Ultrasonics 137 (2024) 107179

Contents lists available at ScienceDirect

Ultrasonics

journal homepage: www.elsevier.com/locate/ultras

Investigating pulse-echo sound speed estimation in breast ultrasound with deep learning

Walter A. Simson [b,a,*], Magdalini Paschali [b], Vasiliki Sideri-Lampretsa [c], Nassir Navab [a,d], Jeremy J. Dahl [b]

[a] Chair for Computer Aided Medical Procedures and Augmented Reality, Technical University of Munich, Munich, Germany
[b] Department of Radiology, Stanford University School of Medicine, Stanford, CA, USA
[c] Institute for Artificial Intelligence and Informatics in Medicine, Technical University of Munich, Munich, Germany
[d] Chair for Computer Aided Medical Procedures, Whiting School of Engineering, Johns Hopkins University, Baltimore, MD, USA

ARTICLE INFO

Keywords:
Ultrasound
Sound speed
Simulation
Deep learning
Breast imaging

ABSTRACT

Ultrasound is an adjunct tool to mammography that can quickly and safely aid physicians in diagnosing breast abnormalities. Clinical ultrasound often assumes a constant sound speed to form diagnostic B-mode images. However, the components of breast tissue, such as glandular tissue, fat, and lesions, differ in sound speed. Given a constant sound speed assumption, these differences can degrade the quality of reconstructed images via phase aberration. Sound speed images can be a powerful tool for improving image quality and identifying diseases if properly estimated. To this end, we propose a supervised deep-learning approach for sound speed estimation from analytic ultrasound signals. We develop a large-scale simulated ultrasound dataset that generates representative breast tissue samples by modeling breast gland, skin, and lesions with varying echogenicity and sound speed. We adopt a fully convolutional neural network architecture trained on a simulated dataset to produce an estimated sound speed map. The simulated tissue is interrogated with a plane wave transmit sequence, and the complex-value reconstructed images are used as input for the convolutional network. The network is trained on the sound speed distribution map of the simulated data, and the trained model can estimate sound speed given reconstructed pulse-echo signals. We further incorporate thermal noise augmentation during training to enhance model robustness to artifacts found in real ultrasound data. To highlight the ability of our model to provide accurate sound speed estimations, we evaluate it on simulated, phantom, and in-vivo breast ultrasound data.

1. Introduction

Breast cancer is the most common cancer in women, with 2.3 million cases in 2020 [1]. Ultrasound is used to examine patients with indeterminate lesions as an adjunct to mammography. Sonography quickly and safely aids clinicians in differentiating cysts from solid lesions but lacks high specificity in discerning benign and malignant findings [2].

Research has shown that different types of breast tissue, for instance, breast gland and malignant lesions, can vary substantially in their sound speed by up to 100 m/s [3]. However, in current diagnostic ultrasound imaging systems, a constant sound speed (e.g., 1540 m/s) is assumed for all tissues to create B-mode images [4]. Since breast tissue is heterogeneous in its sound speed, this assumption can degrade image quality via phase aberration and reduce the efficacy of ultrasound imaging in diagnosing breast cancer [5]. Thus, developing a technique that can accurately estimate the sound speed of tissues can substantially aid diagnostics by enabling active phase aberration correction [5] as well as offering a quantitative diagnostic metric to physicians to aid in diagnosis [3,6].

Sound speed estimation in pulse-echo ultrasound imaging remains a challenging problem. Unlike total tomographic sound speed reconstruction, which uses a complete angular sampling from $[-\pi, \pi]$ and a known distance between transmitter and receiver [7–9], pulse-echo sound speed estimation utilizes a limited angular sampling and lacks accurate position or distance information, thereby complicating the sound speed estimation. An accurate pulse-echo sound speed estimation

* Corresponding author at: Department of Radiology, Stanford University School of Medicine, Stanford, CA, USA.
E-mail address: waltersimson@stanford.edu (W.A. Simson).

https://doi.org/10.1016/j.ultras.2023.107179
Received 2 May 2023; Received in revised form 30 September 2023; Accepted 7 October 2023
Available online 29 October 2023

would greatly increase the applicability and utility of such methods in clinical practice where single-sided access is common.

2. Related work and contributions

Pulse-echo sound speed estimation can be performed with a physical model or with a data-driven machine learning model. Either way, radio frequency (RF), the complex analytic signal, or demodulated in-phase and quadrature (IQ) data can be input into the estimation model. The output is a predicted spatial sound speed distribution in the medium of interest.

Anderson and Trahey proposed a method by which a global average sound speed between the transmitting interface and a focal point is derived from amplitude features extracted from the channel data of a focused transmit [10]. The method was experimentally validated on homogeneous phantoms with wire and speckle-generating targets. Building on the work of Anderson and Trahey, Jakovljevic et al. proposed a model that estimated local sound speed along a wave propagation path from a sequential series of global average sound speed measurements at discretized depths [11]. The model was shown to work when the medium was composed of layers with different sound speeds. Ali and Dahl proposed the IMPACT method that was better able to estimate local sound speed in volumes with lateral inhomogeneities by tomographically maximizing the coherence factor along with an innovative phase aberration term given reconstructions with sound speeds ranging from 1400 to 1700 m/s [12]. Jaeger et al. proposed CUTE, a tomographic pulse-echo model for slowness (inverse sound speed) estimation given phase-shift measurements, which was solved for convenience in the spatial frequency domain [13]. This work displayed initial tomographic contrast results for sound speed imaging on in-silico phantoms, demonstrating the potential of pulse-echo sound speed reconstruction as an imaging modality. Sanabria et al. work towards solving this inverse problem directly in the spatial domain without the use of spatial frequencies with a novel anisotropically-weighted total-variation regularization method and displayed high-resolution sound speed reconstructions for accurate quantitative time of flight measurements [6]. The sound speed reconstruction method [14] improved upon the 2015 CUTE method by solving for sound speed maps using a spatial domain model relating phase shift measurements between transmit and receive angles about a common mid-angle to the slowness along a ray. This new model also corrected the erroneous position of the echos in the reconstruction and created accurate sound speed maps in a series of phantoms and a liver model. Later work extended the model to curvilinear probe geometries [15]. Recently, [16] modeled the ultrasound image reconstruction process in an auto-differentiable pipeline and optimized a slowness map to minimize a common mid-point phase shift metric, allowing for joint sound speed estimation and b-mode auto-focusing. Though the above-mentioned methods have greatly contributed to sound speed estimation, the task remains challenging and could be further improved to achieve clinical adoption.

Recently, there has been growing interest in sound speed estimation methods built upon the advances in computer vision with the advent of performant neural networks as universal estimators. Feigin et al. proposed pulse-echo sound speed estimation [17] with a deep neural network based on VGG [18]. The network was trained to map the RF data from three-plane wave transmits from three discrete apertures to a sound speed distribution. Training data was generated by simulating ultrasonic interrogations of media containing randomly positioned ellipses of varying ultrasonic properties with the k-Wave software package [19]. Each plane wave used a separate 64-element sub-aperture of a 128-element transducer to interrogate the medium at a pre-defined angle. Bernhardt et al. presented a novel approach for reconstructing speed-of-sound (SoS) images for breast cancer imaging using Variational Networks (VN) [20]. The proposed method incorporates simulations with varying complexity into training, utilizes loop unrolling of gradient descent with momentum, and regularizes training using exponentially weighted loss and smooth activation functions. The trained network was evaluated on a simulated test set and achieved a mean absolute error of 12.5 ± 16.1 m/s. When applying the method to in-vivo data, the resulting estimated values were outside the training data range and the normal envelope of healthy human tissue, indicating that the model could be further improved to generalize to the large domain shift between the simulation and in-vivo data. A new investigation on the use of deep learning for ultrasound sound speed reconstruction was presented by [21], in which the multi-input network architecture of [17] is extended to map IQ data with separate I and Q branches to sound speed distribution maps. The k-Wave suite was again used to simulate random ellipses in media, but angular coherence was not considered as only one plane wave transmission was simulated. The evaluation showed accurate sound speed estimations, but the method was evaluated solely on simulated data similar to the training set data and did not incorporate features commonly observed in real-world ultrasound signals, such as thermal noise.

For supervised deep learning, there is no accurate way to manually annotate ultrasound signals with local sound speed information. Therefore, full-wave simulations are commonly used to create paired data of known sound speed distributions with their respective channel data. This requires the simulations to be carefully parameterized to accurately model transducer characteristics and tissue property distributions. Previously, in-silico phantoms have been generated with randomly positioned ellipses of varying ultrasonic properties to create phantoms that are randomly parameterizable [17,21,22]. These phantoms are fully parameterizable with geometric shapes but do not incorporate the target anatomy's anatomical geometries or tissue properties. Others have sampled a virtual breast model with a combination of image-based landmark structures based on MRI data from the NIH Visible Human Project and randomly distributed structures [23], or evaluated their hybrid ultrasound simulation approach segmented MRI volumes of patient data for intra-operative registration [24]. While this allows modeling anatomical geometries, MRIs as anatomical priors are limited in scale, resolution, availability, and quality and cannot be randomly generated. Breast phantoms have been extensively explored for X-ray modeling [25,26].

