MDPI

Article

Fourier Neural Operator Network for Fast Photoacoustic Wave Simulations

Steven Guan †, Ko-Tsung Hsu † and Parag V. Chitnis *

Bioengineering Department, George Mason University, Fairfax, VA 22030, USA
* Correspondence: pchitnis@gmu.edu
† These authors contributed equally to this work.

Abstract: Simulation tools for photoacoustic wave propagation have played a key role in advancing photoacoustic imaging by providing quantitative and qualitative insights into parameters affecting image quality. Classical methods for numerically solving the photoacoustic wave equation rely on a fine discretization of space and can become computationally expensive for large computational grids. In this work, we applied Fourier Neural Operator (FNO) networks as a fast data-driven deep learning method for solving the 2D photoacoustic wave equation in a homogeneous medium. Comparisons between the FNO network and pseudo-spectral time domain approach were made for the forward and adjoint simulations. Results demonstrate that the FNO network generated comparable simulations with small errors and was orders of magnitude faster than the pseudo-spectral time domain methods (~26× faster on a 64 × 64 computational grid and ~15× faster on a 128 × 128 computational grid). Moreover, the FNO network was generalizable to the unseen out-of-domain test set with a root-mean-square error of 9.5×10^{-3} in Shepp–Logan, 1.5×10^{-2} in synthetic vasculature, 1.1×10^{-2} in tumor and 1.9×10^{-2} in Mason-M phantoms on a 64 × 64 computational grid and a root mean squared of $6.9 \pm 5.5 \times 10^{-3}$ in the AWA2 dataset on a 128 × 128 computational grid.

Keywords: photoacoustic imaging; image processing; computer vision; simulation; reconstruction; deep learning

1. Introduction

Photoacoustic imaging is a non-invasive hybrid imaging modality that combines the advantages of optical (e.g., high contrast and molecular specificity) and ultrasound (e.g., high penetration depth) imaging [1]. It has been applied for many preclinical and clinical imaging applications, such as small-animal whole-body imaging, breast and prostate cancer imaging and image-guided surgery [2–6]. Specifically, in breast cancer detection, tumors have been successfully revealed by single-breath-hold photoacoustic computed tomography (SBH-PACT) without the need of ionizing radiation and exogenous contrast agents based on the higher blood vessel density characteristics associated with tumors [7]. Multispectral photoacoustic imaging can be used for functional imaging, such as measuring blood oxygen saturation and metabolism in biological tissues [8]. In addition to applying multispectral photoacoustic imaging to differentiate oxyhemoglobin from deoxyhemoglobin in breast cancer, ultrasound and photoacoustic tomography (US-OT) can reveal differences in lipids and collagen in breast fibroglandular tissue, providing more clinically meaningful insights for diagnosis [9]. Photoacoustic imaging provides both structural and functional information that can potentially reveal novel insights into biological processes and disease pathologies [10].

In photoacoustic tomography (PAT), a tissue medium is illuminated using a short-pulsed laser. Optically absorbing molecules within the medium are excited and undergo thermoelastic expansion, resulting in the generation of photoacoustic waves that are subsequently measured using an array of acoustic sensors [1]. An image representing the initial

pressure distribution can be reconstructed from the measured time-dependent signals using analytical solutions, numerical methods and model-based iterative methods [11–15]. A detailed understanding of parameters describing the imaging medium (e.g., optical, thermal and acoustic properties of the tissue) and the imaging system (e.g., arrangement and characteristics of the laser source and acoustic sensors) is needed to reconstruct a high-quality PAT image.

PAT simulation is a highly useful tool that provides quantitative and qualitative insights into these parameters affecting image quality [16]. It is commonly used prior to experimentation and imaging to optimize the system configuration. It also plays an integral role in PAT image reconstruction and provides numerical phantom data for the development of advanced algorithms, such as iterative methods and deep learning methods [17–22]. Simulating PAT image acquisition comprises two components: optical illumination and photoacoustic propagation. For this work, we are primarily focused on the photoacoustic propagation component. The equation for photoacoustic wave propagation can be solved numerically using classical methods, such as the time domain finite element method [23,24]. These methods generally solve the equation via approximation on a mesh and can be computationally expensive, such as for large three-dimensional (3D) simulations.

