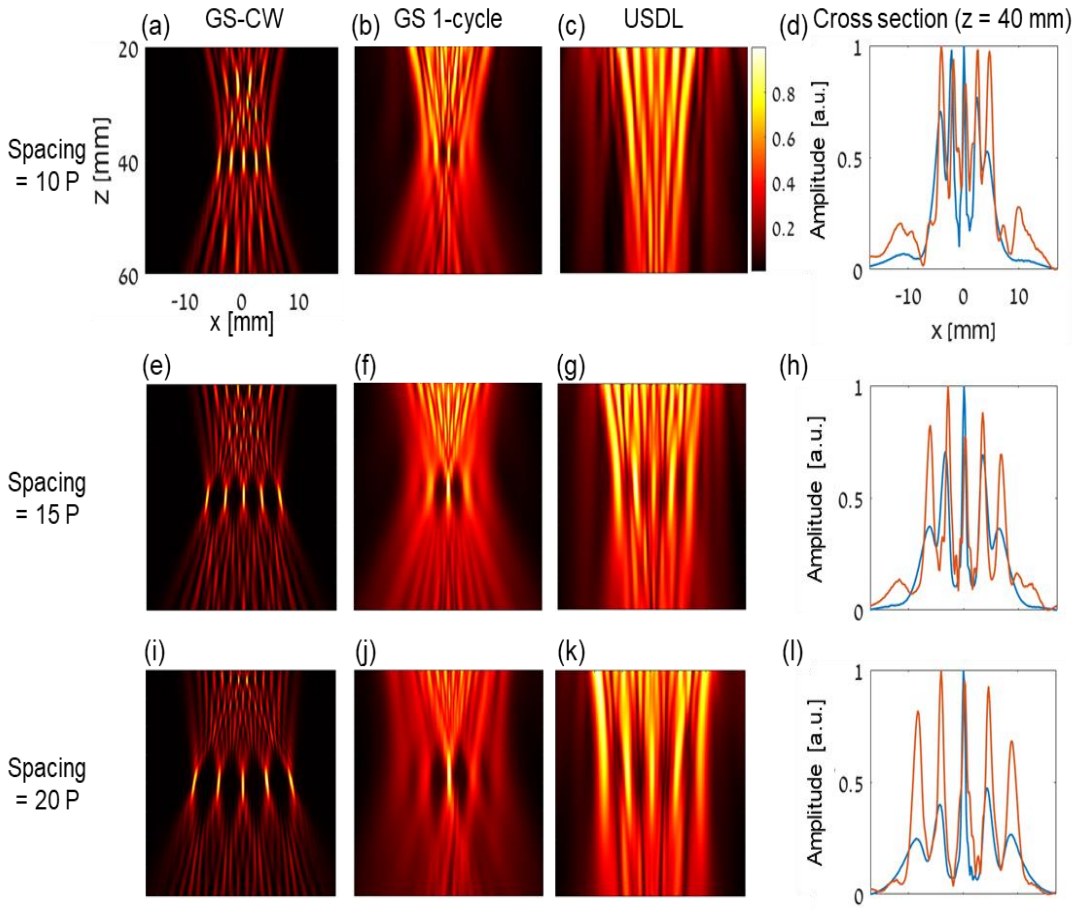

Figure 10. Effect of foci spacing in simulations. (a), (e), and (i) Ideal CW pressure fields computed using GS for patterns with 10, 15, and 20 pitch spacing between foci, respectively. (b), (f), and (j) Pressure fields in (a), (e), and (i) when transmitting using single cycle GS, respectively. (c), (g), and (k) Pressure fields obtained using the USDL model using a single cycle transmission. (d), (h), and (l) are the cross sections at a depth of z = 40 mm for the GS one-cycle (blue line) and USDL (orange line) methods. Axes are common to subfigures (a)–(c), (e)–(g), and (i)–(k) and to (d), (h), and (l). Colorbar is common to all subfigures.

single focus point, meaning we can't compare the quality of the foci themselves at all for this example.

5.2 Success Rate and MSE Comparison for Uniform Amplitude Patterns

In the next step, we aimed to compare the performance of USDL against GS 1-cycle in a robust manner. To this end, we compared the performance of USDL and GS 1-cycle in terms of success

rate and MSE on a large number of samples. As before, the comparison was performed by transmitting in a single cycle transmission mode. A successful transmission was defined as one in which all foci were created successfully when passed to Field ii for forward propagation. All requested foci had a uniform amplitude of 1, meaning the cutoff value was set to 0.5. The MATLAB function find peaks was used to detect and count the generated foci with the cutoff value. Since the intensity used is normalized, the MSE essentially compares the amount of energy lost during the transmission process, providing another measure of the performance of each method.

In order to compare the two methods, three batches of 5,000 samples each were generated. Each sample contained between one to ten focus points, with random spacing between the different samples. The maximal width of the patterns, from the right most foci to the left most foci, was set to 30 mm which roughly corresponds to the width of the transducer.

For each batch the MSE and success rate of both methods were computed, and then averaged over the three different batches. In total, 15,000 samples were evaluated. The variance for both tests were very small, lending higher credibility to the results. The results are available in figure 11.

From observing the graph of the success rate as a function of the number of foci (Fig. 11a), we can see that USDL outperforms GS 1-cycle significantly. While for one or two focus points GS 1-cycle manages to maintain performance, once three or more foci are requested GS suffers from a significant performance drop. The largest improvement in terms of success rate with USDL is achieved for 4 foci, where USDL achieves an approximately 2.5 times better success rate. We note that even though USDL manages to maintain a higher success rate, once more than six foci are requested both methods yield a success rate of less than 20% rendering both of them unusable

in practice. In terms of MSE, USDL surpasses GS 1-cycle consistently, although only by a small margin (Figure 11b).

In addition to the quantitative comparison, an example from the samples used was presented (Figure 11c-f). In this example, 3 foci with approximately equal spacing of 5 mm were requested. As we can see USDL succeeded in creating a homogenous pattern, while the foci generated by

