Neural Operator Learning for Ultrasound Tomography Inversion

Haocheng Dai*, Michael Penwarden*, Mike Kirby, Sarang Joshi. (*equal contribution)
MIDL 2023
More women are diagnosed with breast cancer than any other type of cancer.
More woman are died from breast cancer than any other type of cancer, besides lung cancer.

The 5-year survival rate for breast cancer stands at 90%, yet this rate is closely tied to the (1) timing of detection and (2) treatment.

Mammograms (X-ray)

50% of women between the ages of 40-74 in the US have dense breasts.

❌ Mammograms misses 35.6–52.2% of breast cancers in dense breast tissue.

MRI

✅ works well on extremely dense breast tissue.

❌ high expenses, hard accessibility, and high false positive rates.

Ultrasound

✅ low expenses, radiation free

✅ can detect tiny, node-negative breast tumors

✅ can be used in treatment
The 2D scalar wave equation in the homogeneous medium can be denoted as:

\[
\frac{\partial^2 p(x, t)}{\partial x_1^2} + \frac{\partial^2 p(x, t)}{\partial x_2^2} - \frac{1}{c(x)^2} \cdot \frac{\partial^2 p(x, t)}{\partial t^2} = 0
\]
In steady-state conditions, namely the source pressure is periodic and never fades out, we can Fourier-transform the pressure in time to

\[ p(x, t) = \int_{-\infty}^{\infty} p(x, \omega) e^{-j\omega t} d\omega \]

By substituting the Fourier-transformed pressure into the wave equation, we have

\[ \nabla^2 p(x, t) - \frac{1}{c(x)^2} \cdot \frac{\partial^2 p(x, t)}{\partial t^2} = 0 \]

\[ \nabla^2 \left( \int_{-\infty}^{\infty} p(x, \omega) e^{-j\omega t} d\omega \right) - \frac{1}{c(x)^2} \cdot \frac{\partial^2}{\partial t^2} \left( \int_{-\infty}^{\infty} p(x, \omega) e^{-j\omega t} d\omega \right) = 0 \]

\[ \int_{-\infty}^{\infty} \left[ \nabla^2 (p(x, \omega)e^{-j\omega t}) - \frac{1}{c(x)^2} \cdot \frac{\partial^2}{\partial t^2} (p(x, \omega)e^{-j\omega t}) \right] d\omega = 0 \]

\[ \int_{-\infty}^{\infty} \left[ e^{-j\omega t} \nabla^2 p(x, \omega) - \frac{p(x, \omega)}{c(x)^2} \cdot \frac{\partial^2}{\partial t^2} (e^{-j\omega t}) \right] d\omega = 0 \]

\[ \int_{-\infty}^{\infty} \left[ e^{-j\omega t} \nabla^2 p(x, \omega) - \frac{p(x, \omega)}{c(x)^2} \cdot (-\omega^2 e^{-j\omega t}) \right] d\omega = 0 \]

\[ \int_{-\infty}^{\infty} \left[ \nabla^2 p(x, \omega) + \frac{\omega^2}{c(x)^2} \cdot p(x, \omega) \right] e^{-j\omega t} d\omega = 0 \]

\[ \nabla^2 p(x, \omega) + \frac{\omega^2}{c(x)^2} \cdot p(x, \omega) = 0 \]
The central goal of USCT is to reconstruct the spatial distribution of sound speed within biological tissues, denoted by \( c(x) \), using the measurements obtained from the transducers, represented by \( y \). This inverse problem can be formulated as a PDE-constrained optimization problem:

\[
\min_{c(x), u_k(x)} L = \sum_{k=1}^{M} L_k = \sum_{k=1}^{M} \| y_k - u_k(x_f) \|_2^2 \\
\text{s.t.} \left[ \nabla^2 + \left( \frac{\omega}{c(x)} \right)^2 \right] u_k(x) = -\rho_k(x).
\]

This inverse problem is commonly referred to as frequency domain full waveform inversion (FD-FWI). Akin to many other inverse problems, FD-FWI is typically solved using gradient-based optimizers, but it can be computationally challenging for large-scale problems like 3D reconstructions.