2.1. Our contributions

This study proposes a pipeline for sound speed estimation of medical ultrasound for breast tissue. We create a large-scale synthetic ultrasound dataset of breast ultrasound images, including breast gland, skin, and two types of lesions, which have been accurately parameterized with tissue properties from literature. Next, a Deep Neural Network (DNN) is trained on the complex analytic signal of synthetic aperture data to estimate sound speed. We show the capabilities of the proposed method with an evaluation of simulated, phantom, and in-vivo data. Our contributions are:

- A parameterizable dataset of anatomical breast plane wave ultrasound simulation input data
- An adapted fully convolutional DNN trained on beamformed complex analytic signal to estimate the spatial distribution of sound speed
- Quantitative DNN estimates on phantom and in-vivo data consistent with traditional sound speed measurements
- Evaluation of temporal estimation consistency that displays invariance to artifacts such as thermal noise on sequential ultrasound measurements

Importantly, this work is meant to show the viability of using a neural network to infer sound speed given a small number of angled plane-wave acquisitions from the same aperture when simulated from in-silico phantoms. An exhaustive evaluation and comparison of network architectures and methods are left to future works.

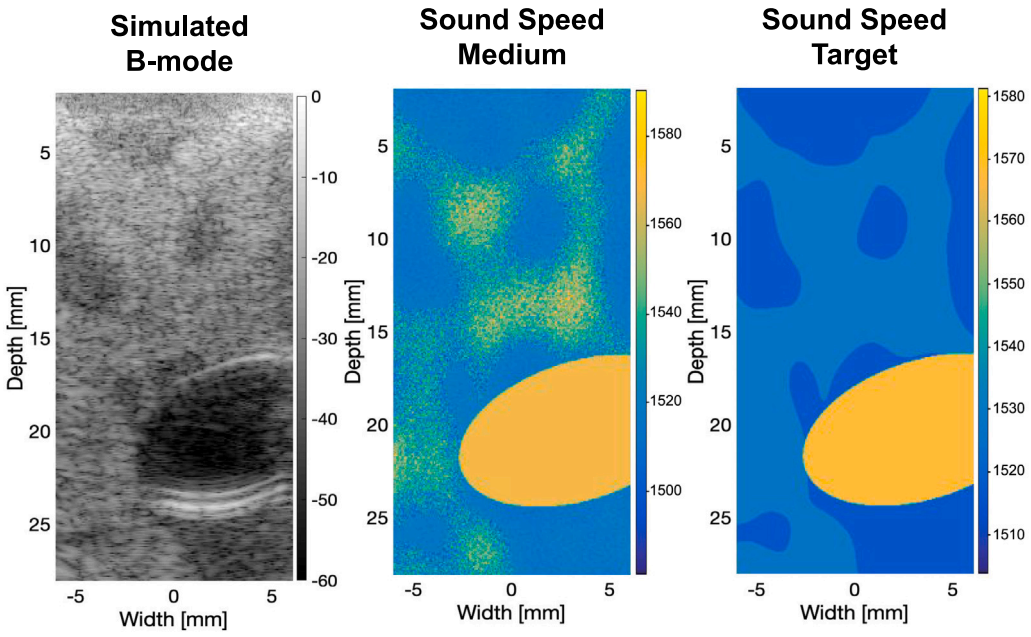

Fig. 1. (a) Simulated ultrasound B-mode image with background approximating glandular breast tissue and an anechoic cyst [2]. (b) Sound speed of the simulated medium. Regions of brighter diffuse scattering display a higher standard deviation of the sound speed map. (c) Sound speed target used for model optimization with region-average sound speed values. Note that two sound speed values are used for the background, and a single sound speed value is used for the cyst.

3. Methods

The following section describes the methodology used to train a DNN on 3D in-silico breast simulations for generalizable sound speed estimation with real transducer data. To this end, we will cover the generation and parameterization of in-silico breast phantom, the simulation process using the k-Wave suite, the proposed data processing and augmentation steps for training a DNN, and the architecture and data flow of the DNN. The methods and notation are formally described in this section, while the selected parameter values are presented in Section 4.

3.1. In-silico phantom

The three-dimensional ultrasound simulations developed in this work are generated to model human breast tissue and comprise three basic elements. A semantic label map defines the spatial distribution of anatomies in the domain, such as skin, lesion, cyst, breast gland, and fat. A scatter distribution field defines the location and relative intensity of all scatterers based on the semantic map. Lastly, a tissue model for a given semantic label and scatterer intensity defines a voxel-wise sound speed, density, non-linearity (B/A), and attenuation in the simulation domain. These elements' combination and random parameterization generate a large heterogeneous dataset for sound speed estimation.

The in-silico phantom domain is described by a Cartesian grid of i points p_i where $p_i \in X \times Y \times Z$. In this context, $X = \{0, x, \dots, x_d\}$, $Y = \{0, y, \dots, y_d\}$, and Z is defined as $\{0, z, \dots, z_d\}$. In this representation, x, y, and z correspond to the spatial resolution of the grid, while x_d, y_d, and z_d represent the dimensions of the grid in their respective directions.

3.1.1. Semantic label map

The in-silico simulations include regions of skin, glandular tissue, fat, breast cysts, and breast lesions resembling fibroadenomas [2]. Every voxel is assigned a semantic label to model an anatomical spatial distribution. A semantic map of the breast gland and fat is modeled by a 2D Gaussian random field (GRF) that mimics the spatial property variation of the breast gland and fat. A 2D Gaussian filter g of the size (x_f, y_f) is defined as $g(u,v) = \frac{1}{2\pi\sigma^2} e^{-\frac{u^2+v^2}{2\sigma^2}}$ where $u \in [-\frac{x_f}{2}, \frac{x_f}{2}]$ and $v \in [\frac{-y_f}{2}, \frac{y_f}{2}]$ and is convolved with a 2D random field of size $F = [0, x_d + x_f] \times [0, y_d + y_f]$ where the field elements $F_i \sim \mathcal{U}_{[0,1]}$. The resulting GRF is then normalized ($\mu = 0$) and scaled to the interval $[-0.5, 0.5]$, and two sub-regions of fatty and glandular tissue via a randomly thresholded to create a random binary map of breast gland regions. Values below the random threshold are assigned to the fatty tissue (low echogenicity), and all above are assigned to the glandular tissue (high echogenicity).

Cysts and lesions are modeled by elliptical inclusions where cysts are anechoic, and lesions display positive or negative echogenicity. The cyst and lesion masks are defined as an ellipse E in space projected onto the aforementioned Cartesian grid and randomly parameterized by the position of its center $(x_c, y_c) \in [X \times Y]$, the lengths of its radii $r_{i=\{1,2\}} \in \mathbb{R} \quad \forall \quad r_i < min(x_d, y_d)$, and an orientation angle $\theta \in [0, \pi]$. The ellipse is projected along the elevational plane to generate a 3D inclusion. The skin anatomy is defined as a linear mask at the top of the domain.

Six combinations of the above-mentioned tissue classes are defined: cyst with skin, lesion with skin, skin, breast tissue (gland and fat), lesion, and cyst. All classes are created over a breast tissue background, which represents a phantom without any other anatomical structures.

3.1.2. Scatterer distribution

The unit-less discretized scatterer density ρ_s defines the fraction of all voxels in a region labeled as scatterers for a semantic class. The discretized scatterer density is defined as:

$$\rho_s = \frac{n_s}{\lambda^3} \cdot x \cdot y \cdot z, \rho_s \in [0, 1],$$

where n_s is the number of scatterers in an imaging resolution voxel (IRV). The size of the IRV in 3D is approximated as λ^3 where λ is the wavelength of the transmitted pulse.

The scatterer intensity distribution $S(x, y, z)$, i.e., how strongly the acoustic properties of a scatterer at a given location deviates from the average, is modeled by a uniform distribution $U \sim \mathcal{U}_{[-0.5,0.5]}$ of intensity for every scatterer in the domain [27]. This distribution is jointly sampled by a Bernoulli distribution $B \sim \mathcal{B}(1, \rho_s)$ to determine

Table 1
Mean sound speed range and scatter contrast per class used for our breast ultrasound dataset simulation.