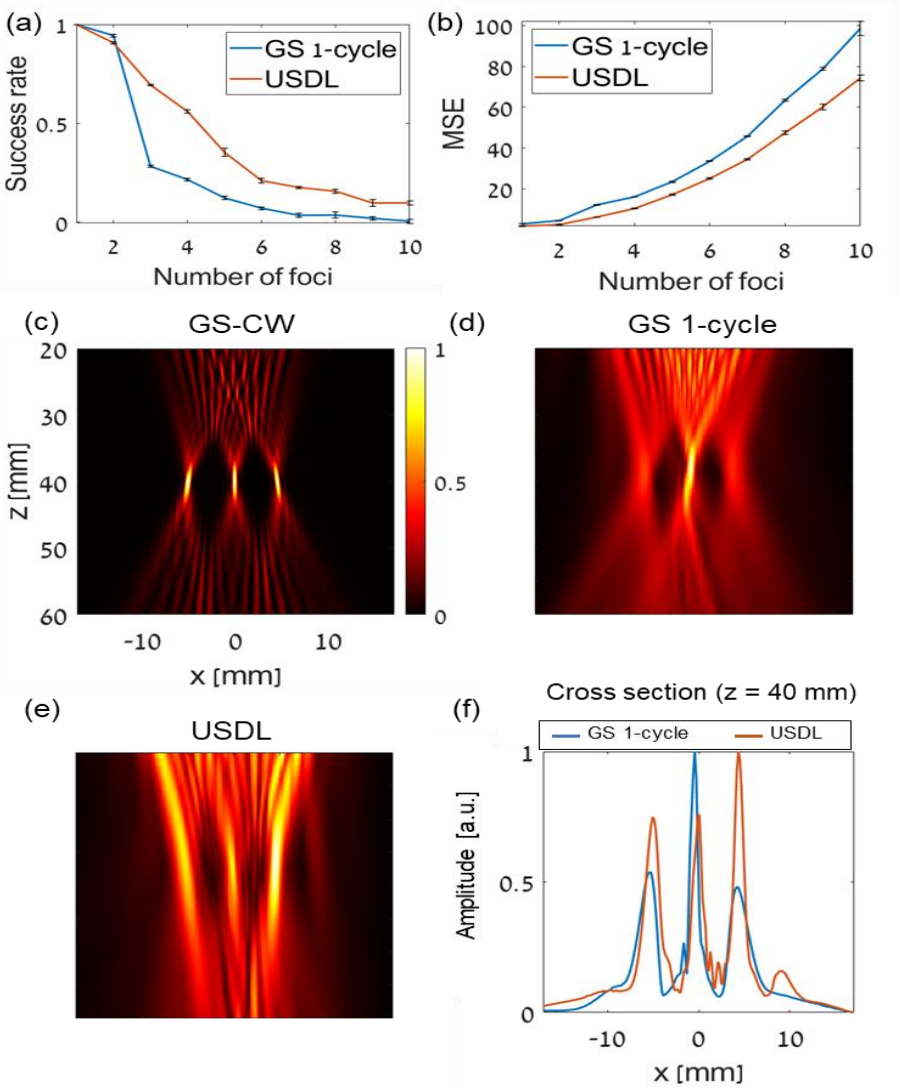

Figure 11. Constant pressure multifoci pattern generation. (a) Success rate for constant pressure multifoci patterns. (b) MSE of the resulting pressure fields. (c)–(f) Example of a multifoci pattern generation with three focal spots that were evenly spaced. (c) Pressure field generated using GS algorithm via CW insonation. (d) GS algorithm assuming single cycle insonation. (e) USDL method assuming single cycle insonation. (f) Cross sections at a depth of z = 40 mm for the GS one cycle (blue line) and USDL (orange line). Axes and colorbar are common to subfigures (c)–(e)

GS 1-cycle aren't uniform and it only succeeded in creating two out of the three requested foci. As can be expected, GS-CW surpassed both methods significantly, as it isn't limited by temporal constraints.

5.3 Success Rate and MSE Comparison for Varying Amplitude Patterns

In order to further challenge our model, we tested in a similar manner the performance of the two different methods on inputs with varying pressure amplitudes. In this case, the generated patterns were generated with a random amplitude between 0.5 and 1. Similarly to the previous comparison, 3 batches of 5000 samples were generated and compared against. In contrast to the previous case, a successful foci was defined here as half the original requested intensity, and not just a value of 0.5. The results are displayed in figure 12.

Similar results to the constant pressure case were observed in the varying pressure case. USDL manages to outperform GS 1-cycle both in terms of success rate and in terms of MSE (Figure 12a-b). It is important to note that the gains USDL yields are less pronounced in terms of success rate in this case.

Similarly, to the constant pressure case, an example from the comparison is also given. In this case, a pattern with six foci of unequal spacing was chosen. The requested foci amplitudes were, from left to right, 0.58, 0.7, 0.88, 1, 0.76, and 0.64. The GS 1-cycle managed to create only four foci, while USDL succeeded in creating all six foci. Furthermore, In contrast to the example given above where GS managed to create peaks which simply didn't reach the detection threshold but were in the correct spot, in this case GS didn't even manage to create peaks in all the spots. Instead, GS combined adjacent focus points into "bumps" that try to catch both peaks at once.

An interesting point is the drop in success rate from which GS 1-cycle suffers in the case of two foci. As can be seen in the Figure 12a, GS suffers from a sharp drop to approximately 40% success rate at 2 focus points, and goes back up to 60% for 3 focus points. While GS suffers from this problem, it seems USDL doesn't encounter such a change. From our observations, it seems that the drop in performance at 2 focus points stems from GS converging to the trivial solution of splitting the transducer into two sub apertures and using the geometrical solution for each one. This effect was observed mainly when one focus point is requested to be significantly weaker

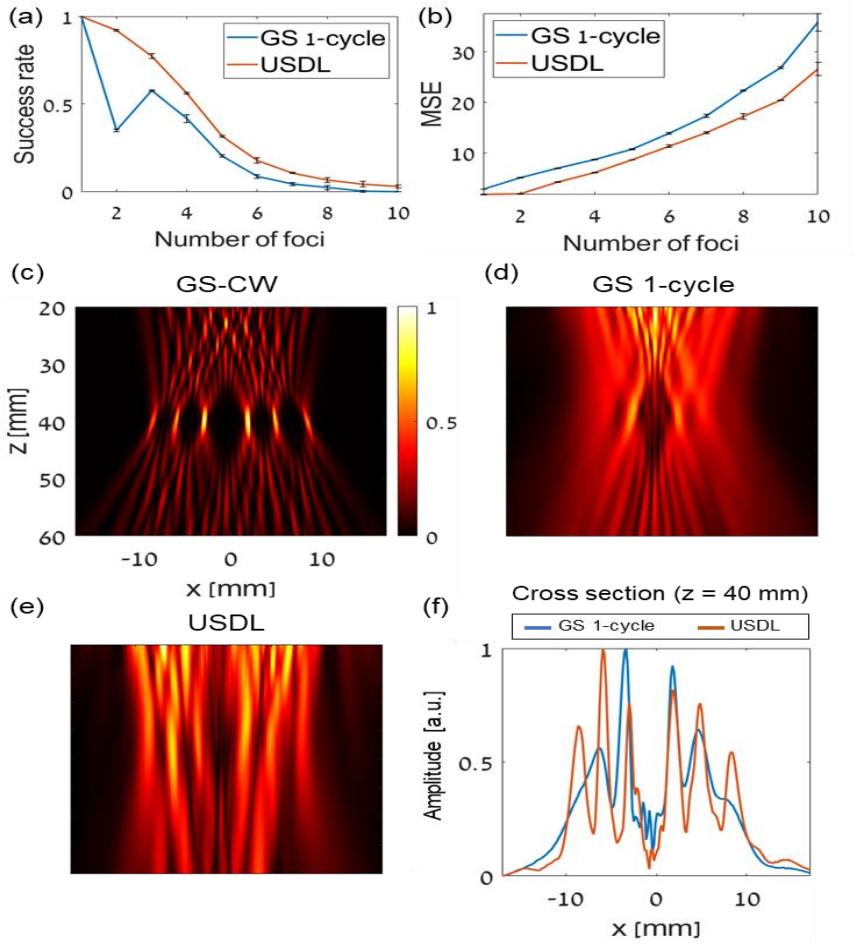

Figure 12. Varying pressure multifoci pattern generation. (a) Success rate for varying pressure multifoci patterns. (b) MSE of the resulting pressure fields. (c)–(f) Example of a multifoci pattern generation with six focal spots that were evenly spaced. (c) Pressure field generated using GS algorithm via CW insonation. (d) GS algorithm assuming single cycle insonation. (e) USDL method assuming single cycle insonation. (f) Cross sections at a depth of z = 40 mm for the GS one cycle blue line) and USDL (orange line). Axes and colorbar are common to subfigures (c)–(e).

than the other one. In this case, GS essentially allocates a small portion of the aperture for the weak focus point, and the rest for the strong focus point. Since anyways we get weaker focus points due to single cycle transmission, if GS directly tries to weaken a focus point it becomes very weak, and goes under the detection threshold. For the case of more than two foci, this effect wasn't observed, and it seems that GS can't utilize this method for them.