Recently, deep learning has been explored as an alternative method for solving partial differential equations (PDE) [25,26]. It has the potential to greatly impact scientific disciplines and research by providing fast PDE solvers that approximate or enhance conventional ones. Applications requiring repeated evaluations of a PDE can greatly benefit from the reduced computation times offered by deep learning. Here, we provide a brief overview of three deep learning methods for solving PDEs—finite dimensional operators, neural finite element models and Fourier Neural Operators (FNO).

Finite dimensional operators use a deep convolutional neural network (CNN) to solve the PDE on a finite Euclidean Space [27,28]. This approach is mesh-dependent, meaning the CNN needs to be retrained for solving the PDE at different spatial resolutions and discretization. Neural finite element models are mesh-independent and closely resemble traditional finite element methods [25,29]. It replaces the set of local basis functions in the finite element models with a fully connected neural network. It requires prior knowledge of the underlying PDE and is designed to solve for one specific instance of the PDE. The neural network needs to be retrained for new instances where the underlying PDE is parameterized with a different set of functional coefficients. FNO is a mesh-free approach that approximates the mapping between two infinite dimensional spaces from a finite collection of input–output paired observations [30,31]. The neural operator is learned directly in the Fourier and image space using a CNN. The same learned operator can be used without retraining to solve PDEs with different discretization and parameterization. Fourier Neural Operators have been demonstrated to achieve state-of-the-art results for a variety of PDEs (e.g., Burgers' equation, Darcy Flow and Navier–Stokes) and outperform other existing deep learning methods [31].

To the best of our knowledge, this is the first paper to apply deep learning for solving the photoacoustic wave equation for simulating PAT. FNOs were chosen for this task given their flexibility in discretization and superior performance compared to other deep learning methods. Prior work with FNOs demonstrated solutions to the Navier–Stokes and Burgers' equations, which have relatively smooth spatio-temporal solutions [31]. Unlike these works, photoacoustic signals have high broadband frequencies and contain sharp transitions. Specifically, this paper highlights the following innovative contributions:

- Adapting the FNO neural network and applying it as a fast PDE solver for simulating the forward and adjoint 2D photoacoustic operator.
- Simulations from the FNO network and the widely used k-Wave toolbox for time domain acoustic wave propagation [16] were compared in terms of accuracy and computation times.

- Further experiments were also conducted to evaluate the generalizability of the FNO network beyond the training data and the impact of key hyperparameters on network performance and complexity.

The remainder of the article is organized as follows. The forward problem and the inverse problem of PAT are described in Section 2. Acoustic wave simulation techniques for the PAT based on conventional methods and FNO networks are presented in Section 2. The data generation and training process of the FNO network and its detailed implementation are also given in Section 2. Simulation results of the FNO network on the different test set with different spatial grid sizes and hyperparameter optimizations are provided in Section 3. Furthermore, the simulation results of zero-shot super-resolution with the FNO network are also given in Section 3. The conclusion and discussion are presented in Sections 4 and 5, respectively.

2. Materials and Methods

2.1. Photoacoustic Signal Generation and Imaging

The photoacoustic signal is generated by irradiating the medium with a nanosecond laser pulse (Figure 1). Chromophores within the image medium are excited by the laser and undergo thermoelastic expansion to generate acoustic pressure waves. Assuming negligible thermal diffusion and volume expansion during illumination, the initial photoacoustic pressure x can be defined as

$$x(r) = \Gamma(r)A(r), \tag{1}$$

where $A(r)$ is the spatial absorption function, and $\Gamma(r)$ is the Grüneisen coefficient describing the conversion efficiency from heat to pressure [32]. The photoacoustic pressure wave $p(r, t)$ at position r and time t can be modeled as an initial value problem, where c is the speed of sound [33].