Tissue class	Mean sound speed range	Scatter contrast
Cyst [3]	[1500, 1620]	–
Lesion [3]	[1488, 1512]	± 10–30 dB
Skin	[1540, 1670]	10 dB
Breast Gland [3]	[1480, 1528]	12 dB
Fat	[1480, 1528]	0 dB

the location of the scatterers in the 3D grid.

$$S(x,y,z) = T(x,y,z) \cdot \sigma_0(x,y,z) + \mu_0(x,y,z), \text{ where}$$

$$T(x,y,z) = \begin{cases} U(x,y,z), & B(x,y,z) = 1 \\ 0, & \text{otherwise} \end{cases}$$

Every point in $S(x,y,z)$ is assigned a sample value from the distribution $T(x,y,z)$, which is subsequently scaled by the variance and mean property value $\sigma_0(x,y,z)$ and $\mu_0(x,y,z)$ to create a spatial white noise distribution for both sound speed and density. The Bernoulli distribution represents the likelihood that a given point in 3D space is a scatterer and is parameterized by the scatterer density ρ_s. By jointly sampling these two distributions, a final scatterer distribution $S(x,y,z)$ is characterized by randomly located discretized scatterers with random intensities.

3.1.3. Tissue model

The tissue model for this work is a semantic lookup table of tissue properties belonging to a semantic class. These include sound speed, density, non-linearity (B/A), attenuation and scatterer contrast. For each semantic class, tissue properties are assigned to the respective spatial location in the k-Wave simulation.

First, a random mean sound speed is selected for breast fat and gland classes as given in Table 2 following [3]. Using the spatial scatterer distribution from above, μ_0 is set to the mean sound speed. For the breast fat and gland sound speed regions, the μ_0 is smoothly modulated with the convolved GRF used to generate the semantic class to ensure sound speed heterogeneity. The scatterer amplitude σ_0 is based on the desired relative brightness (c.f. Table 2). For each class, the desired sound speed can be assigned element-wise. The scatterer distribution scales this sound speed map to create diffuse scattering in the simulation. As shown empirically in [28], the values of sound speed and density can be approximated to be linear, and the density map is set to be proportional to the sound speed map by a factor of α_ρ. To increase variability in the simulation and reduce the linear relationship between sound speed and contrast, α_ρ is randomly scaled when generating the training simulations. The attenuation and non-linearity values are constant, as listed in Table 2.

Lastly, the in-silico phantom sound speed map in Fig. 1(a) is averaged for a given semantic label to form a target average sound speed map, as shown in Fig. 1(c), suitable for training our deep model. The averaged sound speed map is used only as a training label and not for the k-Wave simulation. The original in-silico phantoms are then utilized in k-Wave to generate simulated RF channel signals from pulse-echo ultrasound.

3.2. Data processing

Neural networks perform best when the dataset they are trained on is statistically representative of the dataset they will see at "inference time" or when they are deployed. Therefore, a large heterogeneous dataset is desirable to train robust neural networks. To increase the heterogeneity of the training data, data augmentation is often applied. In computer vision, these augmentations can include rotations, flips, and deformations of the training data. Augmentation is applied with a likelihood at train time and does not increase the number of training samples but rather randomly modifies the training samples. Each augmentation is randomly parameterized at every invocation.

Building on [29], we propose using ultrasound-specific augmentation techniques. We propose augmenting with filtering the simulated RF data with the transducer's impulse response given a random relative bandwidth to make the network robust to transducer sensitivity. We further augment the channel data with broadband thermal noise via thermal noise augmentation (TNA) [30]. Thermal noise is an artifact resulting from electronic noise in ultrasound devices but is missing from ultrasound simulations [31]. TNA is performed by adding white thermal noise to channel data with an augmentation probability p_{TNA}. Given the original RF signal $s(t)$, and random white thermal noise $n(t)$ the received signal $s_{TNA}(t)$ is given by:

$$s_{TNA}(t) = s(t) + n(t)$$

TNA is added on top of the clutter and aberration noise generated by the forward process of the k-Wave simulation.

TNA is parameterized by an upper and lower bound in noise amplitude relative to the transmit signal's Root Mean Square (RMS). This uniform parameterization distribution is randomly sampled via the method proposed by [31]. The constant TNA and attenuating tissue model lead to a realistic reduction of SNR over depth.

The complex analytic signal is generated via the Hilbert transform, and each plane wave is beamformed individually via dynamic receive beamforming with an assumed sound speed $c_0 = 1540$ m/s to generate complex beamformed images similar to [14]. Lastly, the complex components of the analytic signal from the same spatial location are mapped to the channel dimension of the convolutional neural network, which is different than the approach taken in [17,22].

3.3. Network architecture

We modify a deep, a fully convolutional neural network F based on DenseNet [32] with modifications from [33,34] described below. DenseNet consists of an encoder and decoder and is comprised of stacked dense blocks connected via 2D max pooling and unpooling blocks, respectively. Every dense block consists of three convolutional layers, the first two of which have a kernel size of 5×5 with stride 1 and the third one a kernel size of 1×1 and stride 1. Skip connections [35] are added between each dense block of the encoder and decoder to prevent vanishing gradients, enhance network trainability, and maintain feature quality.

The base network is modified to take, as input, three complex beamformed images of a medium (one for each angled plane wave transmission) and output an estimated sound speed map of the medium defined as

$$F : \mathbb{C}^{N \times M} \mapsto \mathbb{R}^{N \times M},$$

for an image size of $N \times M$ pixels. This network consists of a multi-head three input dense blocks (one for each angled plane wave), a bottleneck, and four decoder dense blocks that output the model sound speed estimation. The overall architecture can be seen in Fig. 2. Unlike previous work where the real signal (RF data) is passed to the network, each head is passed the complex analytical signal in this work. This preprocessing is performed to extract phase information for the network via the well-defined Hilbert transform so that this operation does not have to be performed by network weights. The separate processing of each plane wave ensures the extraction of robust features, such as phase shift and spatial coherence. A further modification of [32] is the concatenation along the channel dimension and 1×1 convolutional layer to reduce data dimensionality. This modification merges features from the first encoder layers for downstream comparison. In this work, we replace ReLu activations with PReLu [33], which has been shown to improve model fitting and reduce the risk of overfitting. Furthermore, batch normalization layers are replaced with instance

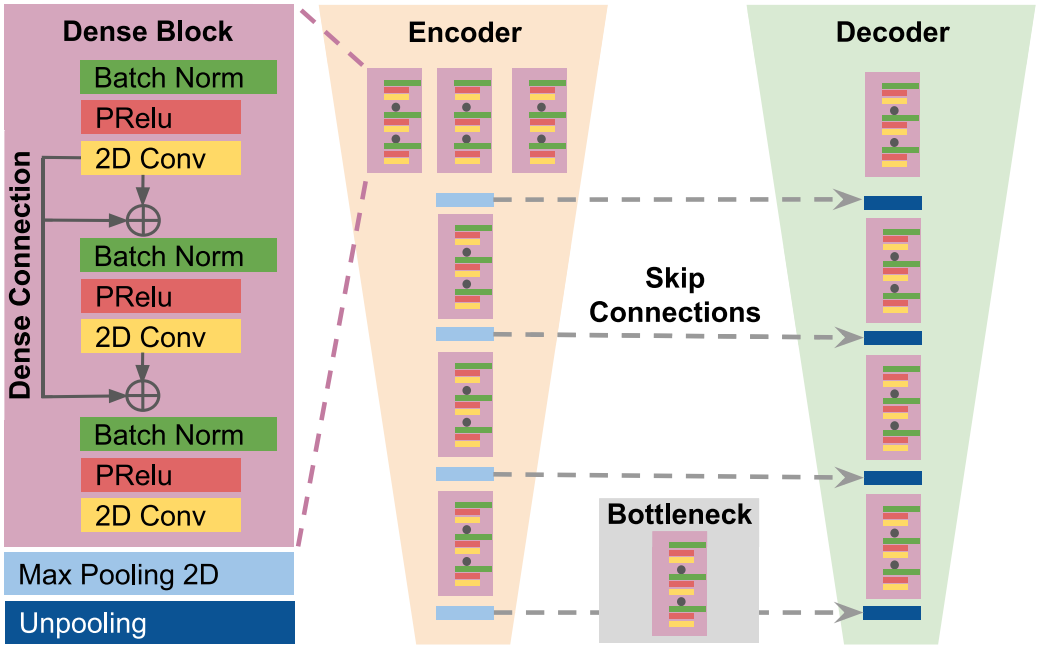

Fig. 2. Overview of the proposed architecture. Our model is composed of an encoder that individually processes three complex beamformed images, whose features are concatenated after being passed through individual dense blocks, a bottleneck, and a decoder that utilizes unpooling and produces the sound speed estimations. Dense skip connections are used within each dense block, and long-term skip connections are placed between encoder and decoder to enhance gradient flow and maintain feature quality.

normalization [34], which has also been shown to enhance training dynamics in noise-sensitive applications.

The Mean Square Error (MSE) between the estimated and target sound speed map is used as the loss function. We use weight decay with L2 regularization to avoid overfitting and maintain weight sparsity [36].