Like the constant pressure case, USDL demonstrated superior performance when compared to GS both in terms of success rate and in MSE. When compared to themselves in the constant pressure case, both models underperform in terms of MSE. On the other hand, in terms of success rate USDL manages to maintain its performance, while GS suffers from a significant reduction in performance. We attributed this reduction to GS generating weaker focus points in general when transmitting in single cycle mode due to the lack of constructive interference. Since the foci points are weaker, when the input pressure requested was lower in advance, the resulting focal point strength is further diminished, causing these weaker focal points to fall below the detection threshold. As a consequence, the GS model's success rate suffered from a significant reduction in performance, while USDL which didn't suffer from this issue maintained its success rate.

5.4 Success rate and MSE comparison for varying number of cycles and axial distance

Two further tests were conducted in order to evaluate the performance of the USDL model. The first parameter tested was the affect of changing the number of transmitted cycles. This comparison is important because in many practical cases a comprise between a one cycle transmission and a continuous cycle transmission is optimal, leading to the transmission of a

small number of cycles in each transmission. The second parameter which was tested is the effect of changing the axial depth to which we are transmitting. This parameter is very important for the USDL model, since if the model can maintain performance across several axial depths, the number of models needed to be trained in advance is reduced. If the model is viable only for a single depth, we will have to train a model for each depth we want to image, making the USDL solution significantly less applicable.

In order to test both parameters, similar tests to the previous cases were applied. As before, 3 batches of 5000 samples were evaluated, such that the MSE and success rate were computed as before. After computing the success rate and MSE, the results were averaged for each of the cases displayed. In total, each one of the samples in the dataset was evaluated for each configuration, either the number of cycles transmitted or the axial depth. We note that the GS was always computed to a depth of 40 mm, even if it was transmitted afterwards to a different depth. This was done in order to maintain consistency w.r.t USDL, which was trained only to 40 mm. The results are available in figure 13.

As can be expected, the more cycles transmitted the closer we get to a continuous wave transmission, meaning the performance of GS improves (Fig. 13a-b). While for the case of one cycle transmission USDL improved on GS by a factor of 3.2 in terms of success rate, the more cycles transmitted the less pronounced the improvement was. For 2 cycles, the improvement is still 2.5 fold, for 6 cycles the improvement is reduced to 17%, and for 8 cycles GS and USDL are equivalent. An interesting point is that USDL reached a saturation in it's performance in terms of success rate at around 4 cycles, while GS passed it after 8 cycles were transmitted. It is important to note that the MSE is still better for USDL, but this can be attributed to the MSE of both of the methods being very low for a high number of cycles. When comparing the affect of

changing the axial depth, we see that USDL is significantly superior to GS when the imaging depth is changed (Fig. 13c-d). This increase in performance was maintained across all depths, both in terms of MSE and in terms of success rate. Due to the depth of field, the model succeeds to generate the requested patterns in an area around 40 mm, even though it was trained specifically for 40 mm. This effective range was calculated as the FWHM and was found to be ± 3 mm for GS, and ± 5 mm for USDL. Thus, even though USDL was trained for 40 mm, it can be effectively applied to the range of 35 mm to 45 mm. Notably, while USDL manages to

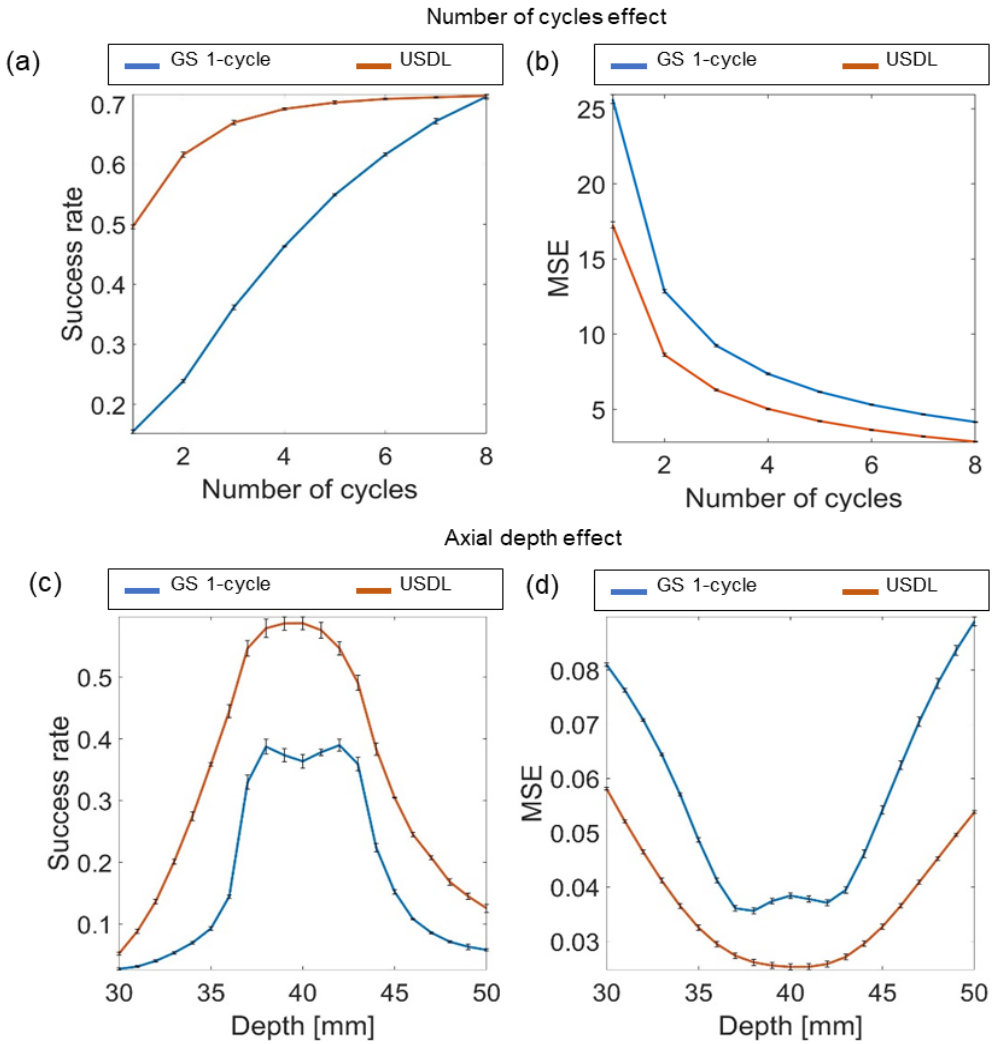

Figure 13. Number of cycles and axial depth effect. (a) Success rate as a function of the number of cycles. (b) MSE as a function of the number of cycles. (c) Success rate as a function of the axial depth. (d) MSE as a function of axial depth. GS 1-cycle (blue line) and USDL (orange line) for all subfigures.

maintain performance for this range of imaging depths, once the imaging depth exceeds this range the performance drops significantly.

5.5 Acoustic pressure mapping experiments

Up until now, all the evaluation of the model has been conducted via numerical simulations. In order to further test the validity of the model, measurements of the emitted fields were performed with a needle hydrophone (Figure 14). To this end, patterns consisting of 3, 4 and 6 foci with constant pressure were tested. The pattern of 3 foci contained focus points located at -5 mm, 4 mm and 6 mm w.r.t to the center of the transducer. The pattern of 4 foci consisted of 4 foci with a uniform spacing of 1 mm between them. These two patterns were chosen in order to demonstrate a pattern with uniform spacing and a pattern with non-uniform spacing. An additional pattern with 6 foci was generated, with foci spacing of ~1.5 mm. This pattern aimed to demonstrate the viability of using many foci in a single pattern, even if the pattern isn't optimally generated.