$$\left(\partial_{tt} - c_0^2 \Delta\right) p(r, t) = 0, \quad p(r, t = 0) = x, \quad \partial_t p(r, t = 0) = 0 \tag{2}$$
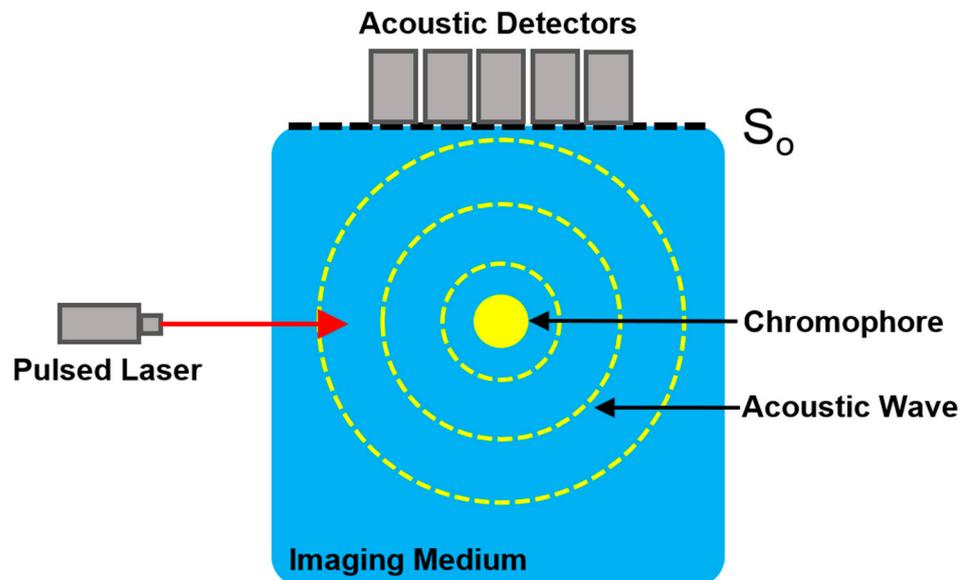

Figure 1. Diagram illustrating the process of photoacoustic signal generation and detection. Chromophores absorb the incident pulsed laser light and undergo thermoelastic expansion to generate acoustic waves. Acoustic detectors along the measurement boundary So are used to measure the acoustic waves.

In photoacoustic imaging, sensors located along a measurement surface S_o, surrounding the medium, are used to measure a time series signal. The linear operator $\mathcal{M}$ acts on $p(r, t)$ restricted to the boundary of the computational domain Ω over a finite time T

and provides a linear mapping to the measured time-dependent signal y. The forward photoacoustic operator W maps the initial acoustic pressure to the measured signal.

$$y = \mathcal{M}_{p|\partial\Omega \times (0,T)} = Wx \tag{3}$$

The measured sensor data are then used to form an image representing the initial acoustic pressure distribution. Photoacoustic image reconstruction is a well-studied inverse problem that can be solved using analytical solutions, numerical methods and model-based iterative methods [11–13,15,34]. The adjoint photoacoustic operator W^* maps the measured signal to the initial acoustic pressure.

$$x = W^*y \tag{4}$$

Time reversal is a robust reconstruction method that works well for homogenous and heterogeneous mediums and also for any arbitrary detection geometry [15,34]. The acoustic waves that are generated are measured along the measurement surface S_o. After a long period of time T, the acoustic field within the medium becomes zero, which is guaranteed by Huygens' principle in homogeneous mediums [35]. A PAT image is formed by running a numerical model of the forward problem and transmitting the measured sensor data in a time-reversed order into the medium, where the detectors along S_o are time-varying pressure sources. Thus, time reversal is modeled as a time-varying boundary value problem, and the resulting acoustic field at $t = 0$ is the initial acoustic pressure distribution to be recovered.

2.2. Conventional Solvers for the Wave Equation

Numerical approaches, such as the finite difference and finite element methods, are commonly used to solve PDEs by discretizing the space into a grid [36]. However, these methods are often slow for time domain modeling broadband or high-frequency waves due to the need for a fine grid with small time steps [16]. Computational efficiency can be improved using pseudo-spectral and k-space methods. The pseudo-spectral method fits a Fourier series to the data and reduces the number of grid points per wavelength required for an accurate solution [37]. The k-space method incorporates a priori information regarding the governing wave equation into the solution [38]. This allows for larger time steps and improves numerical stability in the case of acoustically heterogeneous media. The k-Wave toolbox, a widely used MATLAB tool for photoacoustic simulations, uses the pseudo-spectral k-space approach for solving time domain photoacoustic wave simulations [39].