4. Experimental setup

4.1. In-silico simulations

The in-silico tissue models described in Section 3.1 are parameterized with the values for sound speed, non-linearity (B/A), density, and attenuation as listed in Table 1. The skin depth is parameterized for varying thickness within anatomical norms of 0.7 to 3 mm [37]. The k-Wave simulation parameters are summarized in Table 2. The Gaussian filter for the breast tissue generation uses the parameters $x_f = y_f = 400$ and $\sigma = 600$. In this work, the density ratio α_ρ is set to be $1.5 \pm 10\%$, i.e. $[1.35, 1.65]$ uniformly sampled. The transducer modeled for the simulations is the Cephasonics CPLA12875 (Cephasonics Ultrasound Solutions, Santa Clara, California, USA) with 128 elements with a total aperture width of 37.5 mm, an element height of 7 mm, an element width of 0.293 mm, and a kerf of 0 mm between elements. The transmit frequency is set to 5 MHz with transmit duration of one tone-burst cycle. A medium sound speed of 1540 m/s is used to calculate the transmit delays for steering angles of −8, 0, and 8 degrees for each plane wave transmission. All three transmissions are simulated from the same aperture of the center 64 elements, as is more commonly the case in plane wave imaging pulse sequences for coherent compounding. On both transmit and receive, rectangular apodization is employed. The medium dimensions in grid points are $N_x = 548$, $N_y = 648$, and $N_z = 126$, with a grid spacing of 58.594 µm in all directions such that five grid points could fit laterally within one modeled piezo element. The simulation domain's total dimensions (x_d, y_d, z_d) are 32 mm × 38 mm × 7.4 mm. The time step size δt of the simulation is 11.41 ns, leading to a simulation sampling frequency of 87.6 MHz. A Perfectly Matched Layer (PML) of 7 × 17 × 9 grid points is added to the medium to prevent signal wraparound [19]. The modeled transducer is centered on top of the phantom grid. In total, 5996 samples consisting of three plane wave simulations are generated using the k-Wave Toolbox [19], and the C++ accelerated binary on an NVIDIA Quadro RTX 6000 GPU with 64 CPU threads. The expected GPU run time per simulation is 620 s, 43 days 38 min, and 40 s for the entire dataset.

Table 2
Simulation Properties.

Property	Value
Transmit Frequency	5 Mhz ± 10%
Center Frequency	5 Mhz
Density Ratio	1.5% ± 10%
Alpha Coeff	0.75 dB/MHz cm
Alpha Power	1.5
B/A	6
Bandwidth	60%
Tone Burst Cycles	1
Sampling Frequency	87.6 Mhz
# Elements	128 elements
Pitch	293 µm
Kerf	0 µm

4.2. Data processing parameters

The simulated channel data is resampled from 87.6 MHz to 40 MHz. A Gaussian band-pass filter centered at 5 MHz with a fractional bandwidth uniformly sampled between 50% and 90% is applied to model the transducer's impulse response. TNA was performed with a magnitude range from −120 dB to −80 dB relative to the ballistic pulse and an augmentation probability of $p_{TNA} = 20\%$. A t_0 was set to 2.75 µs for the center transmission and 5.0 µs for the ±8 degree plane waves.

4.3. Network training

Our deep model is trained with a batch size of 6, for 138 epochs and with a learning rate of 0.001 with early stopping based on the loss of the simulated validation set. The Adam optimizer [38] is used with weight decay activated with a decay rate of e^{-4}. The network

Table 3
Sample distribution for the simulated Training and Validation sets.

	Cyst & Skin	Lesion & Skin	Skin	Breast Gland	Lesion	Cyst	Total
Training	915	888	913	971	849	946	5482
Validation	82	85	92	83	75	97	514
All	997	973	1005	1054	924	1043	5996

is programmed in Python using the PyTorch Library v1.7 [39] and the PyTorch Lightning framework v1.2.10 [40] and Weights and Biases [41] for experimental tracking. Our models are trained on an NVIDIA Quadro RTX 6000 GPU.

4.4. Datasets

The proposed simulation and network training method is evaluated on three separate datasets. The in-silico simulation dataset is the basis for a training and validation set. A classwise breakdown can be seen in Table 3. As an independent and out-of-distribution test set, we utilized channel data collected from a real transducer on ex-vivo tissue samples and calibrated phantoms. Lastly, in-vivo measurements were performed to evaluate the performance of the proposed pipeline to domain shift.

4.4.1. Simulation evaluation

Our model is evaluated on a simulated validation set of 514 samples equally drawn from all classes. We report the Mean Absolute Error (MAE) between the predicted and target sound speed for each class. Furthermore, to showcase the importance of TNA, we compare the error distributions over all classes for two otherwise identical models trained with and without TNA. Lastly, we further investigate the effect of thermal noise on the models by comparing the error over depth for three levels of additive thermal noise and our baseline without noise.

Moreover, we compare the proposed architecture with and without the dense skip connections to evaluate the benefit of added dense skip connections as proposed by [32]. Finally, we compare the proposed model against [17], which utilized a fully convolutional architecture with 3.9M trainable parameters without skip connections. The model and experimental setup were replicated following [17].

4.4.2. Phantom evaluation

To evaluate the predictive efficacy of our model, phantom and in-vivo studies are performed using a Cephasonics Griffin with 64 channels and a CPLA12875 transducer. The sound speed of a homogeneous CIRS Phantom Model 040GSE (CIRS Inc, Norfolk, VA USA) is calibrated via the speckle brightness method [42] following the approaches in Ali et al. [12] and Hyun et al. [43] due to the phantom age to ensure the phantom sound speed is consistent with the factory-specified sound speed. The sound speed in the phantom is determined to be 1558 m/s using this method.

Next, a bovine steak is prepared, and its sound speed is measured to be 1566 m/s in a distilled water bath (water temperature 24.6°C, 1495.8 m/s [44]) using the method described in [45]. The steak is cut into two slices of 8 mm and 4 mm thickness and stacked on the CIRS phantom to create a two-layered model. The regional mean sound speed error is estimated for regions of interest (ROI) in the steak and proximal and distal locations in the CIRS phantom. The proximal and distal regions are differentiated to showcase the effect of depth-dependent SNR on the model predictions. Furthermore, to reduce selection bias and evaluate the temporal consistency of our model, the regional sound speed estimates are averaged over 100 consecutive static frame measurements to quantify the influence of thermal noise on the sound speed estimations and stability of the estimates. In total, two sets of phantom data were collected for evaluation, each containing 100 frames. Each frame consists of three plane wave transmits.

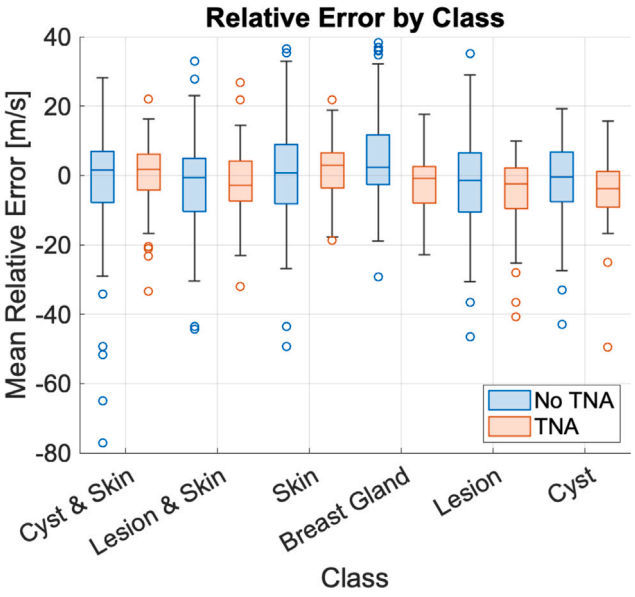

Fig. 3. Boxplot comparing sound speed relative estimation error distributions per class for the simulated validation set. The central mark on the box indicates the median value, and the top and bottom edges of the box indicate the quartile range. The black whiskers indicate the extent of the distribution without the outliers, denoted by circles on the plot. Overall, the model trained with TNA achieves a lower relative error standard deviation and fewer outliers for all classes.

4.4.3. In-vivo demonstration

In-vivo imaging is performed on the left breast of a healthy volunteer (Age: 28, BMI: 22.4) in three regions. The volunteer was selected and provided written informed consent under a protocol approved by an ethical committee from the Technical University of Munich. Channel data for each region are acquired with the same configuration as for the phantom experiments. In total, three in-vivo datasets were collected from a healthy volunteer for evaluation, each containing 100 frames. Each frame consists of 3 plane waves.

5. Results

5.1. Validation set evaluation

Table 4 shows the classwise MAE on the validation set for a baseline comparison against [17], an ablation study of dense skip connections from [32] as well as the proposed network trained with and without TNA. The model from [17] trained on our simulated dataset has an overall error of 62.0 ± 13.2 m/s, while the proposed and ablation study models perform better in all classes of the validation set. This comparison highlighted the impact of skip connections and TNA in our proposed model.

