
A deep learning approach to improving ultrasonic plane wave imaging

James Long

Department of Biomedical Engineering
Duke University

Abstract

This work explores the use of a deep learning approach to improving ultrasonic plane wave imaging. By training a U-Net based CNN, images of quality comparable to that produced with 20 plane waves were able to be reproduced from data of images produced with 3 plane waves. Qualitatively, improvements in border definition and resolution were observed. Additionally, this network improves preservation of native contrast, but does not show the ability to resolve point targets. These initial findings provide evidence that information captured with many plane waves may be redundant, and the movement towards unblurred ultrafast imaging appears feasible.

1 Introduction

1.1 Ultrasonic plane wave imaging

In conventional pulse-echo ultrasound, acoustic waves are transmitted from an array of piezoelectric elements to insonify a target of interest. The received echoes are then delayed and summed across the aperture to reconstruct images of the target based on the intensity of the backscattered wave. The propagating acoustic wave can also be spatially delayed on transmit to create a variety of focal geometries. Plane wave imaging is one such geometry, by which an irradiated plane wave can be used to reconstruct an entire image. **Figure 1** shows a simple schematic for insonification and detection. This technique is similar to optical holography¹, by which light originating from a known source is recorded on a holographic plate with a referenced plane.

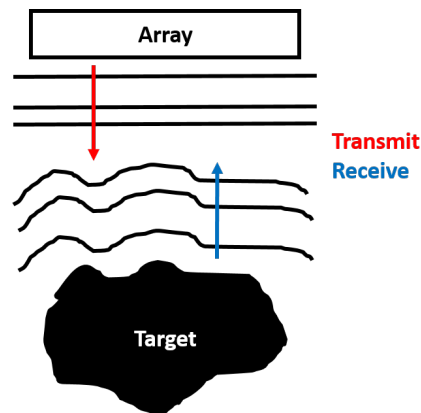


Figure 1: Schematic of ultrasonic plane wave imaging of a target

Single plane wave imaging has the distinction of achieving the fastest possible frame rate allowable, but suffers from low signal-to-noise ratio (SNR) and resolution in traditional delay-and-sum (DAS) beamforming. Multiple plane waves can be coherently compounded to increase SNR and image quality, but they come at the expense of frame rate. There exists several alternative beamforming methods beyond DAS to improve SNR and selected image quality metrics²⁻⁴, but they come with their own limitations, including limited resolution, poor preservation of native contrast, and introduction of low-energy artifacts.

Ultrafast imaging, a technique introduced by Tanter and Fink⁵, seeks to circumvent the trade-off between image quality and frame rate by reconstructing images using recycled plane wave transmissions to create subsequent images. For example, for an image generated with 7 plane waves, the last 6 transmitted waves (and reconstructed images) can be combined with a new plane wave, transmitted at the same angle as the newly omitted wave, to reconstruct a new image. By refreshing the image with each new plane wave, an image with the quality of that reconstructed with several plane wave can be acquired in the time required of a single plane wave image. Despite the theoretical improvements in image quality without the expense of frame rate, ultrafast imaging remain susceptible to temporal blurring due to coherently compounding images over a non-transient time period. While this effect is not expected to be a factor in conventional B-mode imaging, it certainly has implications for techniques requiring very high frame rates, including shear wave elastography and Doppler imaging.

This work seeks to overcome temporal blurring and other trade-offs associated with alternative beamforming methods, while retaining the improvements in image quality seen in images acquired with a high number of plane waves, by leveraging deep learning approaches. The objective of this work is reproduce images acquired with 20 plane waves from data acquired with only 3 plane waves. This work is novel in that it combines non-medical images with diffraction-limited performance constraints to arrive at images that are still physically possible to recreate, but with a more vast set of trained features.

