# Data Driven Methods for Ultrasound Computed Tomography

Luke Lozenski[1], Hanchen Wang[2], Brendt Wohlberg[3], Umberto Villa[4], and Youzuo Lin[2]

[1]Department of Electrical and Systems Engineering,
Washington University in St. Louis, St. Louis, MO 63130, USA
[2]Earth and Environmental Sciences Division,
Los Alamos National Laboratory, Los Alamos, NM 87545, USA
[3]Theoretical Division,
Los Alamos National Laboratory, Los Alamos, NM 87545, USA
[4]Oden Institute for Computational Engineering & Sciences,
The University of Texas at Austin, TX, 78712

## ABSTRACT

Ultrasound computed tomography (USCT) is an emerging imaging modality that holds great promise for breast imaging. Full-waveform inversion (FWI)-based image reconstruction methods incorporate accurate wave physics to produce high spatial resolution quantitative images of speed of sound or other acoustic properties of the breast tissues. However, FWI reconstruction is computationally expensive, which limits its application in a clinical setting. This contribution investigates using the use of a convolutional neural network (CNN) to learn a mapping from USCT data to speed of sound estimates. The CNN was trained using a supervised approach that employed a large set of anatomically and physiologically realistic numerical breast phantoms (NBPs) and simulated USCT measurements. Once trained, this CNN can then be evaluated for real-time FWI image reconstruction from USCT data. The performance of the proposed method was assessed and compared against FWI using a hold-out sample of 41 NBPs and corresponding USCT images. Accuracy was measured using relative mean square error (RMSE) and structural self-similarity index measure (SSIM). This numerical experiment demonstrates that a supervised learning model can achieve accuracy comparable to FWI, while significantly reducing computational time and memory requirements.

## 1. INTRODUCTION

a In particular, USCT holds a great promise for breast cancer diagnosis.[1] Full waveform inversion (FWI) is an image reconstruction that allow to estimate quantitatively accurate values of speed of sound distribution, as well as of other tissue acoustic properties, from USCT data. FWI is a model-based iterative image reconstruction method that incorporate the solution of the wave equation in the evaluation of the imaging operator.[2,3] However, the computational burden associate with solving the wave equation at each FWI iteration has hampered the use of the FWI reconstruction approach in a clinical setting, in favor of less accurate but faster reconstruction methods such as bent-ray approaches.[4,5]

This work investigates a deep learning based image reconstruction method, InversionNet, to alleviate the computational burden of FWI. InversionNet is a convolutional neural network architecture originally proposed for seismic imaging applications.[6] This work proposes modifications of InversionNet for use in USCT image reconstruction, including the use of training/testing sets relevant to medical imaging and architectural changes to accommodate the larger amount of waveform data in USCT compared to seismic imaging. Once trained, InversionNet neural network can be used to reconstruct tissue speed of sound estimates from USCT data in just a few seconds, assuming that the data acquisition parameters are the same as those used for training.

A simulation study is conducted to compare InversionNet to an FWI. The objects used in this study are a large set of anatomically-realistic numerical breast phantoms with acoustic properties assigned to the corresponding tissues.[7] USCT measurements are virtually acquired over a circular aperture by numerical simulation of wave propagation in acoustically heterogeneous media. After training, InversionNet is then compared, in terms of RMSE and SSIM, to an FWI method on a hold-out sample of 41 numerical breast phantoms.

Further author information: (Send correspondence to Umberto Villa and Youzou Lin)
E-mail: uvilla@oden.utexas.edu, ylin@llnl.gov; Telephone: (512) 232-3453, (505) 667-7335

## 2. METHOD

**Imaging Operator** In USCT imaging, the object is insonified from several directions (shots). Acoustic pressure waves propagating through the object are then measured on an aperture $\mathcal{S}$ surrounding the object.