Table 4 further shows the proposed model trained with and without TNA displayed a small classwise MAE relative to the wide sound speed ranges on which the model is trained. The skin class displayed an MAE of 8.50 m/s for a model trained with TNA and 16.40 m/s for the base lesion class for the model trained without TNA. Overall TNA substantially improves estimation error across the classes by 2.2 m/s for the cyst class to 5.5 m/s for the skin and cyst class. The model trained with TNA displayed the lowest overall MAE of 10.3 ± 5.60.

Table 4
Sound speed estimation MAE and standard deviation per class for various models. Estimations are shown in m/s.

Class	Feigin et al. [17]	No dense skips No TNA [35]	Proposed No TNA	Proposed TNA
Cyst & Skin	63.5 ± 12.6	15.1 ± 10.7	16.1 ± 12.4	10.6 ± 5.10
Lesion & Skin	62.5 ± 12.1	17.0 ± 10.1	15.5 ± 8.70	12.0 ± 5.70
Skin	60.5 ± 13.6	14.5 ± 14.3	12.9 ± 9.10	8.50 ± 4.00
Breast Gland	61.1 ± 14.8	12.9 ± 8.34	12.8 ± 8.84	7.90 ± 3.70
Lesion	61.3 ± 13.1	18.1 ± 13.0	16.4 ± 8.70	12.7 ± 7.30
Cyst	62.9 ± 12.9	13.3 ± 7.38	12.8 ± 6.30	10.6 ± 5.90
Overall	62.0 ± 13.2	15.1 ± 11.0	14.3 ± 9.20	10.3 ± 5.60

Fig. 3 shows the relative average error for each class for models trained with and without TNA. Though the performance of both DNNs is good, TNA contributes toward reducing the standard deviation (signified by box size) of the relative error and the number of outliers (signified by circles).

Effect of thermal noise over depth

In Fig. 4, we show the effect of additive thermal noise over transmission depth on the predictions of networks trained with and without TNA. We select three scales of additive noise, specifically −80 dB, −100 dB, and −120 dB relative to the transmit signal RMS, along with a baseline measurement without noise. First, it can be seen that the network trained with TNA (Bottom) is robust to thermal noise since the error remains low over the entire transmission depth of the measurements. The network trained without TNA (Top) is severely affected by thermal noise present in the channel signals for all noise levels, with increasing error from −120 dB to −80 dB. The baseline sound speed error shows that the network trained without TNA is able to accurately estimate the sound speed over transmission depth on the validation set when no noise is added. For noise levels −120 and −100 dB, the model trained without TNA underestimates the sound speed in the medium. For the noise level of −80 dB, the model underestimates to a depth of 1.6 mm and then overestimates the sound speed in the medium. The network trained with TNA only marginally underestimates the medium sound speed after 2 mm depth.

Qualitative evaluation

Qualitative results of simulated B-modes for all classes and their respective sound speed estimates are shown in Fig. 5. Our proposed simulation pipeline creates B-mode images with strong similarity to real-world breast-tissue B-modes. Consistent with the results shown above, our model is able to successfully estimate sound speed distributions throughout the simulated domain for all data classes. Contours of both anechoic and echogenic features in the images are successfully recovered, and the sound speeds within the regions are correctly estimated. In classes with cysts, reverberation artifacts are often present on the boundaries of the simulated cysts. Nonetheless, the model can generate accurate sound speed estimates successfully.

5.2. CIRS phantom evaluation

In order to evaluate the generalization ability of our model, we evaluate its performance on a layered phantom that was not represented in the training set. One hundred ultrasound frames are acquired with a real transducer, and the results for the layered phantoms with both a 4 mm and 8 mm bovine steak phantom are shown in Table 5. As stated in Section 4.4.2, the measured sound speed of the steak is 1566 m/s, and that of the CIRS phantom is 1558 m/s. The evaluation of the model estimation is performed in three discrete ROIs that extend across the entire image aperture and are depicted by red, yellow, and green boxes in Fig. 6 in order to evaluate the influence of thermal noise and attenuation along with other real-world factors on the model's estimation at varying depths.

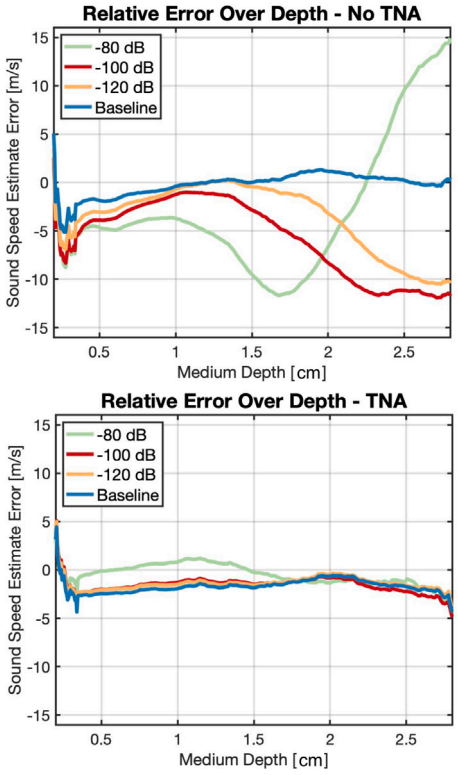

Fig. 4. Relative sound speed estimation error over depth for the simulated validation set for three levels of additive thermal noise and baselines without added noise. Our model trained with TNA (Bottom) is substantially more robust to the addition of thermal noise. In contrast, the network trained without TNA (Top) is more noise-sensitive, and its performance substantially decreases with decreased SNR over depth.

Even though our model is solely trained on simulated breast ultrasound signals, it is still able to infer the sound speed of these two-layered phantoms in agreement with in-vitro sound speed measurement. The mean error for the steak layers ranges from 1.60 m/s for the 4 mm steak to 1.40 m/s for the 8 mm one. Furthermore, the sound speed for the top ROI of the CIRS phantom is also successfully estimated with a mean error of 2.40 m/s for the 4 mm steak and 0.90 m/s for the 8 mm one. As seen in Table 5, the standard deviation values of the estimations among the 100 consecutive frames are also low, ranging from 2.70 m/s to 6.90 m/s, showcasing the temporal consistency of the model predictions for real-world data.

Finally, we can see that the prediction for the bottom 2.9 mm of the CIRS phantom (green ROI) has a larger error than the top, ranging from 13.90 m/s for the 4 mm steak phantom to 15.30 m/s for the 8 mm layer steak phantom. The total range of the 8 mm steak phantom is 1516.20–1653.70 m/s and 1518.90–1645.90 m/s for the 4 mm steak phantom. The prediction of the model trained without TNA for the bottom region of the CIRS phantom is 1536.70 m/s ± 4.72 m/s for the 4 mm steak phantom and 1510.70 m/s ± 8.11 m/s for the 8 mm steak phantom. This improvement in accuracy and reduced standard deviation shows model superiority when trained with TNA, which decreases the error from 29.30 m/s to 13.90 for the 4 mm steak phantom and from 55.30 m/s to 15.30 m/s for the 8 mm steak phantom. These results are in agreement with those on the validation set and are promising for the generalization of our trained model beyond simulations to out-of-distribution heterogeneous tissues

5.3. In-vivo demonstration

Fig. 7 shows the predictions of our model for three breast regions in a healthy volunteer. As with the phantom evaluation, we calculate the