2 Methods

2.1 Network

Consider a set of plane wave images, \hat{x}_k , in pixel space such that each image $x_i \in \mathbb{R}^N$, where k is the total number of images and N is the number of pixels reconstructed. This series of plane wave images can be coherently compounded to form a final, high-quality image $x_k^* \in \mathbb{R}^N$. This work proposes to train a convolutional neural network (CNN), represented as $f_{\text{CNN}}(\hat{x}_k)$. As explained previously, we can expect higher quality images as k increases; the proposed network will perform the following mapping to extract high-quality images from low-quality data:

$$\hat{x}_{k_H}^* = f_{\text{CNN}}(\hat{x}_{k_L}) \quad (1)$$

where k_H is the number of plane wave images required of a high-quality image, k_L is the number of plane wave images of a low-quality image, and $k_H > k_L$.

The proposed network architecture to achieve this task will be adapted from the popular U-Net, a well-validated network for segmentation of biomedical images⁶. The advantage of using a U-Net-like architecture is the combination of preserved low-level extracted features with learned high-level extracted features. In the context of ultrasonic imaging, this means that theoretically, the learned model will contain methods to extract system-dependent features (low-level, e.g. carrier frequency, channel size) as well as target-dependent features (high-level, e.g. echogenicity patterns). For the purposes of this project, we will use $k_H = 20$ and $k_L = 3$. A list of network operations is shown in **Table 1**.

The Adam optimizer was used with a learning rate of 5×10^{-5} and the mean-squared error (MSE) was used as the training loss. The labels were envelope-detected images of the high-quality coherently compounded images, meaning the network was expected to learn envelope detection in addition to image matching. This operation is reflected in the larger kernel size in Layer 9 (**Table 1**). This was intentionally chosen to avoid the network relying on the carrier frequency for loss minimization. The network was trained over 490 images in batches of 10, and validated with 10 separate images.

Table 1: CNN parameters, with steps for maxpooling and upsampling skipped for readability

Layer	operation	n_{filters}	filter size	data size
0	input	-	-	$512 \times 192 \times k_L$
1	conv2d	32	3×3	$512 \times 192 \times 32$
2	conv2d	64	3×3	$256 \times 96 \times 64$
3	conv2d	128	3×3	$128 \times 48 \times 128$
4	conv2d	256	3×3	$64 \times 24 \times 256$
5	conv2d	512	3×3	$32 \times 12 \times 512$
6	conv2d	256	3×3	$64 \times 24 \times 256$
7	conv2d	128	3×3	$128 \times 48 \times 128$
8	conv2d	64	3×3	$256 \times 96 \times 64$
9	conv2d	32	16×16	$512 \times 192 \times 32$
10	conv2d	2	2×2	512×192

2.2 Dataset simulation

The Field II Pro simulation package⁷ was used to simulate ultrasound channel data from a linear array. Field II calculates a spatially-variant point spread function (PSF) and convolves it with a field of acoustic scatterers with prescribed backscattering amplitude, an operation based on the Fresnel approximation^{8,9}. The user is allowed to tune the imaging parameters (e.g. transmit frequency, number of array elements, transmit delays) and provide the spatial locations of the scatterers to produce images. While Field II does not take into account non-linear effects due to acoustic propagation, it is computationally efficient and has proven to be an accurate approximation for good acoustic environments.

To quickly simulate a variety of features and echogenicity patterns, 64×64 8-bit images from Tiny ImageNet¹⁰ were converted to echogenicity maps of scatterers. The classification labels were not used. An example of the conversion process is seen in **Figure 2**. From the ground truth images, we can spatially encode echogenicities that are applied to scatterers residing in the respective regions. Received channel data were saved for k_H transmissions, and k_L sub-images were extracted as surrogate low-quality images.