In the case of 3 non-uniform foci, GS fails completely at generating 2 of the foci and focuses on creating a single strong focus point. USDL also creates the same focus, but manages to create something resembling a focus point, although not one that would pass the detection threshold. In the case of 4 uniform foci, again GS manages to create only a single focus points, while in this case USDL manages to create all 4 requested focus points. For 6 foci generation, GS was able to successfully create only the two central foci, while USDL was able to create the 4 central foci. As can be seen from the result, USDL manages to create better foci also in real life measurements when compared to GS. In addition, even when USDL doesn't succeed in creating all the requested foci, it still yields a better solution than GS. In order words, even though the pattern would be

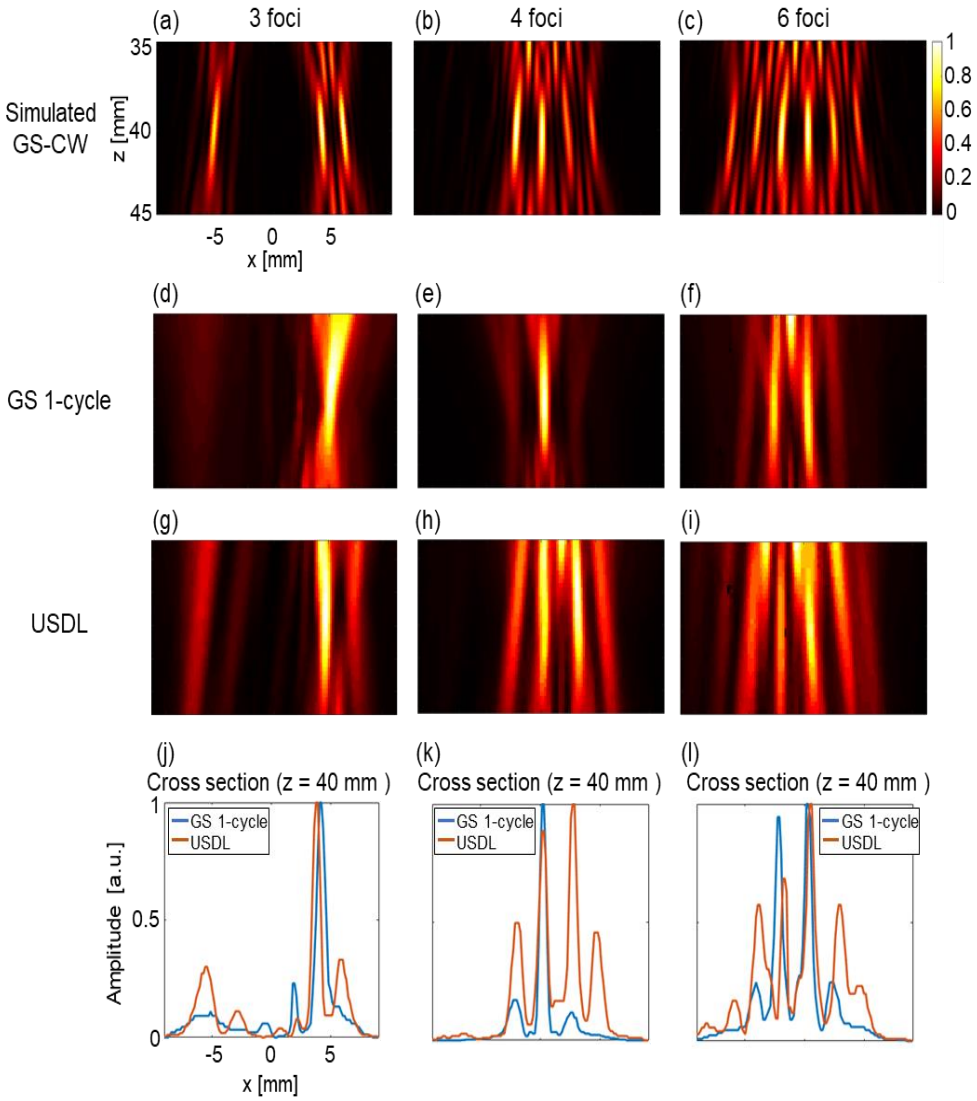

Figure 14. Experimental validation of the generated patterns via hydrophone measurement of the emitted pressure fields. 3,4 and 6 foci patterns were generated at a depth of 40 mm using (a)-(c) Simulated GS CW algorithm; (d)-(f) GS algorithm using single cycle transmission; (g)-(i) USDL method. (j)-(l) Cross sections at a depth of z = 40 mm for the GS 1-cycle (blue line) and USDL (orange line). Axes and colorbar are common to subfigures (a)-(i), and (j)-(l). All pressure fields are normalized.

marked as a failure for both of them, the failure from GS can be called worse when compared to the failure from USDL.

6 Discussion

Acoustical beam shaping is a growing field with many applications. In the field of ultrasound therapy there are several examples including but not limited to generation of acoustic holograms [20], creation of multifoci patterns for thermal ablation [12], [16], and ultrasonic neural stimulation. [15]. Nevertheless, the algorithms employed for manipulating the acoustic field operate under the assumption of CW insonation and neglect the temporal aspects of acoustic wave propagation. To achieve optimal outcomes in imaging applications, an effective beam shaping method should consider the temporal interference resulting from the emitted pulses.

This work introduced a novel approach for manipulating the transmitted acoustical field by utilizing a multi-level residual deep convolutional network termed USDL. The network was trained to learn the propagation function between the transducer and focal planes when transmitting in a single cycle transmission mode. The study primarily focuses on the generation of multifoci patterns, including patterns with uniform and varying intensities, different foci spacing, symmetric and asymmetric patterns, and patterns with varying number of focal spots. Potential use cases for this method are acoustical structured illumination [21], obstacles bypassing [71], point spread function engineering for side lobe reduction [72] extended depth of field imaging [73], and increased imaging frame rate in proportion to the number of generated foci. These applications have already been demonstrated, and should only benefit from using USDL over GS.

Future applications can include practically any method in optics that relies on elaborate spatial patterns which wasn't viable until now. For example, Barker based patterns can facilitate super resolution imaging [74] which may now be feasible on USDL. In addition, with the current progress in the fabrication of matrix arrays that are compatible with the programmable ultrasound system, the method can be scaled to 3D volumetric beam shaping.

Due to GS not taking into consideration the time domain when computing the required phases, wide patterns with a large number of foci cannot be generated reliably. This occurs due to GS relying on constructive interference between different cycles which cannot occur when transmitting in a single cycle transmission mode. As such, distant foci aren't generated or are weaker than the central lobe. In other words, patterns containing foci that aren't centered around a single area are challenging for GS, and many times it will fail at generating the requested pattern. In contrast, USDL was trained on patterns transmitted with single cycle excitation and therefore has effectively learned the time domain in addition to the spatial domain. According to Eq (7), (8) that dictates the maximal pattern width that can be generated with GS 1-cycle, for z = 40 mm, and f = 4.46 MHz, Δx = 3.65 mm. If we choose a pattern with 5 foci, the maximal foci spacing is expected to be 1.46 mm. If we choose a larger foci spacing we will exceed the maximal possible pattern width, yielding suboptimal pattern generation. On the other hand, USDL was successful in creating patterns with large foci spacing (Figure 10), uniform pressure foci (Figure 11), foci with varying amplitudes (Figure 12), and varying number of cycles transmitted and imaging depth (Figure 13).

It is important to note that due to the Talbot effect [75], there were additional generated foci patterns at different axial depths. This effect occurs in periodic diffraction grating generation and is widely known in optical holography. It may be possible to take this effect into account and design an axial multifoci pattern approach for acoustical holography.