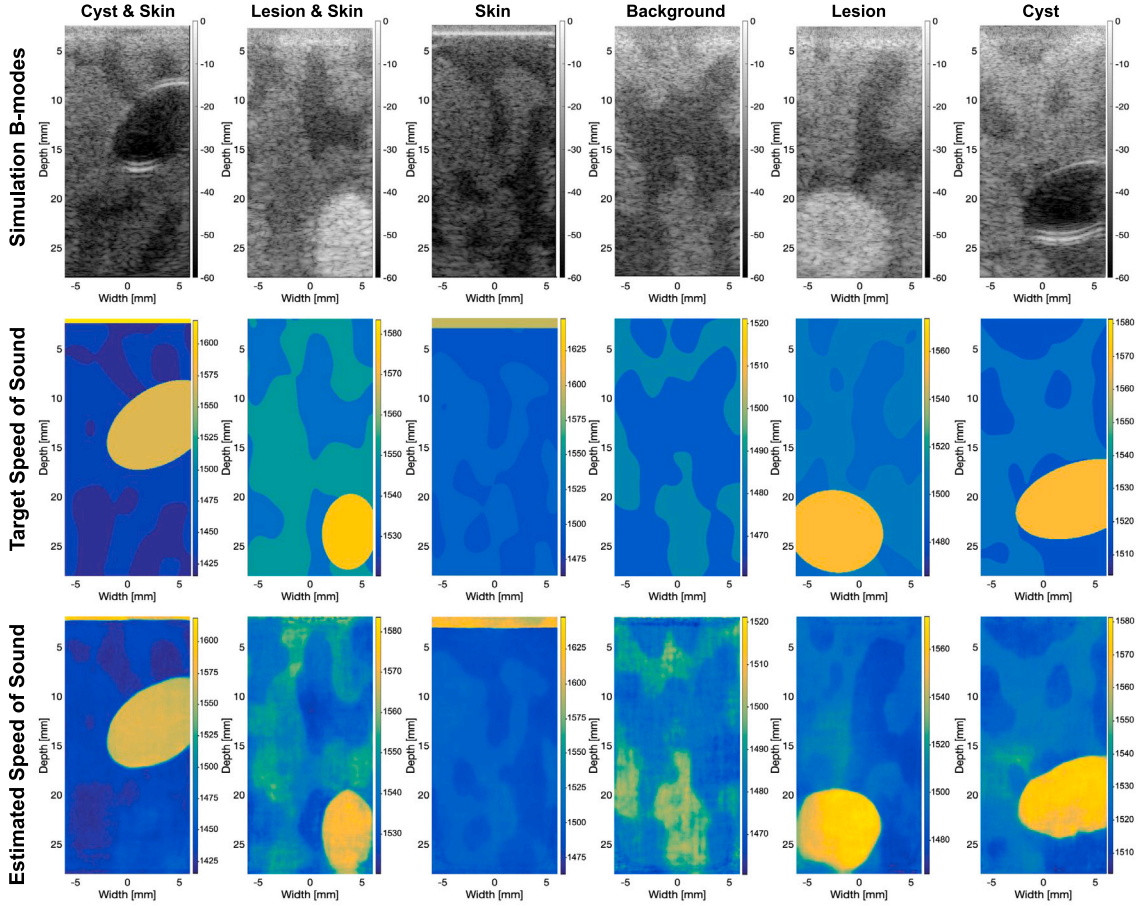

Fig. 5. Simulated B-modes with target sound speed and model estimation from the six classes. Our simulations produce quasi-realistic B-mode images, and our model successfully estimates sound speeds and contours of cysts, lesions, skin, glandular and fatty tissue.

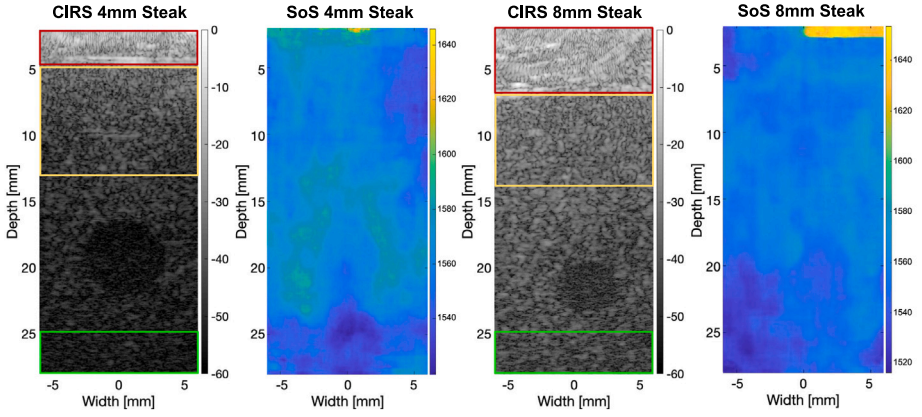

Fig. 6. Sound speed estimations for CIRS and steak layered phantoms along with B-mode images. The red ROI delineates the steak at a depth of 4 mm and 8 mm respectively. The yellow ROI delineates the top of the abutting CIRS layer and is 8.6 mm thick (left) and 6.6 mm thick (right). A green ROI encloses the bottom 2.9 mm of both phantoms. Model estimations are coherent and in agreement with the measured sound speed of 1566 m/s for the steak and 1558 m/s for the CIRS background.

Table 5
Sound speed estimations [m/s] and errors for the CIRS and steak phantom predictions compared with the insertion and speckle brightness methods in m/s. Estimations, errors, and standard deviations are computed over 100 consecutive frames.

	Traditional measurements	4mm steak		8mm steak	
		Estimation	Error	Estimation	Error
Steak (red)	1566	1564.4 ± 3.60	-1.60 ± 3.60	1564.6 ± 2.70	-1.40 ± 2.70
CIRS Background (yellow)	1558	1555.6 ± 4.43	-2.40 ± 4.43	1558.9 ± 2.49	$+0.90 \pm 2.50$
CIRS Background (green)	1558	1544.7 ± 6.90	-13.9 ± 6.90	1542.7 ± 4.80	-15.3 ± 4.80

average sound speed over 100 consecutive frames with a static probe for the in-vivo measurements. With no specific ROI or ground truth, the estimated sound speed of the entire field view is evaluated. The overall mean sound speed over 100 frames for R1, R2, and R3 are 1518.0 ± 5.3, 1500.1 ± 6.1, and 1499.0 ± 3.4 m/s, These values are consistent with each other and the literature on the sound speed of measured glandular breast tissue of 1505.0 ± 47.3 m/s from the Foundation for Research on Information Technologies in Society (IT'IS) [46] and 1510 m/s from [47]. Also, the model predictions align with the values of our simulated dataset, where the breast gland is modeled with sound speed between 1480 m/s and 1528 m/s following [3].

6. Discussion

Our proposed modeling of breast tissue creates B-mode images from simulated signals which closely resemble tissue B-modes due to their accuracy and grounding in ex vivo tissue measurements. Networks trained on these simulated signals can be deployed on real scanners and applied to in-vivo data with interpretable results. The proposed 3D phantom allows the modeling of 3D wave simulations on an in-silico phantom. To have 2D label maps for each 3D phantom, sound speed, and density were projected in the elevational plane. This projection allows for an accurate and known 2D sound speed distribution in the imaging plane despite the use of a 3D phantom used for the non-linear wave simulations. The assumption of elevational consistency is considered a fair approximation of many in-vivo tissue distributions. It is essential to note that all simulations modeled the 3D wave propagation within a 3D medium. Critically, the scatterer distribution field was modeled in 3D, ensuring that every slice in the elevational plane was independent despite the sound speed and density projection.

The proposed model with the addition of the dense skip connections and TNA was able to outperform the model without dense skip connections and the baseline proposed in [17]. This could be attributed to the smaller model size of [17] consisting of 3.9M trainable parameters, in comparison to the 15M parameters of the proposed model. Moreover, long-term skip connections and dense skip connections have been used by various state-of-the-art architectures and have led to improved training and gradient flow [48]. Finally, the simulated dataset in our work contained higher tissue variability and high frequency features, not included in the dataset proposed in [17], that increased the complexity of our sound speed reconstruction task.

A possible advantage of the complex beamformed representation is that our model can more easily process both the magnitude and phase information when predicting spatial sound speed distributions, considering slight phase shifts or drops in spatial coherence within the network. When generating sound speed estimates, our fully convolutional architecture considers a multi-scale context, i.e., large anatomical features, local phase shift, and coherence features. High-frequency filters in shallow layers and a large receptive field in deeper layers of the proposed architecture [36] enable multi-scale feature extraction. These multi-scale features are collected via skip connections and combined with the decoder weights to generate the final sound speed estimates.

Due to the constant sound speed assumption used for beamforming, the geometry of some B-mode images in Fig. 5 can be spatially distorted compared to the true geometrical layout of their respective media. Geometric deformation from gross phase aberration is especially prominent below the lesion regions for the Lesion & Skin and Lesion classes in Fig. 5 where the lower lesion boundary is not pictured in the B-mode but is visible in the sound speed simulation medium. Importantly, the estimated sound speed maps correspond correctly to the spatial distribution of the target sound speed maps and not the distribution of brightness in the B-modes, indicating that signal B-mode contours do not dictate the geometric layout of the sound speed estimate.

It could be expected that the brightness of the speckle in the B-mode is correlated with the underlying sound speed and would therefore be a strong indicator of average sound speed in the medium. In this work, the echogenicity of the medium was not correlated with the mean sound speed in the training data, and random sound speed variations occurred in highly echogenic and anechoic regions. The scatterer sound speed standard deviation and density contribute to brightness and were sampled independently of mean sound speed. This made the trained model robust to gross echogenicity variations. Furthermore, the sound speed values of anechoic cysts are also accurately estimated. Since anechoic cysts contain no reflectors, there is no local spatial information in the pulse-echo signal to indicate their sound speed. The correct estimation of the sound speed of anechoic cysts shows that global context is used to estimate sound speed and not only local echogenicity.