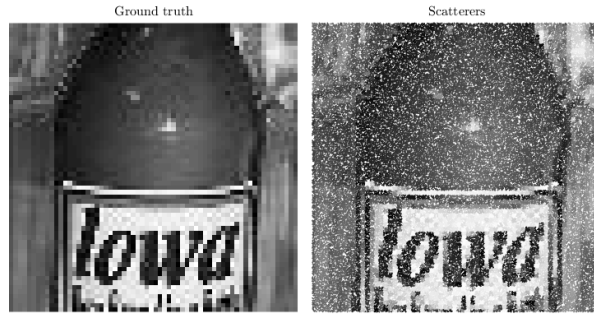


Figure 2: Conversion from ground truth images to echogenicity map of scatterers

To validate the final performance of the network, cysts phantoms with $-\infty$, -6 , and $+6$ dB native contrasts, as well as a point target phantom, were generated. Contrast was calculated as follows:

$$C = 20 \log_{10} \left(\frac{\mu_{in}}{\mu_{out}} \right) \quad (2)$$

where μ_{in} is the mean signal intensity of envelope-detected data inside the cyst, and μ_{out} is that outside the cyst. Lastly, resolution was calculated as the full-width half maximum (FWHM) of the PSF. Metrics were compared between the low-quality image ($\hat{x}_{k_L}^*$), low-quality image passed through the CNN ($\hat{x}_{k_L + \text{CNN}}^*$), and high-quality image ($\hat{x}_{k_H}^*$).

3 Results and discussion

3.1 Network performance

Figure 3 shows the trends in validation loss over epochs. As expected, roughly monotonic decreases in validation loss were observed with each successive epoch. The network was allowed to train for up to 100 epochs, but due to an early stopping condition (minimum change in validation loss of 5×10^{-5} with a patience of 5 epochs), it trained for only 63 epochs. The final validation loss was 1.428×10^{-3} , although the smallest validation loss was observed at epoch 58 (1.421×10^{-3}).

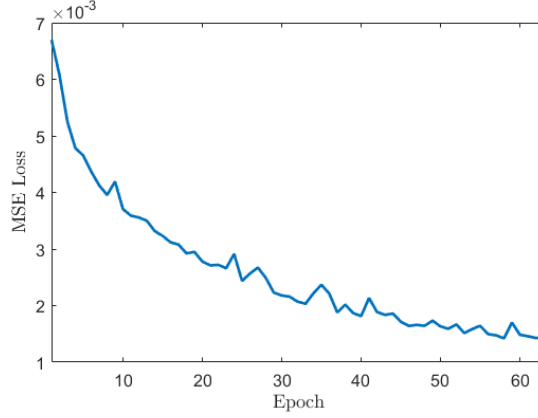


Figure 3: MSE validation loss per epoch

3.2 Qualitative assessment

Figure 4 shows $\hat{x}_{k_L}^*$, $\hat{x}_{k_L+CNN}^*$, $\hat{x}_{k_H}^*$, and the ground truth tiny ImageNet image for selected validation cases. The improvements from $\hat{x}_{k_L}^*$ to $\hat{x}_{k_L+CNN}^*$ are appreciable; there is noticeable recovery of border delineation and native contrast. $\hat{x}_{k_L+CNN}^*$ also appears to contain higher spatial frequencies, approaching that of $\hat{x}_{k_H}^*$. Furthermore, there is suppression of artifacts originating from poor SNR associated with DAS beamforming.

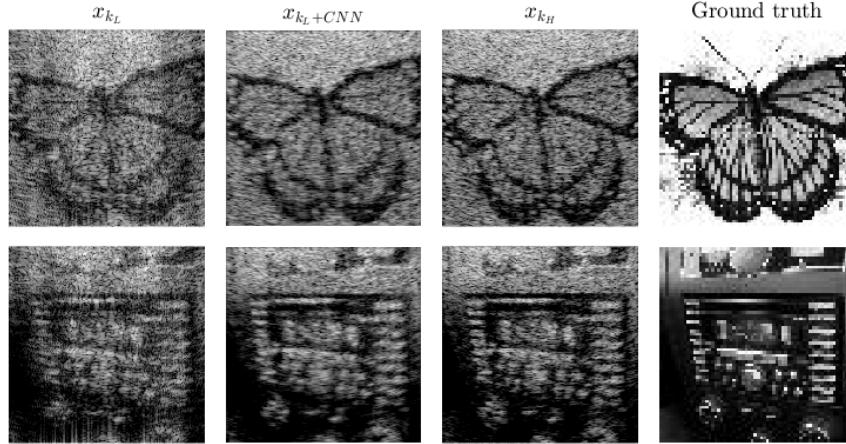


Figure 4: Columns left to right: $\hat{x}_{k_L}^*$, $\hat{x}_{k_L+CNN}^*$, $\hat{x}_{k_H}^*$, and ground truth images