Assuming wave propagation in a two-dimensional (2D) acoustic nonlossy medium with homogeneous density and spatially varying speed of sound $c = c(\mathbf{r})$, the canonical USCT continuous to continous (C-C) imaging operator is defined as

$$\mathcal{H}^c s \coloneqq p(\mathbf{r}, t) \quad (\mathbf{r}, t) \in \mathcal{S} \times [0, T], \tag{1}$$

where the notation $\mathcal{H}^c$ is used to underline the dependence of the imaging operator on the medium speed of sound $c$. Above, $T$ denotes the acquisition time for a single shot, $s$ denotes the excitation pulse, and the acoustic pressure $p = p(\mathbf{r}, t)$ solves the wave equation

$$\frac{1}{c(\mathbf{r}, t)^2} \frac{\partial^2}{\partial t^2} p(\mathbf{r}, t) - \Delta p(\mathbf{r}, t) = s(\mathbf{r}, t) \quad (\mathbf{r}, t) \in \mathbb{R}^2 \times [0, T]. \tag{2}$$

Assuming that $M$ idealized point-like transducers are distributed along the measurement aperture at locations $\mathbf{r}_j \in \mathcal{S}$ $(j = 1, \dots M)$,the sampling operator $\mathcal{M}$ mapping the pressure $p(\mathbf{r}, t)$ to the vector $\boldsymbol{g} \in \mathbb{R}^{N_t M}$ is defined as

$$[\mathcal{M}p]_{k+(j-1)N_t} \coloneqq [\boldsymbol{g}]_{k+(j-1)N_t} = p(\mathbf{r}_j, k\Delta T) \quad j = 1, \dots, M; k = 1, \dots, N_t, \tag{3}$$

where $\Delta T = T/N_t$ is the sampling interval and $N_t$ is the number of pressure samples measured over the acquisition interval $[0, T]$.

Using the the continuous-to-discrete imaging operator $\mathcal{M}\mathcal{H}^c$, the USCT data acquisition process is modeled as

$$\boldsymbol{d}_i = \mathcal{M}\mathcal{H}^c s_i + \boldsymbol{n} \quad i = 1, \dots, m, \tag{4}$$

where $s_i \coloneqq s_i(\mathbf{r}, t)$ is the $i^{\text{th}}$ excitation pulse, $m$ is the number of shots, and $\boldsymbol{n} \in \mathbb{R}^{N_t \times M}$ is additive noise.

Finally, the discrete-to-discrete (D-D) imaging operator is established by introducing a Cartesian grid with $N$ pixels. Denoting with $\boldsymbol{r}_n$ the center of the $n^{\text{th}}$ pixel, the finite dimensional vectors $\boldsymbol{c} \in \mathbb{R}^N$ and $\boldsymbol{s}_i \in \mathbb{R}^{N_t M}$ are defined as

$$[\boldsymbol{c}]_n = c(\mathbf{r}_n), \ [\boldsymbol{s}_i]_{k+(n-1)N_t} = s_i(\mathbf{r}_n, k\Delta T) \quad n = 1, \dots, N; \ k = 1, \dots, N_t.$$

With the above notation, the D-D USCT model is given by

$$\boldsymbol{d}_i = \boldsymbol{M}\boldsymbol{H}^{\boldsymbol{c}}\boldsymbol{s}_i + \boldsymbol{n} \quad i = 1, \dots, m, \tag{5}$$

where $\boldsymbol{M} : \mathbb{R}^{N_t N} \mapsto \mathbb{R}^{N_t M}$ is the discrete counterpart of the sampling operator $\mathcal{M}$ defined via nearest neighbor interpolation of transducer coordinates to the pixel centers of the Cartesian grid, and $\boldsymbol{H}^{\boldsymbol{c}} : \mathbb{R}^{N_t N} \mapsto \mathbb{R}^{N_t N}$ stems from finite difference approximation of the C-C imaging operator $\mathcal{H}$.[8]

**Full Waveform Inversion** One method of reconstructing the speed of sound map $c$ given pressure traces $\boldsymbol{d}$ is full waveform inversion (FWI).[9] By use of the D-D USCT imaging model in Eq. (5), FWI seeks a speed of sound estimate $\hat{\boldsymbol{c}}$ such that

$$\hat{\boldsymbol{c}} \coloneqq \underset{\boldsymbol{c} \in \mathbb{R}^N}{\operatorname{argmin}} \frac{1}{2} \sum_{i=1}^{m} \|\boldsymbol{d}_i - \mathbf{H}(\mathbf{c})\mathbf{s}_i\|^2. \tag{6}$$

This discrete optimization problem can then be solved using a gradient-based method to update estimates of $\boldsymbol{c}$. This optimization process can be accelerated by leveraging multi-channel measurements and linearity of the wave equation with respect to the excitation pulse via a stochastic optimization method known as *source-encoding*.[2] Even with acceleration, these methods are still slow to converge and reconstructions for individual images may take several minutes or hours.