Conventional numerical approaches are typically used to solve a single instance of PDEs and require the PDEs' explicit form. Because these approaches solve the PDE via approximation on a mesh, there is a trade-off between accuracy and computation time. In comparison, the FNO network is a data-driven and black-box approach that learns a solution for a family of PDEs from the training data and does not require the PDEs' explicit form. It is also resolution- and mesh-invariant, meaning that the trained network can be used for solving PDEs at varying levels of resolution and discretization. However, the FNO network is dependent on the quality of the data and is slow to train. Although modifications to the functional form of the PDE can be easily accounted for in conventional approaches to solve a single instance, the FNO network would need to be retrained with a new training dataset.

2.3. Fourier Neural Operator Networks

The FNO network [31] was adapted for solving the 2D photoacoustic wave equation. In the original implementation, the input a to the FNO was the first several time steps of the solution acquired using a conventional solver. Using insights from the physics of photoacoustic imaging, we determined that only the initial pressure distribution (time = 0) was needed as the input a to the FNO for solving the wave equation because the generated acoustic pulses propagate in an omni-directional manner. This removed the need

for conventional solvers, and the full spatio-temporal solution could be obtained using only the FNO. To avoid instabilities and keep our solution bounded, we replaced previously used element-wise Gaussian normalizers with a peak-normalization scheme, where the initial source distribution map was normalized by its maximum value. We empirically found that the Gaussian normalizer was not appropriate for simulating wave propagation because the distribution of pressures over time at any given element in the computational grid was not normally distributed.

The network begins by projecting the input *a* (initial pressure distribution) onto a higher dimensional latent representation using the fully connected layer FC_1 with a single shallow layer (Figure 2). The dimensionality of this latent representation is defined by the hyperparameter termed channels. Four Fourier layers are then used to iteratively update the projected features. In each Fourier layer, the features initially undergo a Fourier transform, which plays a key role in enabling the network to efficiently learn mesh- and resolution-invariant features for solving the PDE. Features learned in the Fourier space are global by nature and represent patterns spanning the whole computational grid. In contrast, features learned in a standard CNN are local by nature and represent patterns spanning over a local region (e.g., edges and shapes).

Pseudo-code for the FNO

```
def  FNO(x) :
    x  =  fc 0(x)

    for  i  in  range( 3) :
        x1 = SpectralConv  3d(x)
        x2 = conv 3d(x)
        x  = x1 + x2
        x  = gelu (x)

    x1 = SpectralConv  3d( x)
    x2 = conv 3d(x)
    x  = x1 + x2

    x  = fc 1(x)
    x  = gelu (x)
    x  = fc 2(x)

    return    x
```
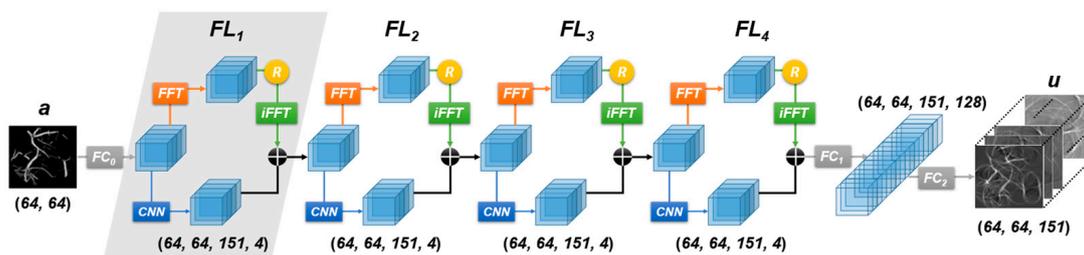

Figure 2. Neural network architecture for the FNO network. The input a (initial pressure distribution a) is mapped to a higher dimensional space using a fully connected layer (FC_0). The transformed feature is passed through four Fourier Layers (FLs). Finally, a fully connected layer (FC_2) is used to obtain the final output u (solution to the wave equation u) with the desired dimensions. The input goes through two paths in each Fourier layer. In the top path, the input undergoes a Fourier Transform FFT, linear transform R and inverse Fourier Transform iFFT. In the bottom path, the input undergoes a linear transform. Outputs from each path are summed together and undergo GeLU activation. The dimension of the feature representation for each operation is given in the parentheses.