Note that $\hat{x}_{k_H}^*$, despite being the label for a high-quality image, does not transfer all features from the ground truth images. This is due to diffraction limits for resolution, as well as the phenomenon of speckle, which arises from all forms of coherent imaging. These resolution limits and speckle further

obscure boundaries and distort image contrast by redistributing energy from the backscattered echoes to the side lobes of the associated PSF. Therefore, techniques for speckle reduction and improving border definition are largely confined to operations post-beamforming. Matching $\hat{x}_{k_L+CNN}^*$ to the ground truth is not a focus of this work, but it is an active area of research¹¹. This work, however, is synergistic with such efforts; the goal is to be able use these developed algorithms on $\hat{x}_{k_L+CNN}^*$ to arrive at even higher-quality images.

3.3 Quantitative assessment

Figure 5 shows $\hat{x}_{k_L}^*$, $\hat{x}_{k_L+CNN}^*$, and $\hat{x}_{k_H}^*$ for the anechoic, -6 dB, +6 dB, and point target phantoms. Quantitatively and qualitatively, we observe recapturing of the native contrasts between $\hat{x}_{k_L}^*$ and $\hat{x}_{k_L+CNN}^*$. The measured contrast is shown in **Table 2**.

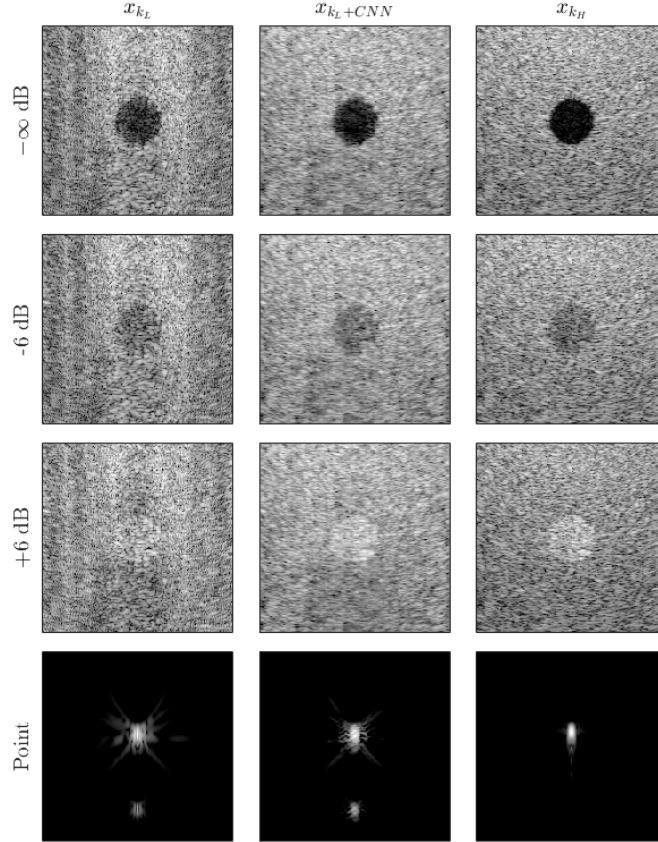


Figure 5: Columns left to right: $\hat{x}_{k_L}^*$, $\hat{x}_{k_L+CNN}^*$, and $\hat{x}_{k_H}^*$ images

Table 2: Contrast across conditions, all values in dB relative to background tissue

C_{native}	$\hat{x}_{k_L}^*$	$\hat{x}_{k_L+CNN}^*$	$\hat{x}_{k_H}^*$
$-\infty$	-18.3	-21.8	-22.5
-6	-8.3	-6.9	-6.4
+6	+0.2	+5.2	+5.5

We see that across all conditions that $\hat{x}_{k_L+CNN}^*$ more closely approximates both the native contrast and that of $\hat{x}_{k_H}^*$ compared to $\hat{x}_{k_L}^*$. Interestingly, the contrast in $\hat{x}_{k_L}^*$ was lower than that of both $\hat{x}_{k_L+CNN}^*$ and $\hat{x}_{k_H}^*$; however, it does not match the native contrast of -6 dB.