A significant advantage included in the training method of USDL is that all the training was performed on simulated data. This means that training a new model doesn’t require extensive experimental data collection, but rather running an automated process. Moreover, the time to compute an example took on average 0.1 milliseconds, meaning the training time itself should

also be relatively small. If needed, this process can be further speed up using methods such as transfer learning [76] or knowledge distillation [77] in order to further reduce the network size. Finally, since the patterns used during the training process are random, the model is flexible and generalizes well. An example of this can be seen in the results in how USDL managed to handle nonstandard patterns, such as nonuniform distances or varying input intensities.

The model effectiveness was evaluated by creating 3 batches of 5,000 random patterns, with up to 10 foci in each one. The generated patterns were then fed into USDL and GS, forward propagated to the requested depth and compared to the original requested pattern in order to compute the success rate and MSE. Due to the usage of normalized amplitudes, the success rate essentially measures the uniformity of the generated pattern, since the amplitude is computed w.r.t to the strongest amplitude. On the other hand, the MSE acts as a measure of how much energy was lost outside of the requested foci. It is important to note that the pressure levels between USDL and GS were similar, thus the comparison of the MSE is valid. In total, we get that the combination of the two methods should act as a measure of the similarity between the generated pattern and the requested pattern.

We found that when GS 1-cycle and USDL both succeeded in creating the requested pattern, the foci amplitudes were more or less equal. However, in many of the cases GS wasn't able to create all of the requested foci, succeeding in creating only some subset of the total requested foci. In turn, the foci that were actually generated had a significantly higher amplitude while the distal foci were weak, such that the total energy was preserved. The smaller MSE values which the USDL method obtained suggest that the foci generated by USDL are both tighter, and that more energy is located in the desired region and actually used for the pattern generation. An implication of this result is that the dynamic range of the USDL patterns also increases. In terms

of success rate, the greatest improvement was achieved for 4 foci generation where GS achieved a success rate of 25%, while USDL achieved a success rate of 60%. However, a significant improvement was achieved for all numbers of foci.

The model was further tested by evaluating the effect of changing different parameters in the data on which the model is evaluated. The different use cases tested were all based on configurations which have been demonstrated previously. Changing the foci intensity is common in High Intensity Focused US, in order to allow finer control over the intensity in different areas of the treatment. Varying the number of cycles transmitted has also been in use in therapeutic ultrasound, in order to have further control of the on the treatment. The final parameter that was tested was changing the imaging depth, which is quite common since different organs require different imaging depths. As such, testing the performance of USDL across several depths is critical for understanding its applicability. Across all these different cases the model displayed superior performance when compared to GS, demonstrating better generalization and applicability.

We note that a significant caveat should be noted about the imaging depth test. While USDL outperformed GS under similar conditions, GS can be adapted to different depths very easily, simply by changing the depth to which the acoustic wavefront is propagated. This would allow GS to optimize itself for the specific depth required, while USDL is preconfigured to a specific depth during training. Thus, while USDL can be used for ± 5 mm of the imaging depth used in training, GS can effectively be used also for other depths by changing the depth of the propagation.

Additional limitations of the method should also be noted. While USDL indeed improved the number of foci that were created, the improvement still has limitations. For example, in the case

of 6 foci the USDL success rate was less than 20% which isn't practical in most use cases. Nonetheless, even in the cases where USDL didn't manage to create the required pattern, it may still be a viable option since the generated pattern many times are close to the required, as seen in Figure 14.

Another drawback is that USDL was trained for the specific transducer that was used in the experiments. The P6-3 phased array transducer, which was used during the data generation stage, was chosen to maximize element directivity and facilitate pattern generation. In order to use the model in a different setting, a new model needs to be trained for the new parameters. Several models can be trained in advance, and then chosen during the usage of the US system according to the required task. For example, in order to allow usage in an imaging depth of between 30 to 70 mm, a set of four models should be trained to depths of 35 mm, 45 mm, 55, and 65 mm in order to cover the full range required.

The process of training a new model is fairly straight forward, since all the datasets are generated via simulation and thus we have complete control over all of the parameters. These include the distance to the focal plane, the patterns that are generated (number of foci, amplitude and spacing of each foci, and the location of the pattern), center frequency, transducer size and any other parameter which we wish to change. In order to train a new model, the appropriate parameters need to be chosen, and then the data generation process should be performed again. Once the dataset is generated appropriately, the model should be trained on the new data until it reaches sufficient performance. We note that it may be necessary to modify the training hyperparameters in order to reach optimal performance. By utilizing transfer learning, it should be fairly easy to speed up this training process, although this was outside of the scope of this work.

An additional problem from which USDL suffers from is the computation burden it requires. Since it is required to work in real time, some US systems may not have the necessary hardware in order to use it. Thus, USDL needs to be benchmarked in realistic settings, and not only on a GPU accelerated computer. It is important to note that many modern US machines have been receiving more and more processing power, including optimization for deep learning models, in order to enable new modulation and imaging schemes. If the performance of USDL isn't sufficient, a lookup table of patterns and phases can be precomputed and used in real time. This option is viable since most US schemes repeat the same transmissions, meaning the transmission which will be used can be precomputed when the transmission is configured, and then simply looked up during transmission.

Two further caveats should be noted. The first, is that the USDL method was tested only during the transmission phase, and wasn't incorporated in a full pulse-echo imaging scheme. While this is sufficient for the therapeutic setting, in order to fully test the method, an appropriate beamforming method that considers the multifoci pattern should be implemented on receive. Once this is done, the actual affect of using USDL on imaging can be evaluated, both in terms of FPS and in terms of image quality.

The second relevant drawback is the differences in speed of sound when propagating through the medium. Similarly to standard ultrasound imaging, during the training of USDL the speed of sound is assumed to be constant and known. This affects both the pattern generation in GS and the US propagation in Field II. For most US applications, this speed is assumed to be ~1540 m/s. This assumption is approximately true, but in practice the speed may vary slightly. In addition, if there is an object in the field of view, such as a bone, the generated pattern will be distorted. In order to compensate for this, USDL can be combined with aberration correction techniques,

similarly to traditional US methods. Alternatively, the model could be trained on a dataset that takes into account the potential change in speed of sound into account and learn to optimize the transmitted phases to compensate for that. A reasonable way to accomplish this would be to transmit a plane wave, and adding the result to the inputted data. This would allow the model to get an impression of how the medium responds to transmissions, while maintaining a low computational and time cost.

7 Conclusion

In conclusion, we have developed a deep learning based approach to beam shaping the acoustical field in order to generate a variety of multifoci patterns. While deep learning has been used previously in ultrasound imaging, it has been applied mainly on receive for beamforming or postprocessing. In this work we have successfully incorporated these methods into the transmission stage for acoustical wave front manipulation in diagnostic applications. This method is uniquely designed for acoustical hologram generation with a single cycle transmission and can be used for ultrasound imaging applications.

8 References

[1] C. J. Harvey, J. Pilcher, J. Richenberg, U. Patel, and F. Frauscher, “Applications of transrectal ultrasound in prostate cancer,” *British Journal of Radiology*, vol. 85, no. SPEC. ISSUE 1, Nov. 2012, doi: 10.1259/bjr/56357549.

[2] A. T. Stavros, D. Thickman, C. L. Rapp, M. A. Dennis, S. H. Parker, and G. A. Sisney, “Solid breast nodules: use of sonography to distinguish between benign and malignant lesions.,” *Radiology*, vol. 196, no. 1, pp. 123–134, 1995.

[3] Jayne Cleve and Marti L. McCulloch, “Conducting a Cardiac Ultrasound Examination,” in *Echocardiography*, J. Nihoyannopoulos Petros and Kisslo, Ed., Cham: Springer International Publishing, 2018, pp. 33–42.