After the Fourier transform, the resulting Fourier modes can be truncated to optimize computational efficiency. This is useful as a regularization technique and for PDEs with smooth solutions that can be accurately represented with fewer Fourier modes, as previously demonstrated for the 2D Navier–Stokes equation [31]. Following the Fourier layers, the updated features are projected in FC_2 to a higher dimensional representation with 128 channels in the hidden layer and finally to the desired dimensions in the final shallow layer to obtain the output u (solution to the wave equation). Through a combination of Fourier, linear and non-linear transformations, the FNO network can approximate highly complex and non-linear operators in PDEs.

Channels and modes are the two main hyperparameters for the FNO network. Channels represent the dimensionality of the latent representation in the FNO network, and modes define the number of Fourier modes retained in each Fourier layer. Increasing the channel parameter generally increases the representational power of the model to learn more complex operators but can lead to issues of overfitting. There is no upper limit to the number of channels that can be used. Choosing the number of Fourier modes to retain largely depends on the smoothness of the PDE's solution. The maximum number of modes is defined by the size of the computational grid.

PDE solvers using Fourier methods assume a periodic boundary condition. Although the FNO network heavily uses the Fourier transform, it is not limited by this assumption and can be applied to solve PDEs with non-periodic boundary conditions, such as Burgers' equation and the Navier–Stokes equation [31]. This is important because the photoacoustic wave equation also has non-periodic boundary conditions.

The photoacoustic wave equation can be solved using either a 2D or 3D FNO architectural implementation. In the 2D architecture, the FNO network performs 2D convolutions in space and finds a solution for some fixed interval length Δt. The solution is then recurrently propagated in time and used to solve for the next interval length. In the 3D architecture, the FNO network performs 3D convolutions in space-time and can directly output the full time series solution with any time discretization. Both implementations were demonstrated to have similar performance. In this work, the 3D FNO network was used because it was found to be more expressive and easier to train [31].

2.4. Data Generation

The MATLAB toolbox k-Wave was used for photoacoustic wave simulation and to generate data for training and testing the FNO network [16]. The simulation medium was defined as a 64×64 computational grid, non-absorbing and homogenous with a speed of sound of 1480 m/s and density of 1000 kg/m^3. Forward simulations were performed with a time step of 20 ns for T = 151 steps. The initial photoacoustic pressure was initialized using anatomically realistic breast vasculature phantoms that were numerically generated [40]. The training dataset (n = 500) and testing dataset (n = 100) comprised images representing the initial photoacoustic pressure (the input to the FNO network) and the corresponding simulation of the photoacoustic wave propagation (output of the FNO network). Simulations for the Shepp–Logan, synthetic vasculature, tumor and Mason-M phantoms were also generated to evaluate the generalizability of the FNO network [16,41].

A second dataset was generated based on images from the Animals with Attributes 2 (AWA2) dataset originally developed for zero-shot image classification [42]. The AWA2 dataset has over 32,000 images categorized into 50 animal classes. This highly diverse dataset with many varied animals and backgrounds provides a more challenging task to train and evaluate the FNO. Forward simulations were performed with a similar medium as described previously, except with a 128×128 computational grid for T = 302 steps. With the same medium, the adjoint simulations were also performed using a 128-sensor linear array at the top of the computational grid. Images from 10 animal classes were used to create a training dataset (n = 1500). Images from 5 different animal classes were used to create a testing dataset (n = 500). The forward and adjoint simulations had their own respective datasets.

2.5. Model Training and Evaluation

The FNO network was implemented in PyTorch v1.7.1, a popular open-source deep learning library for Python [43]. The Adam optimizer with a mean squared error loss function was used to train the FNO network for 2000 epochs over approximately two days on an NVIDIA Tesla K80 GPU. The trained model was used to solve the wave equation for all time steps in a single forward pass.