Table 1. Virtual imaging system and discretization parameters

| Ultrasound system | | Discretization | |
|---|---|---|---|
| Number of receivers $M$ | 256 | Grid size $Nx$ | 360 |
| Number of transmitters $m$ | 64 | Grid intervals $\delta x$ | 0.6 mm |
| Transducer radius $R$ | 110.4 mm | Points per wavelength $\frac{c_{min}}{f_0 \delta x}$ | 2.33 |
| Pulse frequency $f_0$ | 1 MHz | Number of time steps $N_t$ | 400 |
| Sampling frequency | 5 MHz | Time steps $\delta t$ | 0.2 $\mu$s |
| Acquisition time (per source) | 640 $\mu s$ | CFL Number $\frac{c_{max} \delta t}{\delta x}$ | 0.53 |

**Learned FWI via InversionNet** This work proposes leveraging a convolutional neural network (CNN) to learn an inversion map from USCT measurements to the speed of sound maps of breast tissues. Once trained, this CNN can be evaluated much quicker than FWI methods, significantly reducing the computational burden of reconstructions.

The specific CNN architecture used here is InversionNet.[6] InversionNet utilizes an end-to-end trained encoder-decoder structure. In this scheme, pressure traces are encoded to a high-dimensional topological space and then decoded to the space of images. Specifically, the input to InversionNet is a 3-d tensor $\boldsymbol{d} = \begin{pmatrix} \boldsymbol{d}_1 & \dots & \boldsymbol{d}_n \end{pmatrix} \in \mathbb{R}^{m \times N_t \times M}$ corresponding to USCT measurements collected for multiple excitation pulses and its output is a 2-d tensor $\boldsymbol{c} \in \mathbb{R}^{N \times N}$ corresponding to pixel values of speed of sound estimates over the field of view.

InversionNet was originally developed for FWI of seismic tomography data in geophysics, which presents several similarities to USCT imaging but with a few key differences. Primarily, measurements in seismic imaging are sparse and expensive to acquire while USCT measurements are quite large. This means that modifications to the InversionNet architecture are needed to accommodate a larger input space and also more training examples are needed to learn a larger number of network parameters.

Additionally, InversionNet was trained with a supervised approach which requires pairs of ground truth speeds of sound and pressure traces. Since InversionNet implicitly learns an image prior from the examples in the training set,[10] it is mandatory to train InversionNet on a set of images that are relevant to USCT and display sufficient variability. However, a large data sets of accurate speed of sound estimates from clinical USCT data are not readily available. Therefore the training and testing sets were constructed using anatomically accurate numerical breast phantoms (NBPs). These NBPs were constructed using tools from the Virtual Imaging Clinical Trial for Regulatory Evaluation (VICTRE) project at the US Food and Drugs Administration[11] and adapted for use in USCT virtual imaging studies.[7,12] In particular, the size and breast density of such phantoms is randomly sampled based on the four BI-RADS categories of breast density type: A] almost all fatty breasts, B] breasts with scattered fibroglandular density, C] breasts with heterogeneous density, and D] extremely dense breasts. Then anatomically realistic breast tissue structures are stochastically generated by use of the VICTRE tools and the speed of sound maps corresponding to physiological variations in breast tissues are stochastically generated.[7]