[4] D. L. Miller *et al.*, “Overview of Therapeutic Ultrasound Applications and Safety Considerations,” *Journal of Ultrasound in Medicine*, vol. 31, no. 4, pp. 623–634, 2012.

[5] O. H. Baltarowich *et al.*, “National ultrasound curriculum for medical students,” *Ultrasound Q*, vol. 30, no. 1, pp. 13–19, 2014, doi: 10.1097/RUQ.0000000000000066.

[6] K. K. Shung, “Diagnostic ultrasound: Past, present, and future,” *J Med Biol Eng*, vol. 31, no. 6, pp. 371–374, 2011, doi: 10.5405/JMBE.871.

[7] T. L. Szabo, “Diagnostic Ultrasound Imaging: Inside Out: Second Edition,” *Diagnostic Ultrasound Imaging: Inside Out: Second Edition*, pp. 1–549, 2004, doi: 10.1016/C2011-0-07261-7.

[8] R. Abiteboul and T. Ilovitsh, “Optimized Simultaneous Axial Multifocal Imaging via Frequency Multiplexed Focusing,” *IEEE Trans Ultrason Ferroelectr Freq Control*, vol. 69, no. 10, pp. 2930–2942, Aug. 2022.

[9] 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 Trans Ultrason Ferroelectr Freq Control*, vol. 66, no. 2, pp. 273–284, Feb. 2019.

[10] R. Mallart and M. Fink, “Improved imaging rate through simultaneous transmission of several ultrasound beams,” in *New Developments in Ultrasonic Transducers and Transducer Systems*, F. L. Lizzi, Ed., SPIE, Nov. 1992, pp. 120–130. doi: 10.1117/12.130591.

[11] G. Matrone, A. Ramalli, A. S. Savoia, P. Tortoli, and G. Magenes, “High frame-rate, high resolution ultrasound imaging with multi-line transmission and filtered-delay multiply and sum

beamforming," *IEEE Trans Med Imaging*, vol. 36, no. 2, pp. 478–486, Feb. 2017, doi: 10.1109/TMI.2016.2615069.

[12] M. S. Ibbini and C. A. Cain, "A Field Conjugation Method for Direct Synthesis of Hyperthermia Phased-Array Heating Patterns," *IEEE Trans Ultrason Ferroelectr Freq Control*, vol. 36, no. 1, pp. 3–9, 1989, doi: 10.1109/58.16962.

[13] E. S. Ebbini and C. A. Cain, "Multiple-Focus Ultrasound Phased-Array Pattern Synthesis: Optimal Driving-Signal Distributions for Hyperthermia," *IEEE Trans Ultrason Ferroelectr Freq Control*, vol. 36, no. 5, pp. 540–548, 1989, doi: 10.1109/58.31798.

[14] R. Gerchberg, "A practical algorithm for the determination of phase from image and diffraction plane pictures," *Optik (Stuttg)*, vol. 35, pp. 237–246, 1972.

[15] Y. Hertzberg, O. Naor, A. Volovick, and S. Shoham, "Towards multifocal ultrasonic neural stimulation: Pattern generation algorithms," *J Neural Eng*, vol. 7, no. 5, p. 56002, Oct. 2010, doi: 10.1088/1741-2560/7/5/056002.

[16] J. Liu, J. Foiret, D. N. Stephens, O. Le Baron, and K. W. Ferrara, "Development of a spherically focused phased array transducer for ultrasonic image-guided hyperthermia," *Phys Med Biol*, vol. 61, no. 14, pp. 5275–5296, Jun. 2016, doi: 10.1088/0031-9155/61/14/5275.

[17] D. Gourevich *et al.*, "Ultrasound-Mediated Targeted Drug Delivery Generated by Multifocal Beam Patterns: An In vitro Study," *Ultrasound Med Biol*, vol. 39, no. 3, pp. 507–514, 2013.

[18] S. Jimenez-Gambin, N. Jimenez, A. Pouliopoulos, J. M. Benlloch, E. Konofagou, and F. Camarena, "Acoustic Holograms for Bilateral Blood-Brain Barrier Opening in a Mouse Model," *IEEE Trans Biomed Eng*, vol. 69, no. 4, pp. 1359–1368, Apr. 2022.

[19] S. Jiménez-Gambín, N. Jiménez, J. M. Benlloch, and F. Camarena, "Holograms to Focus Arbitrary Ultrasonic Fields through the Skull," *Physical Review Applied* , vol. 12, no. 1, p. 14016, 2019.

[20] K. Melde, A. G. Mark, T. Qiu, and P. Fischer, “Holograms for acoustics,” *Nature*, vol. 537, no. 7621, pp. 518–522, Sep. 2016.

[21] T. Ilovitsh, A. Ilovitsh, J. Foiret, B. Z. Fite, and K. W. Ferrara, “Acoustical structured illumination for super-resolution ultrasound imaging,” *Commun Biol*, vol. 1, no. 1, pp. 1–11, Dec. 2018.

[22] J. A. Jensen, “FIELD: A program for simulating ultrasound systems,” in *Nordic-Baltic Conference on Biomedical Imaging Published in Medical & Biological Engineering & Computing*, 1996, pp. 351–353.

[23] Jϕ. A. Jensen and N. B. Svendsen, “Calculation of Pressure Fields from Arbitrarily Shaped, Apodized, and Excited Ultrasound Transducers,” *IEEE Trans Ultrason Ferroelectr Freq Control*, vol. 39, no. 2, pp. 262–267, 1992.

[24] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[25] B. Ait Skourt, A. El Hassani, and A. Majda, “Lung CT Image Segmentation Using Deep Neural Networks,” *Procedia Comput Sci*, vol. 127, pp. 109–113, Jan. 2018.

[26] H. Yang *et al.*, “Unpaired brain mr-to-ct synthesis using a structure-constrained cyclegan,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11045 LNCS, pp. 174–182, 2018.

[27] K. Umehara, J. Ota, and T. Ishida, “Application of Super-Resolution Convolutional Neural Network for Enhancing Image Resolution in Chest CT,” *J Digit Imaging*, vol. 31, no. 4, pp. 441–450, Aug. 2018.

[28] X. Liu, J. Shi, and Q. Zhang, “Tumor classification by deep polynomial network and multiple kernel learning on small ultrasound image dataset,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9352, pp. 313–320, 2015.

[29] J. Shi, S. Zhou, X. Liu, Q. Zhang, M. Lu, and T. Wang, “Stacked deep polynomial network based representation learning for tumor classification with small ultrasound image dataset,” *Neurocomputing*, vol. 194, pp. 87–94, Jun. 2016.

[30] X. Liu, J. Song, S. Hong Wang, J. Zhao, and Y. Chen, “Learning to Diagnose Cirrhosis with Liver Capsule Guided Ultrasound Image Classification,” *Sensors*, vol. 17, no. 1, p. 149, Jan. 2017.

[31] J. Lu and W. Liu, “Unsupervised Super-Resolution Framework for Medical Ultrasound Images Using Dilated Convolutional Neural Networks,” *IEEE International Conference on Image, Vision and Computing, ICIVC*, pp. 739–744, Oct. 2018.

[32] M. Hossein Eybposh, N. W. Caira, M. Atisa, P. Chakravarthula, and N. C. Pégard, “DeepCGH: 3D computer-generated holography using deep learning,” *Opt Express*, vol. 28, no. 18, p. 26636, Aug. 2020.

[33] R. Horisaki, Y. Nishizaki, K. Kitaguchi, M. Saito, and J. Tanida, “Three-dimensional deeply generated holography,” *Appl Opt*, vol. 60, no. 4, p. A323, Feb. 2021.

[34] H. Azhari, *Basics of biomedical ultrasound for engineers*. Wiley, 2010.