Measurements of the PSF resolution were not attempted due to a clear failure to reproduce $\hat{x}_{k_H}^*$ for the point target. To a degree, this is expected: the training set did not contain images with sparsely spaced point targets. Therefore, there was no opportunity for the network to learn how to resolve individual scatterers.

3.4 Limitations and future work

The most noticeable shortcoming of this work is represented by vertical banding present in $\hat{x}_{k_L}^*$. This is a result of incomplete spatial coverage of the region shown by the 3 selected plane waves. The implications of this are two-fold: (1) the SNR outside of the central bright band is extremely poor, and (2) it is highly likely that information for the two outer plane waves is not used at all. This suggests that a network trained on only one plane wave may have comparable results to the one shown in this work. This issue is easily solved by moving the depth of the region-of-interest shallow to the array and forcing more finely spaced angles.

Another limitation of this work is the use of only 1 speckle realization for all images trained. Therefore, there is the possibility that the network relies heavily on learning the speckle pattern to extract echogenicity information. However, the images shown in **Figure 5** were produced with a different speckle pattern, and appear to still show improvements in image quality. Future iterations of this work will randomize speckle patterns per image realization for a more robust training set.

The next step for this work is to explore the use of plane wave selection within the model for optimal extraction of spatial frequencies. For example, an image with significant energy at high lateral spatial frequencies will see marked improvements in resolving borders when moving from a plane wave at 0° (i.e. propagating orthogonal to the array) to a plane wave at 5° .

4 Conclusion

This work explores the use of a deep learning approach to improving ultrasonic plane wave imaging. By training a U-Net based CNN, images of quality comparable to that produced with 20 plane waves were able to be reproduced from data of images produced with 3 plane waves. Qualitatively, improvements in border definition and resolution were observed. Additionally, this network improves preservation of native contrast, but does not show the ability to resolve point targets. These initial findings provide evidence that information captured with many plane waves may be redundant, and the movement towards unblurred ultrafast imaging appears feasible.

Acknowledgments

The author would like to thank R. Horstmeyer, O. Huang, and K. Zhou for extensive guidance, as well as W. Long, M. Morgan, and N. Bottenus for insightful comments and suggestions.

References

- [1] Gabor, D. (1948). A new microscopic principle. *Nature*, 161, 777-778.
- [2] Li, P. C., & Li, M. L. (2003). Adaptive imaging using the generalized coherence factor. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 50(2), 128-141.
- [3] Lediju, M. A., Trahey, G. E., Byram, B. C., & Dahl, J. J. (2011). Short-lag spatial coherence of backscattered echoes: Imaging characteristics. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 58(7), 1377-1388.
- [4] Fink, M. (1992). Time reversal of ultrasonic fields. I. Basic principles. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 39(5), 555-566.
- [5] Montaldo, G., Tanter, M., Bercoff, J., Benech, N., & Fink, M. (2009). Coherent plane-wave compounding for very high frame rate ultrasonography and transient elastography. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 56(3), 489-506.
- [6] Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (234-241). Springer, Cham.
- [7] Jensen, J. A. (2014, September). A multi-threaded version of Field II. In *2014 IEEE International Ultrasonics Symposium* (2229-2232). IEEE.
- [8] Goodman, J. W. (2005). *Introduction to Fourier optics*. Roberts and Company Publishers.
- [9] Cobbold, R. S. (2006). *Foundations of biomedical ultrasound*. Oxford university press.
- [10] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (248-255). IEEE.
- [11] Hyun, D., Brickson, L. L., Looby, K. T., & Dahl, J. J. (2019). Beamforming and speckle reduction using neural networks. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*. Pre-print.