The simulations from k-Wave served as the ground truth and were used to evaluate the quality of the FNO simulations. The root-mean-square error (RMSE) was used to quantitatively measure FNO simulation quality. Prior to calculating the RMSE, the k-Wave and FNO simulations were normalized to have values between 0 and 1 based on the peak value in the entire time series data. Normalization was applied to the entire time series and not to each individual time step. The RMSE was calculated at each time step and for the whole simulation.

For further validation, an in silico experiment of PAT imaging with a 64-sensor linear array was conducted. In general, any sensor array geometry could be used because the photoacoustic simulation and data sampling were independent events. A linear geometry was chosen because it is a widely available sensor array that is often used in experimental and clinical settings. Sensor data for the image reconstruction experiment were generated from the k-Wave and FNO network simulations by sampling the photoacoustic pressures at the top of the computational grid. The sampled time series sensor data were then used to reconstruct an image with the time reversal method [16].

The execution times to run the k-Wave and FNO simulations were measured on the same machine with an NVIDIA GeForce GTX 1080 Ti GPU and an Intel i7-12700K CPU. Simulations were repeated for 200 iterations, and the mean execution time was recorded. Torchinfo v1.6.5 was used to estimate the model size and GPU memory requirements.

3. Results

3.1. Breast Vasculature Simulation

An FNO network with the hyperparameters of 5 channels and 64 Fourier modes was trained using the breast vasculature dataset. The trained FNO network was used to predict and simulate photoacoustic wave propagation for $n = 100$ initial photoacoustic sources in the breast vasculature testing dataset. The photoacoustic wave simulations produced by the FNO network and k-Wave were remarkably similar and essentially identical to the naked eye (Figure 3). The FNO simulations successfully maintained the sharp edges and fine image details of the acoustic waves as they propagated throughout the medium. This demonstrated that the FNO network can model the broadband and high-frequency waves required for photoacoustic simulations.