[35] M. D. J. Quinn *et al.*, “Dense speed-of-sound shift imaging for ultrasonic thermometry,” *Phys Med Biol*, vol. 68, no. 21, p. 215004, Oct. 2023, doi: 10.1088/1361-6560/ACFEC3.

[36] T. Mano, T. Grutman, and T. Ilovitsh, “Versatile Ultrasound-Compatible Microfluidic Platform for In Vitro Microvasculature Flow Research and Imaging Optimization,” *ACS Omega*, Dec. 2023, doi: 10.1021/ACSOMEGA.3C05849.

[37] A. Anvari, F. Forsberg, and A. E. Samir, “A Primer on the Physical Principles of Tissue Harmonic Imaging,” *Radiographics*, vol. 35, no. 7, pp. 1955–1964, Nov. 2015, doi: 10.1148/RG.2015140338.

[38] J. L. Prince and J. M. Links, *Medical Imaging Signals and Systems* , 2nd ed. Pearson, 2005. Accessed: Apr. 29, 2023. [Online]. Available: https://books.google.com/books/about/Medical_Imaging_Signals_and_Systems.html?hl=iw&id=HHrangEACAAJ

[39] Thomas L. Szabo, *Diagnostic Ultrasound Imaging: Inside Out*, 2nd ed. Elsevier Science & Technology, 2013.

[40] R. Abiteboul and T. Ilovitsh, "Optimized Simultaneous Axial Multifocal Imaging via Frequency Multiplexed Focusing," *IEEE Trans Ultrason Ferroelectr Freq Control*, vol. 69, no. 10, pp. 2930–2942, Oct. 2022, doi: 10.1109/TUFFC.2022.3200468.

[41] F. Prieur, B. Dénarié, A. Austeng, and H. Torp, "Multi-line transmission in medical imaging using the second-harmonic signal," *IEEE Trans Ultrason Ferroelectr Freq Control*, vol. 60, no. 12, pp. 2682–2692, Dec. 2013, doi: 10.1109/TUFFC.2013.2868.

[42] M. G. L. Gustafsson, "Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy.," *J Microsc*, vol. 198, no. 2, pp. 82–87, May 2000.

[43] A. G. York *et al.*, "Resolution doubling in live, multicellular organisms via multifocal structured illumination microscopy," *Nat Methods*, vol. 9, no. 7, pp. 749–754, Jul. 2012.

[44] K. Kumagai, S. Hasegawa, and Y. Hayasaki, "Volumetric bubble display," *Optica*, vol. 4, no. 3, p. 298, Mar. 2017.

[45] K. Dholakia and T. Čižmár, "Shaping the future of manipulation," *Nat Photonics*, vol. 5, no. 6, pp. 335–342, Jun. 2011.

[46] D. L. Liu and R. C. Waag, "Propagation and backpropagation for ultrasonic wavefront design," *IEEE Trans Ultrason Ferroelectr Freq Control*, vol. 44, no. 1, pp. 1–13, 1997, doi: 10.1109/58.585184.

[47] S. O. Arik *et al.*, “Deep Voice 2: Multi-Speaker Neural Text-to-Speech,” *Adv Neural Inf Process Syst*, vol. 2017-December, pp. 2963–2971, May 2017, Accessed: Jun. 07, 2023. [Online]. Available: https://arxiv.org/abs/1705.08947v2

[48] A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, and M. Chen, “Hierarchical Text-Conditional Image Generation with CLIP Latents,” Apr. 2022, Accessed: Jun. 07, 2023. [Online]. Available: https://arxiv.org/abs/2204.06125v1

[49] E. Schwartz, R. Giryes, and A. M. Bronstein, “DeepISP: Towards Learning an End-to-End Image Processing Pipeline,” *IEEE Transactions on Image Processing*, vol. 28, no. 2, pp. 912–923, Jan. 2018, doi: 10.1109/TIP.2018.2872858.

[50] T. B. Brown *et al.*, “Language Models are Few-Shot Learners,” *Adv Neural Inf Process Syst*, vol. 2020-December, May 2020.

[51] D. Perrin *et al.*, “Application of convolutional artificial neural networks to echocardiograms for differentiating congenital heart diseases in a pediatric population,” *Medical Imaging 2017: Computer-Aided Diagnosis*, vol. 10134, no. 3, pp. 779–787, Mar. 2017.

[52] S. Goudarzi, A. Asif, and H. Rivaz, “Fast Multi-Focus Ultrasound Image Recovery Using Generative Adversarial Networks,” *IEEE Trans Comput Imaging*, vol. 6, pp. 1272–1284, 2020.

[53] R. J. G. Van Sloun *et al.*, “Super-Resolution Ultrasound Localization Microscopy through Deep Learning,” *IEEE Trans Med Imaging*, vol. 40, no. 3, pp. 829–839, Mar. 2021.

[54] Vinod Nair and Geoffrey E. Hinton, “Rectified Linear Units Improve Restricted Boltzmann Machines,” in *ICML*, 2010.

[55] B. Xu, N. Wang, H. Kong, T. Chen, and M. Li, “Empirical Evaluation of Rectified Activations in Convolutional Network,” May 2015.

[56] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, Dec. 2015, pp. 770–778.

[57] K. Zhang, M. Sun, T. X. Han, X. Yuan, L. Guo, and T. Liu, “Residual Networks of Residual Networks: Multilevel Residual Networks,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 6, pp. 1303–1314, Aug. 2016.

[58] W. H. Lopez Pinaya, S. Vieira, R. Garcia-Dias, and A. Mechelli, “Autoencoders,” *Machine Learning: Methods and Applications to Brain Disorders*, pp. 193–208, Mar. 2020, doi: 10.1016/B978-0-12-815739-8.00011-0.

[59] S. Ioffe and C. Szegedy, “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,” in *International conference on machine learning*, Feb. 2015, pp. 448–456. doi: 10.48550/arxiv.1502.03167.

[60] I. Sutskever, J. Martens, G. E. Dahl, and G. E. Hinton, “On the importance of initialization and momentum in deep learning,” *International Conference on Machine Learning*, Jul. 2013.

[61] M. H. Lee, H. M. Lew, S. Youn, T. Kim, and J. Y. Hwang, “Deep learning-based framework for fast and accurate acoustic hologram generation,” *IEEE Trans Ultrason Ferroelectr Freq Control*, vol. 69, no. 12, pp. 3353–3366, Dec. 2022, doi: 10.1109/TUFFC.2022.3219401.

[62] Q. Lin *et al.*, “A deep learning approach for the fast generation of acoustic holograms,” *J Acoust Soc Am*, vol. 149, no. 4, p. 2312, Apr. 2021, doi: 10.1121/10.0003959.

[63] R. Horisaki, R. Takagi, and J. Tanida, “Deep-learning-generated holography,” *Appl Opt*, vol. 57, no. 14, p. 3859, May 2018.

[64] D. Schein, T. Grutman, and T. Ilovitsh, “Deep Learning-Based Ultrasound Beam Shaping for Spatiotemporal Acoustic Holograms Generation,” *IEEE Trans Ultrason Ferroelectr Freq Control*, vol. 70, no. 6, pp. 551–561, Jun. 2023, doi: 10.1109/TUFFC.2023.3264580.

[65] W. Weng and X. Zhu, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” *IEEE Access*, vol. 9, pp. 16591–16603, May 2015, doi: 10.1109/ACCESS.2021.3053408.

[66] A. Paszke *et al.*, “PyTorch: An Imperative Style, High-Performance Deep Learning Library,” in *Advances in neural information processing systems*, Dec. 2019.

[67] P. Virtanen *et al.*, “SciPy 1.0: fundamental algorithms for scientific computing in Python,” *Nat Methods*, vol. 17, no. 3, pp. 261–272, Mar. 2020, doi: 10.1038/S41592-019-0686-2.