The real-world predictions of the layered CIRS and steak phantoms in Fig. 6 show that both cases are consistent with a homogeneous background unseen in the training set. The sound speed difference between the steak and CIRS layers is measured with the insertion method of 8 m/s. Our model estimates a sound speed with an accuracy of 8.8 m/s for the 4 mm steak phantom and 5.7 m/s for the 8 mm phantom, close to the insertion and speckle brightness methods. In the case of the homogeneous medium, the network did not infer a sound speed distribution with the appearance of the breast tissue from the training set but rather correctly inferred a homogeneous sound speed, indicating that the network had learned a collection of robust features that generalize beyond the training data to real transducer data and out of distribution property geometries. Normally, it would be expected for network performance to deteriorate on out-of-distribution samples. Still, the macro sound speed estimate here is accurate even for out-of-distribution homogeneous samples. Two regions of over-estimation (1620 m/s) can be seen in the top 1–2 mm of both phantoms, especially the 8 mm steak phantom, resembling the skin class from the training set. Despite this fact, when median absolute distance outlier removal is applied frame-wise to the sound speed in the region of interest, the sound speed estimate is 1562.3 ± 2.6 over 100 frames, only modestly increasing the regional error by 2.3 m/s.

The in-vivo demonstration showed global sound speed estimates aligned with reference values from the literature. Unlike the homogeneous phantom models, the in-vivo estimates displayed the expected tissue variation in the sound speed estimate, which resembles the underlying breast tissue distribution. Furthermore, all estimated values in the in-vivo estimation were within the expected range for in-vivo breast tissue. It is important to note that the same model, trained on the simulated dataset, is evaluated on both phantom and in-vivo data after being trained only on simulated data. The model could differentiate the sound speed regimes of ~ 1550 m/s and ~ 1500 m/s, respectively.

We hypothesize that the slight sound speed underestimation in the bottom of both the phantom and in-vivo scans results from lower SNR deeper in the medium. Thermal noise is present in real-world transducers, and TNA contributes toward bridging the performance gap but does not entirely alleviate the problem. Further, the thermal noise amplitude used in the TNA might not directly match the amplitude in the US device, partly due to mismatched attenuation values. Other noise sources in the signals from the lower regions could lower the SNR and contribute to performance loss.

Increasing SNR via, e.g., higher angular sampling frequency by an increased number of plane wave firings could potentially alleviate the problem of low SNR at deeper imaging depths and lead to increased performance on less superficial anatomies. An increase in the number of transmissions passed to the network comes at a computational cost when generating simulation data, which is why it was not performed in this work.

Our phantom and in-vivo results display the proposed method's robustness by correctly predicting sound speed on out-of-distribution data and under real-world factors. This robustness can be attributed to the proposed modestly anatomically realistic simulations and the pre-processing pipeline with TNA that improves generalization to real-world signals.

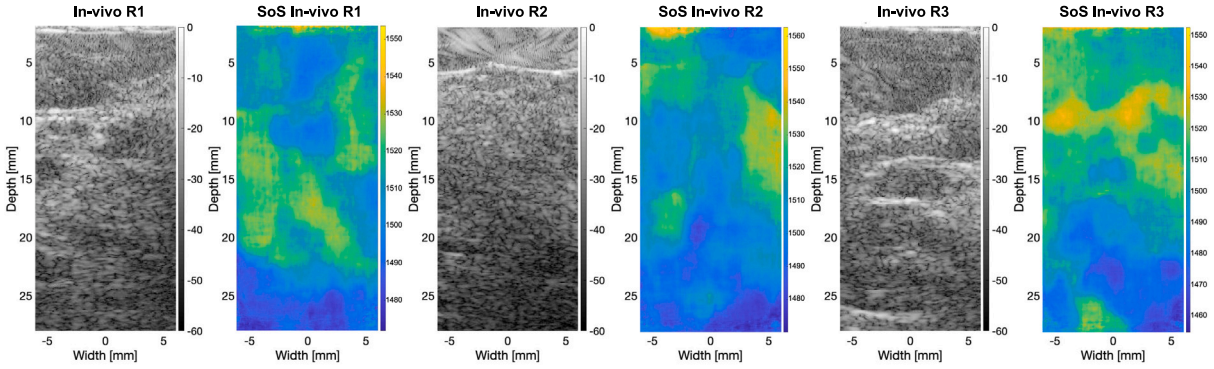

Fig. 7. In-vivo sound speed estimations and B-mode images for three breast regions R1–R3. Our model is able to estimate all three breast regions, with breast gland sound speed values within the sound speed range measured in [3,46,47].

Furthermore, the robust evaluation of our method goes beyond the standard protocol for deep model evaluation in medical imaging. This evaluation pipeline includes testing our model on external data sources of real-world phantom and real-world in-vivo data not included in the training distribution and reporting the model predictions over 100 US sequential frames. The real-world inference on in-vivo data showed that the proposed method could generalize beyond the elliptical and linear contours of the training dataset and infer sound speed on biological sound speed distributions. Our errors' low standard deviation shows our predictions' stability over 100 consecutive frames. This approach could set a new precedent for evaluating the consistency of sound speed estimation for physics-based and deep-learning models.

Future work includes more realistic modeling of real transducers and in-vivo artifacts. The dataset could be extended to include irregularly shaped lesions to model malignant tissue with irregular boundaries. Such modeling will be crucial for developing robust and generalizable sound speed estimation models with DNNs. Furthermore, the presented method utilizes three plane waves, which reduces the SNR of the signal at both training and inference time. It is expected that with the simulation of more plane waves to a comparable number to [14], the performance could increase further along with the computational cost. Finally, our dataset could be used as a benchmark for sound speed estimation methods to increase their comparability, similar to challenges in beamforming such as PICMUS and CUBDL [49,50].

7. Conclusion

In this paper, we proposed a novel pipeline for sound speed estimation of breast ultrasound imaging. A large-scale simulation dataset was created, processed, and used to train our tailored fully convolutional network architecture to produce sound speed estimations from beamformed analytic signal plane wave ultrasound data. Our model, which will become publicly available along with our simulated dataset, is a promising step toward reproducible sound speed estimation for ultrasound imaging and could be further extended to other anatomies, such as thyroid or liver, or used as an initialization for traditional ultrasound estimation techniques such as [14]. The proposed method is simulator agnostic given the input phantoms presented in this work since many full wave simulations display similar performance [51]. Our method was evaluated on simulated, phantom, and in-vivo breast ultrasound data. The estimated sound speeds were temporally consistent among frames and agreed with traditionally measured sound speeds and clinical literature. Future work could also utilize our predicted sound speeds to improve beamforming quality by removing constant sound speed and straight ray propagation assumptions.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Jeremy Dahl reports financial support was provided by National Institute of Health.

Data availability

Data will be made available on request.

Acknowledgments

The work of Walter Simson was supported by grant ZF4190502CR8 of the Zentrale Innovationsprogramm Mittelstand (ZIM). The work of Jeremy Dahl was supported by grant R01-EB027100 from the National Institute of Biomedical Imaging and Bioengineering, United States.

References

[1] World Health Organization (WHO), Breast cancer, 2021, https://www.who.int/news-room/fact-sheets/detail/breast-cancer. (Accessed: 22 Aug 2021).
[2] R. Smithuis, L. Wijers, I. Dennert, Ultrasound of the breast, 2010, https://radiologyassistant.nl/breast/ultrasound/ultrasound-of-the-breast. (Accessed: 22 Aug 2021).
[3] J. Bamber, Ultrasonic propagation properties of the breast, Ultrason. Exam. Breast (1983) 37–44.
[4] M.K. Feldman, S. Katyal, M.S. Blackwood, US artifacts, Radiographics 29 (4) (2009) 1179–1189.
[5] R. Ali, T. Brevett, L. Zhuang, H. Bendjador, A.S. Podkowa, S.S. Hsieh, W. Simson, S.J. Sanabria, C.D. Herickhoff, J.J. Dahl, Aberration correction in diagnostic ultrasound: A review of the prior field and current directions, Zeitschrift für Medizinische Phys. (2023).
[6] S.J. Sanabria, E. Ozkan, M. Rominger, O. Goksel, Spatial domain reconstruction for imaging speed-of-sound with pulse-echo ultrasound: Simulation and in Vivo study, Phys. Med. Biol. 63 (21) (2018) 215015.
[7] J.F. Greenleaf, S.A. Johnson, S.L. Lee, G. Hermant, E. Woo, Algebraic reconstruction of spatial distributions of acoustic absorption within tissue from their two-dimensional acoustic projections, in: Acoustical Holography. Vol. 5, Springer, 1974, pp. 591–603.
[8] J.F. Greenleaf, R.C. Bahn, Clinical imaging with transmissive ultrasonic computerized tomography, IEEE Trans. Biomed. Eng. 2 (2) (1981) 177–185.
[9] A.C. Kak, M. Slaney, Principles of Computerized Tomographic Imaging, SIAM, 2001.
[10] M.E. Anderson, G.E. Trahey, The direct estimation of sound speed using pulse–echo ultrasound, J. Acoust. Soc. Am. 104 (5) (1998) 3099–3106.
[11] M. Jakovljevic, S. Hsieh, R. Ali, G. Chau Loo Kung, D. Hyun, J.J. Dahl, Local speed of sound estimation in tissue using pulse-echo ultrasound: Model-based approach, J. Acoust. Soc. Am. 144 (1) (2018) 254–266.
[12] R. Ali, A.V. Telichko, H. Wang, U.K. Sukumar, J.G. Vilches-Moure, R. Paulmurugan, J.J. Dahl, Local sound speed estimation for pulse-echo ultrasound in layered media, IEEE Trans. Ultrason. Ferroelectr. Freq. Control 69 (2) (2021) 500–511.