The parameters $\boldsymbol{\xi} \in \mathbb{R}^W$ of the InversionNet $\Phi_{\boldsymbol{\xi}}$ was then trained by minimizing the loss function

$$\min_{\boldsymbol{\xi} \in \mathbb{R}^w} \frac{1}{2N} \sum_{j=1}^{N} \|\Phi_{\boldsymbol{\xi}}(\boldsymbol{d}^j) - c^j\|^2, \tag{7}$$

where $\{(c^j, \boldsymbol{d}^j)\}_{j=1}^N$ and $\{\boldsymbol{d}^j\}_{j=1}^N$ are the speed of sounds maps and corresponding USCT measurements. Above, the $l^2$ norm of the difference between the speed of sound map and the output of $\Phi_{\boldsymbol{\xi}}$ is used for training, however other choices are possible.[6]

## 3. NUMERICAL STUDIES

A simulation study was performed to demonstrate the feasibility of the proposed method. The image quality achieved by InversionNet was compared to that of FWI with stochastic source-encoding.[2] The parameters of the virtual imaging study are presented in Table 1.

The training set and testing set consists of 1,353 and 41 NBPs and corresponding USCT data, respectively. Ultrasound measurements are simulated using the two-dimensional virtual imaging system pictured in Figure 1. The virtual imaging system has a circular measurement aperture with $M = 256$ idealized point-like transducers.

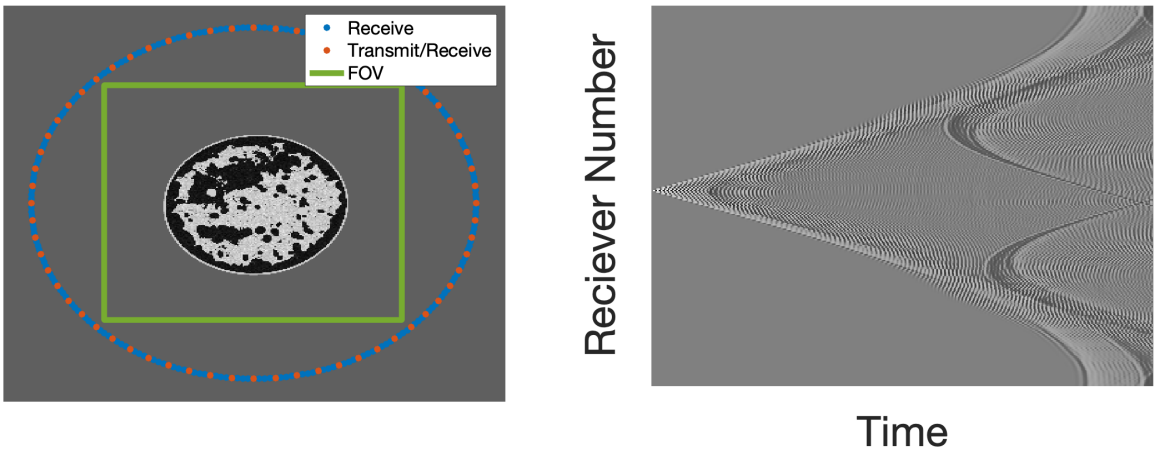


Figure 1. Virtual 2D USCT imaging system in (left): Receiving transducers (shown in blue and red), transmitting transducers (shown in red) are distributed along a circular aperture surrounding the field of view (shown in green). Example measurement (right) resulting from an excitation pulse from one transmitter and recorded pressure trace across all receiving transducers

Emitters are also equidistributed along the measurement aperture (every fourth transducers acts in both emitter and receiver mode). The excitation pulse of the $m = 64$ emitters is given by

$$s_i(\mathbf{r}, t) = \delta(\mathbf{r} - \mathbf{r}_j) \exp\left(-\frac{(t - t_0)^2}{2\sigma^2}\right) \sin(2\pi f_0 t) \quad i = 1, \ldots, m,$$

where $\mathbf{r}_j$ is the location of the emitter, $f_0$ is the frequency of the signal, $t_0$ is the time shift, and $\sigma$ controls the signal width. Measurements are collected by firing one transmitter at a time and recording data at every receiver. This is repeated for each transmitter and results in multi-channel measurements.