Errors in the FNO network were quantitatively measured using the RMSE. The distribution of normalized photoacoustic pressures decreased as the simulation continued forward because energy dissipated within the medium as the acoustic waves propagated and exited the medium (Figure 4). For the testing dataset, the RMSE was several orders of magnitude smaller than the distribution of photoacoustic pressures in the simulations. This indicated that the errors in the FNO network simulations were small compared to the actual acoustic pressures, and the FNO network simulations were highly accurate approximations of the k-Wave simulations for photoacoustic wave propagation. Here, the RMSE was orders of magnitude smaller than the photoacoustic pressure distribution in the simulations, which can be attributed to the inherent model properties of globally learning frequency domain features and using the entire Fourier mode without any truncation. Therefore, the broadband frequency components originally distributed in the photoacoustic time series data can be described by sufficient frequency components.
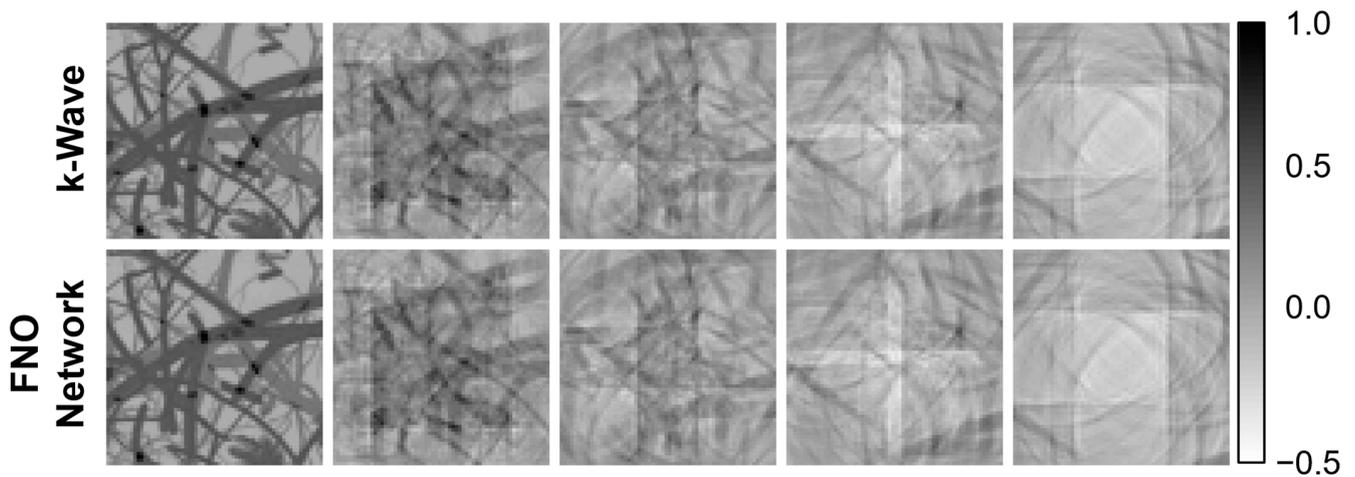
Figure 3. Visual comparison of the ground truth (**Top** Row) using k-Wave and the FNO network (**Bottom** Row) simulated photoacoustic wave propagation for an example vasculature image in a homogeneous medium at t = [1, 20, 40, 60, 80] time steps. The RMSE over all time steps was 3.8×10^{-3} for this example.
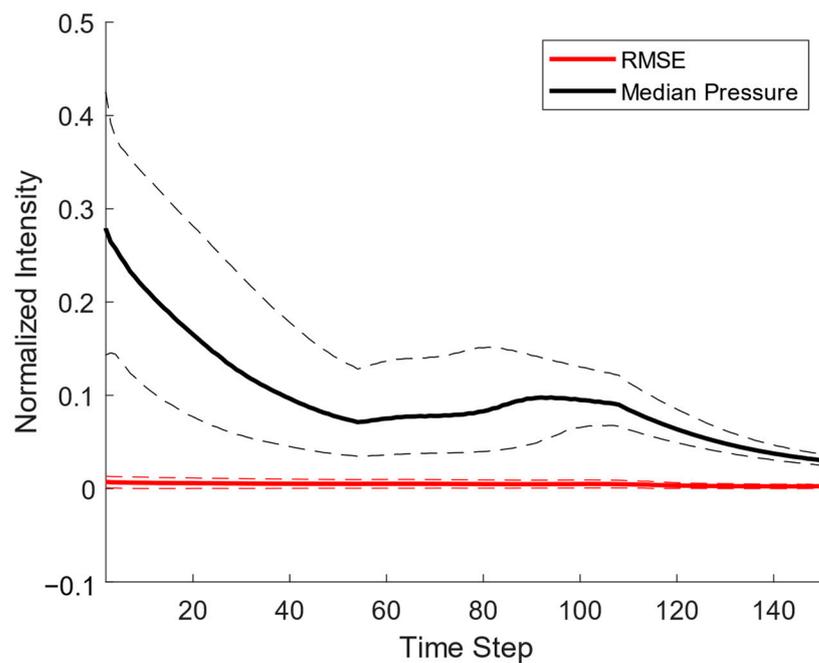

Figure 4. RMSE and standard deviation bands between the k-Wave and FNO simulations on the breast vasculature test dataset. Distribution of pressures (25th, 50th and 75th percentiles) are provided as a frame of reference.

A key advantage of data-driven PDE solvers over traditional ones is the vast reduction in computation time. In general, the time required to solve the photoacoustic wave equation largely depends on the discretization of the computational grid. For a computational grid of 64×64, k-Wave on average required 1.63 s to complete the simulation on a GPU. The FNO network on average required 0.061 s to complete it on the same GPU. This was approximately a $26\times$ reduction in computation time.