Can we find a data-driven method in the era of deep learning to solve USCT problem? Namely learning a mapping between complex function spaces.
K-Wave simulator

kgrid
 Needed
.Nx 
.dx 
.t_array 
.Nt 
.dt 
.k

medium 
 Needed
.c/\rho 
.sound_speed 
.density 
.BonA 
.alpha_power 
.alpha_coeff

source
 Needed
.p0 
.p_mask 
.p 
.u_mask 
.ux 
.uy 
.uz

sensor
 Needed
.mask 
.record

kspaceFirstOrder1D(kgrid, medium, source, sensor)
kspaceFirstOrder2D(kgrid, medium, source, sensor)
kspaceFirstOrder3D(kgrid, medium, source, sensor)

sensor_data

Initial Pressure
Sensor Mask
Sound Speed
Density
Sensor Position
Time Step
K-Wave Simulation

Masking & Skin Wrapping

Scaling & Thresholding

Raw GRF

Emitter Receiver Wave

Speed of Sound

Emitter

Receiver

Wave

Water TOF

TOF Diff

Emitter Axis

Receiver Axis

GRF TOF

Cross Correlation

Time of Flight

Concatenate TOF Diff w/ Positional Encoding

Recovered Speed of Sound

Noised

Clean
Single realization of train/test set examples

T-FNO better captures the structure feature in the phantom, compared to U-Net.

The dashed red line indicates the location of the evenly distributed emitters and receivers. The region outside the transducer ring is masked out for improved training and quantity of interest comparison but is present in the full-wave simulation to mitigate reflection.
MSE comparison between T-FNO and U-Net

<table>
<thead>
<tr>
<th>Model</th>
<th>GRF Correlation</th>
<th>Noise</th>
<th>Testing MSE</th>
<th>Training MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>T-FNO</td>
<td>High</td>
<td>Clean</td>
<td>$2.01 \pm 0.33 \times 10^{-2}$</td>
<td>$0.69 \pm 0.11 \times 10^{-2}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10%</td>
<td>$2.06 \pm 0.34 \times 10^{-2}$</td>
<td>$0.68 \pm 0.11 \times 10^{-2}$</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>Clean</td>
<td>$3.27 \pm 0.16 \times 10^{-2}$</td>
<td>$1.00 \pm 0.06 \times 10^{-2}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10%</td>
<td>$2.67 \pm 0.17 \times 10^{-2}$</td>
<td>$1.53 \pm 0.08 \times 10^{-2}$</td>
</tr>
<tr>
<td>U-Net</td>
<td>High</td>
<td>Clean</td>
<td>$2.52 \pm 0.44 \times 10^{-2}$</td>
<td>$0.12 \pm 0.03 \times 10^{-2}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10%</td>
<td>$2.79 \pm 0.42 \times 10^{-2}$</td>
<td>$0.01 \pm 0.01 \times 10^{-2}$</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>Clean</td>
<td>$3.81 \pm 0.17 \times 10^{-2}$</td>
<td>$0.42 \pm 0.05 \times 10^{-2}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10%</td>
<td>$4.02 \pm 0.21 \times 10^{-2}$</td>
<td>$0.02 \pm 0.01 \times 10^{-3}$</td>
</tr>
</tbody>
</table>

T-FNO outperforms the U-Net at test time under all conditions, whereas the U-Net better fits the training set but does not generalize well.
T-FNO better captures the overall trends in the data, while the U-Net is prone to **overfit** the training data.

The U-Net suffers from considerable **generalization error**, even when additional examples were provided.
Conclusions

1. We have proposed using neural operators to accurately and efficiently solve the full-wave inverse problem on synthetic ultrasound tomography.

2. Our novel application of the T-FNO improves over the baseline U-Net, laying the foundation for real-time accurate predictions of soft tissue distribution for tumor identification on breast imaging.

3. Additionally, the application of both U-Net and T-FNO to this problem formulation is itself novel since both are real-time predictors and do not require computationally expensive ray-based inversion once trained.