[13] M. Jaeger, M. Frenz, Towards clinical computed ultrasound tomography in echo-mode: Dynamic range artefact reduction, Ultrasonics 62 (2015) 299–304.
[14] P. Stähli, M. Kuriakose, M. Frenz, M. Jaeger, Improved forward model for quantitative pulse-echo speed-of-sound imaging, Ultrasonics 108 (2020) 106168.
[15] M. Jaeger, P. Stähli, N.K. Martiartu, P.S. Yolgunlu, T. Frappart, C. Fraschini, M. Frenz, Pulse-echo speed-of-sound imaging using convex probes, Phys. Med. Biol. 67 (21) (2022) 215016.
[16] W. Simson, L. Zhuang, S.J. Sanabria, J.J. Dahl, D. Hyun, Differentiable beamforming for ultrasound autofocusing, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2023, pp. 428–437.
[17] M. Feigin, D. Freedman, B.W. Anthony, A deep learning framework for single-sided sound speed inversion in medical ultrasound, IEEE Trans. Biomed. Eng. 67 (4) (2019) 1142–1151.
[18] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, 2014, arXiv preprint arXiv:1409.1556.
[19] B.E. Treeby, B.T. Cox, K-wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields, J. Biomed. Opt. 15 (2) (2010) 021314.
[20] M. Bernhardt, V. Vishnevskiy, R. Rau, O. Goksel, Training variational networks with multidomain simulations: Speed-of-sound image reconstruction, IEEE Trans. Ultrason. Ferroelectr. Freq. Control 67 (12) (2020) 2584–2594.
[21] F.K. Jush, P.M. Dueppenbecker, A. Maier, Data-driven speed-of-sound reconstruction for medical ultrasound: Impacts of training data format and imperfections on convergence, in: Annual Conference on Medical Image Understanding and Analysis, Springer, 2021, pp. 140–150.
[22] F.K. Jush, M. Biele, P.M. Dueppenbecker, O. Schmidt, A. Maier, Dnn-based speed-of-sound reconstruction for automated breast ultrasound, in: 2020 IEEE International Ultrasonics Symposium, IUS, IEEE, 2020, pp. 1–7.
[23] Y. Wang, E. Helminen, J. Jiang, Building a virtual simulation platform for quasistatic breast ultrasound elastography using open source software: A preliminary investigation, Med. Phys. 42 (9) (2015) 5453–5466.
[24] M. Salehi, S.-A. Ahmadi, R. Prevost, N. Navab, W. Wein, Patient-specific 3D ultrasound simulation based on convolutional ray-tracing and appearance optimization, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 510–518.
[25] B.A. Lau, I. Reiser, R.M. Nishikawa, P.R. Bakic, A statistically defined anthropomorphic software breast phantom, Med. Phys. 39 (6Part1) (2012) 3375–3385.
[26] S.J. Glick, L.C. Ikejimba, Advances in digital and physical anthropomorphic breast phantoms for x-ray imaging, Med. Phys. 45 (10) (2018) e870–e885.
[27] B. Burger, S. Bettinghausen, M. Radle, J. Hesser, Real-time GPU-based ultrasound simulation using deformable mesh models, IEEE Trans. Med. Imaging 32 (3) (2012) 609–618.
[28] T.D. Mast, Empirical relationships between acoustic parameters in human soft tissues, Acoust. Res. Lett. Online 1 (2) (2000) 37–42.
[29] M. Tirindelli, C. Eilers, W. Simson, M. Paschali, M.F. Azampour, N. Navab, Rethinking ultrasound augmentation: A physics-inspired approach, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VIII 24, Springer, 2021, pp. 690–700.
[30] X. Huang, M.A.L. Bell, K. Ding, Deep learning for ultrasound beamforming in flexible array transducer, IEEE Trans. Med. Imaging (2021).
[31] D. Hyun, L.L. Brickson, K.T. Looby, J.J. Dahl, Beamforming and speckle reduction using neural networks, IEEE Trans. Ultrason. Ferroelectr. Freq. Control 66 (5) (2019) 898–910.
[32] S. Jégou, M. Drozdzal, D. Vazquez, A. Romero, Y. Bengio, The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017, pp. 11–19.
[33] K. He, X. Zhang, S. Ren, J. Sun, Delving deep into rectifiers: Surpassing human-level performance on imagenet classification, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1026–1034.
[34] D. Ulyanov, A. Vedaldi, V. Lempitsky, Instance normalization: The missing ingredient for fast stylization, 2016, arXiv:1607.08022.
[35] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241.
[36] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, 2016, http://www.deeplearningbook.org.
[37] S.-Y. Huang, J.M. Boone, K. Yang, A.L. Kwan, N.J. Packard, The effect of skin thickness determined using breast CT on mammographic dosimetry, Med. Phys. 35 (4) (2008) 1199–1206.
[38] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, 2014, arXiv preprint arXiv:1412.6980.
[39] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, et al., PyTorch: An imperative style, high-performance deep learning library, in: Advances in Neural Information Processing Systems. Vol. 32, Curran Associates, Inc., 2019, pp. 8024–8035.
[40] W. Falcon, et al., PyTorch Lightning. Vol. 3, 2019, GitHub. https://github.com/PyTorchLightning/pytorch-lightning.
[41] L. Biewald, Experiment tracking with weights and biases, 2020, Software available from wandb.com, URL https://www.wandb.com/.
[42] L. Nock, G.E. Trahey, S.W. Smith, Phase aberration correction in medical ultrasound using speckle brightness as a quality factor, J. Acoust. Soc. Am. 85 (5) (1989) 1819–1833.
[43] D. Hyun, A. Wiacek, S. Goudarzi, S. Rothlübbers, A. Asif, K. Eickel, Y.C. Eldar, J. Huang, M. Mischi, H. Rivaz, et al., Deep learning for ultrasound image formation: CUBDL evaluation framework & open datasets, IEEE Trans. Ultrason. Ferroelectr. Freq. Control (2021).
[44] W. Marczak, Water as a standard in the measurements of speed of sound in liquids, J. Acoust. Soc. Am. 102 (5) (1997) 2776–2779.
[45] I. Kuo, B. Hete, K. Shung, A novel method for the measurement of acoustic speed, J. Acoust. Soc. Am. 88 (4) (1990) 1679–1682.
[46] P. Hasgall, D. Gennaro, C. Baumgartner, E. Neufeld, et al., IT'IS database for thermal and electromagnetic parameters of biological tissues, version 4.0, 2018, http://dx.doi.org/10.13099/VIP21000-04-0, URL.
[47] J. Nebeker, T.R. Nelson, Imaging of sound speed using reflection ultrasound tomography, J. Ultrasound Med. 31 (9) (2012) 1389–1404.
[48] D. Balduzzi, M. Frean, L. Leary, J. Lewis, K.W.-D. Ma, B. McWilliams, The shattered gradients problem: If resnets are the answer, then what is the question? in: International Conference on Machine Learning, PMLR, 2017, pp. 342–350.
[49] D. Hyun, A. Wiacek, S. Goudarzi, S. Rothlübbers, A. Asif, K. Eickel, Y.C. Eldar, J. Huang, M. Mischi, H. Rivaz, et al., Deep learning for ultrasound image formation: CUBDL evaluation framework and open datasets, IEEE Trans. Ultrason. Ferroelectr. Freq. Control 68 (12) (2021) 3466–3483.
[50] H. Liebgott, A. Rodriguez-Molares, F. Cervenansky, J.A. Jensen, O. Bernard, Plane-wave imaging challenge in medical ultrasound, in: 2016 IEEE International Ultrasonics Symposium, IUS, IEEE, 2016, pp. 1–4.
[51] J.-F. Aubry, O. Bates, C. Boehm, K. Butts Pauly, D. Christensen, C. Cueto, P. Gélat, L. Guasch, J. Jaros, Y. Jing, et al., Benchmark problems for transcranial ultrasound simulation: Intercomparison of compressional wave models a, J. Acoust. Soc. Am. 152 (2) (2022) 1003–1019.