For each evaluation of the imaging operator in (5), the wave equation is solved using a finite difference scheme (4th order in space and 2nd order in time) using a spatial grid of size $N_x \times N_x$, with $N_x = 360$, and a spatial grid with $N_t = 400$ samples. Absorbing boundary conditions are implemented to prevent wave reflections at the boundaries of the computational domain. Electronic noise is modeled as a Gaussian additive with noise with a standard derivation of $10^{-5}$. Image quality is assessed using two metrics; relative root mean square error (RMSE) and structural self-similarity index measure (SSIM).[13]

An instance of InversionNET[6] is implemented in PyTorch, an open-source machine learning framework,[14] and trained for 1,000 epochs with a batch size of 50 using Adam optimizer.[15] Training InversionNet took approximately 10 hours on an HPC cluster with 512 GB of memory and 4 NVidia Volta V100 graphic processing units (GPUs). The FWI method is implemented in Python using the acoustic wave solver built into Devito, a Python library for optimized computation of finite difference stencils.[16] Each FWI reconstruction took 18 minutes on a MacBook Pro with an M1 chip, and 8 cores with 16 GB of 3228 MHz memory.

## 4. RESULTS

Estimation of the speed of sound maps for all 41 objects in the testing set took approximately 42 seconds using InversionNet and 12 hours using FWI. Figure 2 displays the RMSE and SSIM box plots for the InversionNet and FWI reconstructions over the testing set. Additionally, reconstructions of four selected NBPs one for each BI-RADS breast type are shown in Figure 3. The FWI method still results in better reconstructions than the InversionNet in terms of both RMSE and SSIM. However, InversionNet results present fewer streak artifacts when compared to FWI and sufficient resolution to capture the primary features and structures of the object.

## 5. CONCLUSION

This work presents a deep learning image reconstruction method for ultrasound computed tomography (USCT). In particular, the InversionNet architecture, originally proposed for seismic imaging, was extended to produce

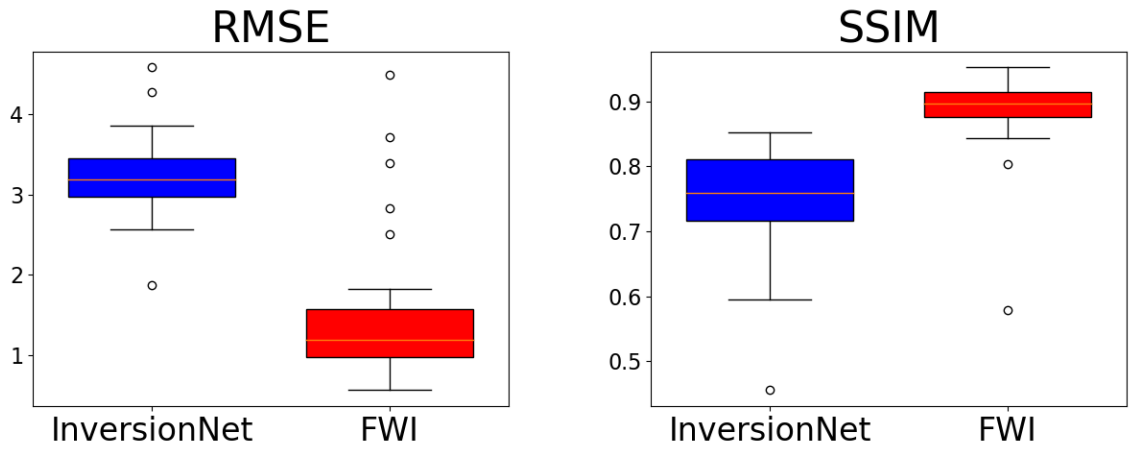


Figure 2. RMSE (left) and SSIM (right) box plots of reconstructed estimates obtained by InversionNet and FWI methods applied to a testing set of 41 NBPs and corresponding USCT data.