[68] P. J. Huber, “Robust Estimation of a Location Parameter,” *The Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 73–101, Mar. 1964.

[69] A. M. Saxe, J. L. McClelland, and S. Ganguli, “Exact solutions to the nonlinear dynamics of learning in deep linear neural networks,” *2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings*, Dec. 2013, Accessed: Jun. 14, 2023. [Online]. Available: https://arxiv.org/abs/1312.6120v3

[70] K. T. Karlinsky, M. Bismuth, R. Aronovich, and T. Ilovitsh, “Nonlinear Frequency Mixing Ultrasound Imaging of Nanoscale Contrast Agents,” *IEEE Trans Biomed Eng*, vol. PP, 2023, doi: 10.1109/TBME.2023.3321743.

[71] T. Ilovitsh, A. Ilovitsh, J. Foiret, and K. W. Ferrara, “Imaging beyond ultrasonically-impenetrable objects,” *Sci Rep*, vol. 8, no. 1, pp. 1–11, Apr. 2018.

[72] A. Ilovitsh, T. Ilovitsh, and K. W. Ferrara, “Multiplexed ultrasound beam summation for side lobe reduction,” *Sci Rep*, vol. 9, no. 1, pp. 1–8, Sep. 2019.

[73] L. Dong *et al.*, “Extended-depth-of-field object detection with wavefront coding imaging system,” *Pattern Recognit Lett*, vol. 125, pp. 597–603, Jul. 2019.

[74] A. Ilovitsh, N. Levanon, T. Ilovitsh, E. Preter, and Z. Zalevsky, “Super-resolution using Barker-based array projected via spatial light modulator,” *Opt Lett*, vol. 40, no. 8, pp. 1802–1805, Apr. 2015.

[75] M. Xiao, Y. Zhang, and J. Wen, “The Talbot effect: recent advances in classical optics, nonlinear optics, and quantum optics,” *Adv Opt Photonics*, vol. 5, no. 1, pp. 83–130, Mar. 2013, doi: 10.1364/AOP.5.000083.

[76] K. Weiss, T. M. Khoshgoftaar, and D. Wang, “A survey of transfer learning,” *J Big Data*, vol. 3, no. 1, p. 9, Dec. 2016, doi: 10.1186/s40537-016-0043-6.

[77] G. Hinton, O. Vinyals, and J. Dean, “Distilling the Knowledge in a Neural Network,” *arXiv e-prints*, Mar. 2015.

[78] “SGD Explained | Papers With Code.” Accessed: Sep. 09, 2023. [Online]. Available: https://paperswithcode.com/method/sgd

תקציר

אולטרסאונד הינה שיטה בטוחה, זולה, מהירה, ואמינה לביצוע הדמיה בעומק גוף האדם, ולכן הפכה לאחת משיטות ההדמיה הנפוצות ביותר לאבחון רפואי. דימות אולטרסאונד סטנדרטי מתבצע על ידי שימוש במתמרים שמכילים מספר רב של אלמנטי שידור וקליטה, תוך שימוש בפיקוס הלוך-חזור. בשידור, חזית הגל האקוסטית מפוקסת לנקודה בודדת על ידי הוספת השהיות לכל אחד מהאלמנטים בצורה פרבולית, כך שכל הפולסים המשודרים יבצעו התאבכות בונה במיקום ספציפי וזמן ספציפי על מנת ליצור נקודת פוקוס. שיטה זאת מניבה רזולוציה לטרלית גבוהה, מגבירה את יחס האות לרעש ומאפשרת עומק חדירה גדול יותר בהשוואה לשיטות ההדמיה האולטראסוניות האחרות.

חסרון משמעותי בפיקוס הלוך-חזור נובע מכך שבכל שידור רק נקודת פוקוס בודדת מיוצרת, ועל מנת ליצור תמונה נדרש להעביר את הנקודה המיוצרת על פני כל המרחב שרוצים לדמות. התהליך הזה לוקח זמן, ומגביל את קצב הפריימים האפשרי בצורה פרופורציונלית למספר הפוקוסים שרוצים לייצר עבור תמונה בודדת. לחילופין, ניתן להשתמש בדימות מולטיפוקלי על מנת להגביר את קצב הפריימים. ניתן להשיג זאת באמצעות עיצוב הגל האקוסטי שמשודר, על ידי הנדסת הפאזות המשודרות על מנת לייצר תבנית רצויה. אלגוריתמים של אחזור פאזה בהשראה אופטית ושיטות סטנדרטיות לעיצוב חזית הגל האקוסטית מצליחים לייצר הולוגרמות אקוסטיות לטובת יישומים טיפוליים אשר כוללים שידור פרצים ארוכים. למרות זאת, נדרשת שיטה להנדסת הפאזה אשר מיועדת לשידור של מחזור בודד ומסוגלת להניב התאבכות בונה במרחב ובזמן של הגלים המשודרים לטובת יישומי הדמיה.

לטובת מטרה זאת, פיתחנו רשת קונבולוציה עמוקה לחישוב התהליך ההופכי לקידום חזית הגל מהמתמר למישור הדימות, כלומר חישוב הפאזות הנצרכות עבור יצירה של תבנית מולטיפוקלית. רשת זאת (USDL) אומנה על זוגות אימון מדומים של תבניות מולטיפוקליות במישור הדימות ותבניות הפאזה התואמות במישור המתמר, כך שקידום חזית הגל התבצע תוך הנחת שידור מחזור בודד. שיטת הUSDL הציגה ביצועים טובים יותר ביחס לשיטה הסטנדרטית Gerchberg-Saxton, כאשר משדרים מחזור בודד, בפרמטרים שכוללים את מספר נקודות הפוקוס שנוצרו, העוצמה שלהם, והאחידות שלהם. בנוסף, USDL הוכחה להיות יותר גמישה בייצור תבניות עם ריווח גדול של פוקוסים, מרווחים לא אחידים, ועוצמות משתנות. בסימולציות, השיפור הכי משמעותי הושג עבור תבניות של 4 פוקוסים כאשר

Gerchberg-Saxton הצליח לייצר סך הכל 25% מהתבניות שהוא התבקש, בעוד USDL הצליח לייצר 60% מהתבניות. התוצאות אומתו על ידי ביצוע מדידות היידרופון.

הממצאים שלנו מצביעים על כך שעיצוב חזית גל מבוסס למידה עמוקה יכול לאפשר את הדור הבא של הולוגרמות אקוסטיות עבור יישומי דימות באולטרסאונד.

אוניברסיטת תל אביב

הפקולטה להנדסה ע"ש איבי ואלדר פליישמן

בית הספר לתארים מתקדמים ע"ש זנדמן-סליינר

עיצוב חזית גל מבוסס למידה עמוקה לטובת הולוגרמות אקוסטיות בשידור גל לא רציף

חיבור זה הוגש כעבודת גמר לקראת התואר "מוסמך אוניברסיטה" בהנדסת חשמל

על-ידי

דרור שיין

העבודה נעשתה במחלקה להנדסה ביו-רפואית

בהנחיית ד"ר טלי אילוביץ

אוניברסיטת תל אביב

הפקולטה להנדסה ע"ש איבי ואלדר פליישמן

בית הספר לתארים מתקדמים ע"ש זנדמן-סליינר

עיצוב חזית גל מבוסס למידה עמוקה לטובת הולוגרמות אקוסטיות בשידור גל לא רציף

חיבור זה הוגש כעבודת גמר לקראת התואר "מוסמך אוניברסיטה" בהנדסת חשמל

על-ידי

דרור שיין

העבודה נעשתה במחלקה להנדסה ביו-רפואית

בהנחיית ד"ר טלי אילוביץ