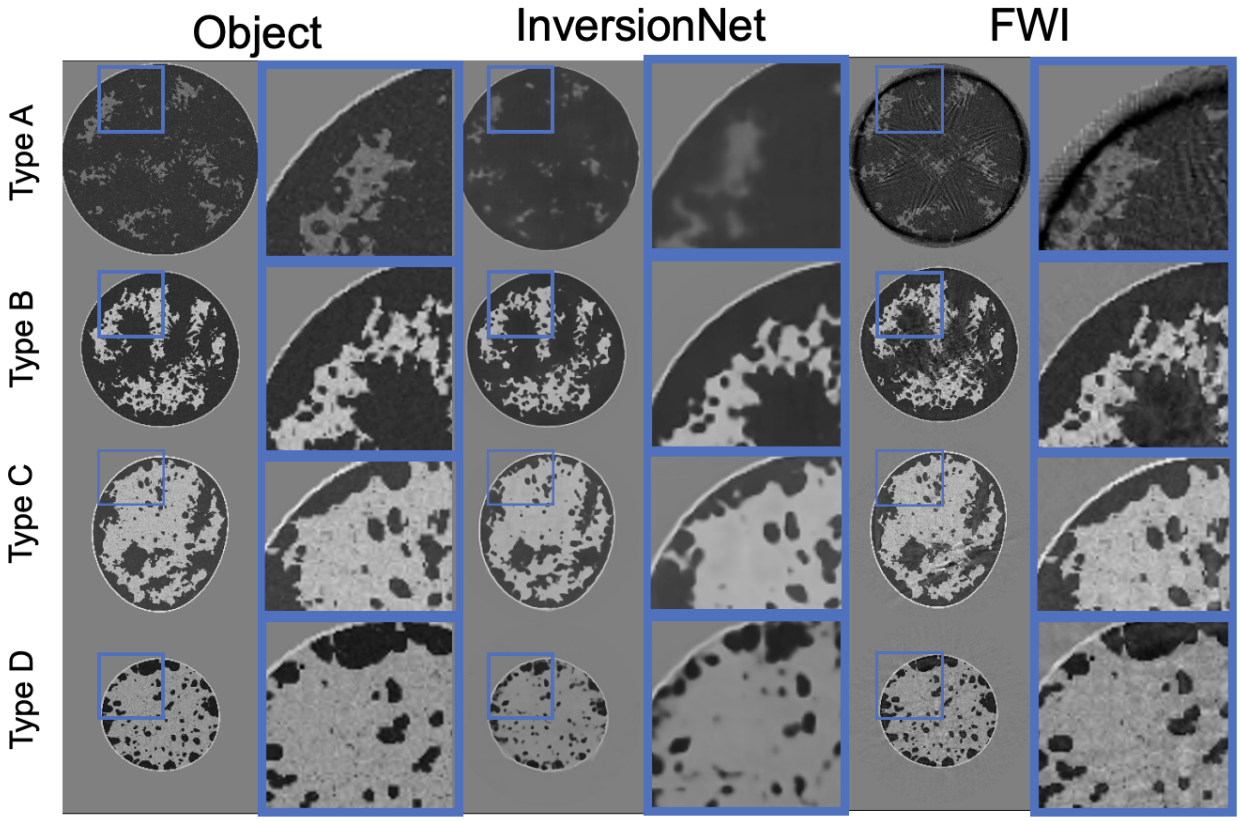


Figure 3. Reconstructions of 4 selected phantoms from the testing set for each BI-RADS breast density type. Type A: Almost all fatty; Type B: Scattered fibroglandular density; Type C: Heterogenous density; Type D: Extremely dense. The speed of sound distributions of the objects are shown in the left column. Reconstructed estimates using InversionNet and FWI methods are shown in the middle and right columns, respectively. Grayscale range is $1400 - 1600$ m/s

quantitatively accurate speed of sound maps of breast tissues from simulated USCT data, without the computational burden of model-based iterative methods, such as full waveform inversion (FWI). The proposed deep learning image reconstruction method was illustrated in a virtual imaging study using a large set of anatomically and physiologically realistic numerical breast phantoms. The numerical results demonstrate that InversionNet is over three orders of magnitude faster than FWI while producing estimates that are only slightly inferior to those produced by FWI in terms of RMSE and SSIM. Notably, speed of sound estimate produced by InversionNet were not affected by any streak artifact, which was present in the FWI estimates.

Future work will compare InversionNet to FWI using task-based measurements of image quality, as well as investigate unsupervised approaches to train InversionNet from clinical USCT measurement data.

## 6. ACKNOWLEDGEMENTS

We would like to thank Dr. Mark Anastasio and Fu Li in the Computational Imaging Science Laboratory at the University of Illinois at Urbana-Champaign for providing us with the anatomically realistic numerical breast phantoms used in the numerical studies.

This work was supported in part by the Center for Space and Earth Science at Los Alamos National Laboratory (LANL) and by the Laboratory Directed Research and Development program of LANL under project number 20210542MFR, and in part by the US National Institutes of Health under grant R01-EB028652.

## REFERENCES

[1] Duric, N., Littrup, P., Poulo, L., Babkin, A., Pevzner, R., Holsapple, E., Rama, O., and Glide, C., "Detection of breast cancer with ultrasound tomography: First results with the computed ultrasound risk evaluation (CURE) prototype," *Medical physics* **34**(2), 773–785 (2007).

[2] Wang, K., Matthews, T., Anis, F., Li, C., Duric, N., and Anastasio, M. A., "Waveform inversion with source encoding for breast sound speed reconstruction in ultrasound computed tomography," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* **62**(3), 475–493 (2015).

[3] Lucka, F., Pérez-Liva, M., Treeby, B. E., and Cox, B. T., "High resolution 3D ultrasonic breast imaging by time-domain full waveform inversion," *Inverse Problems* **38**, 025008 (dec 2021).

[4] Javaherian, A. and Cox, B., "Ray-based inversion accounting for scattering for biomedical ultrasound tomography," *Inverse Problems* **37**(11), 115003 (2021).

[5] Hormati, A., Jovanović, I., Roy, O., and Vetterli, M., "Robust ultrasound travel-time tomography using the bent ray model," in [*Medical Imaging 2010: Ultrasonic Imaging, Tomography, and Therapy*], **7629**, 76290I, International Society for Optics and Photonics (2010).

[6] Wu, Y. and Lin, Y., "InversionNet: An efficient and accurate data-driven full waveform inversion," *IEEE Transactions on Computational Imaging* **6**, 419–433 (2020).

[7] Li, F., Villa, U., Park, S., and Anastasio, M. A., "3-D stochastic numerical breast phantoms for enabling virtual imaging trials of ultrasound computed tomography," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* **69**(1), 135–146 (2022).

[8] Alford, R., Kelly, K., and Boore, D. M., "Accuracy of finite-difference modeling of the acoustic wave equation," *Geophysics* **39**(6), 834–842 (1974).

[9] Virieux, J. and Operto, S., "An overview of full-waveform inversion in exploration geophysics," *Geophysics* **74**(6), WCC1–WCC26 (2009).

[10] Giles, C. L. and Maxwell, T., "Learning, invariance, and generalization in high-order neural networks," *Applied optics* **26**(23), 4972–4978 (1987).

[11] Badano, A., Graff, C. G., Badal, A., Sharma, D., Zeng, R., Samuelson, F. W., Glick, S. J., and Myers, K. J., "Evaluation of digital breast tomosynthesis as replacement of full-field digital mammography using an in silico imaging trial," *JAMA network open* **1**(7), e185474–e185474 (2018).

[12] Li, F., Villa, U., Park, S., and Anastasio, M., "2D acoustic numerical breast phantoms and USCT measurement data: V1." Harvard Dataverse, DOI: 10.7910/DVN/CUFVKE (2021).

[13] Wang, Z., Bovik, A., Sheikh, H., and Simoncelli, E., "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing* **13**(4), 600–612 (2004).

[14] Paszke, A. et al., "Pytorch: An imperative style, high-performance deep learning library," in [*Advances in Neural Information Processing Systems 32*], 8024–8035, Curran Associates, Inc. (2019).

[15] Kingma, D. and Ba, J., "Adam: A method for stochastic optimization," *arXiv e-prints* (2014).
[16] Kukreja, N., Louboutin, M., Vieira, F., Luporini, F., Lange, M., and Gorman, G., "Devito: Automated fast finite difference computation," in [*2016 Sixth International Workshop on Domain-Specific Languages and High-Level Frameworks for High Performance Computing (WOLFHPC)*], 11–19, IEEE (2016).