3.2. Image Reconstruction

The reconstructed PAT images from the FNO network and k-Wave simulations were almost visually identical (Figure 5). Vasculature structures and artifacts arising from the limited-view nature of the linear array can be seen in both the k-Wave and FNO network images. The reconstructed images were quantitatively compared using the RMSE and the structural similarity index metric (SSIM), a metric ranging from 0 to 1 that measures the similarity between two images based on factors relevant to human visual perception (e.g., structure, contrast and luminance) [44]. For the testing dataset ($n = 100$), the similarity between the FNO network and k-Wave images were measured with the RMSE ($7.1 \pm 1.5 \times 10^{-3}$), SSIM ($0.98 \pm 0.01$) and maximum error pixelwise (0.05 ± 0.01). These small errors and high similarity scores demonstrated that the time series sensor data produced using the FNO network and k-Wave simulations were highly similar.
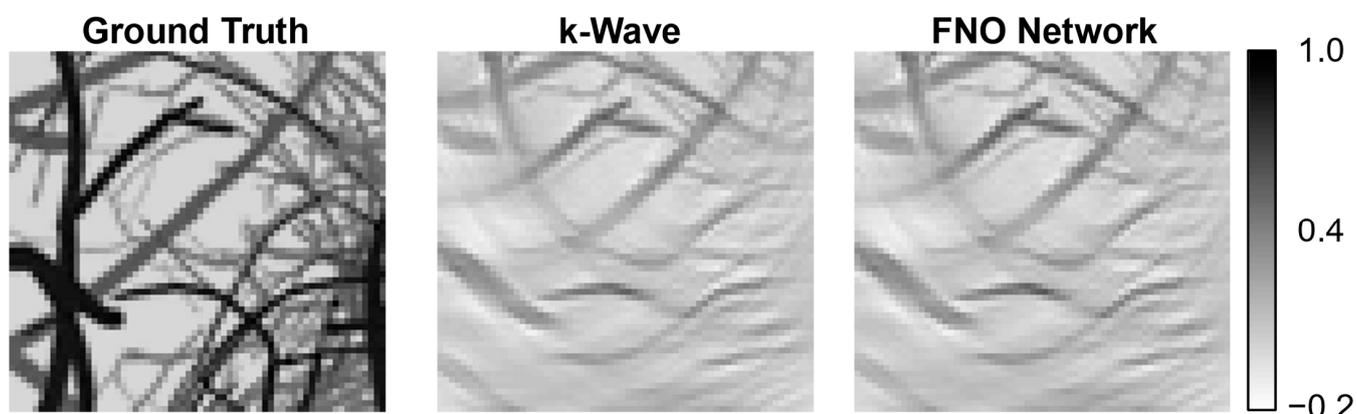

Figure 5. FNO and k-Wave were used to simulate wave propagation for the ground truth image. Sensor data were sampled from the resulting simulations and reconstructed into PAT images. The k-Wave and FNO images are almost identical. RMSE (0.011), SSIM (0.97) and max error (0.06).

3.3. Generalizability of Trained FNO Network

The FNO network was trained on photoacoustic simulations of breast vasculature. To evaluate its generalizability, the trained FNO network was used to simulate photoacoustic wave propagation with initial photoacoustic sources for the Shepp–Logan, synthetic vasculature, breast tumor and Mason-M logo phantoms. These phantoms contain many features not observed in the training dataset. For example, the breast vasculature phantoms typically occupy a majority of the space in the computational grid and have a mixture of large and small vessels, whereas the other phantoms occupy a fraction of the space and have small vessels or non-vasculature structures.

The FNO network and k-Wave simulations for each phantom tested were highly similar, but small visual differences could be observed (Figure 6). For example, the Mason-M logo has a mostly uniform grayscale background in the k-Wave simulation at t = 1, but a small gradient or shading could be seen in the FNO network simulation (average RMSEs across all time steps in the FNO network simulations: Shepp–Logan (9.5×10^{-3}), synthetic vasculature (1.5×10^{-2}), tumor (1.1×10^{-2}) and Mason-M (1.9×10^{-2}) phantoms). The Mason-M FNO simulation likely had the highest RMSE because it was a non-biological phantom unlike the other phantoms and the training dataset.

These results were promising and provided evidence that the trained FNO network is generalizable to other initial photoacoustic sources not in the training data. However, the FNO network did overfit the training data, as shown by the larger RMSE of the additional phantoms. Having a more diverse and larger training dataset can further improve the generalizability of the FNO network.
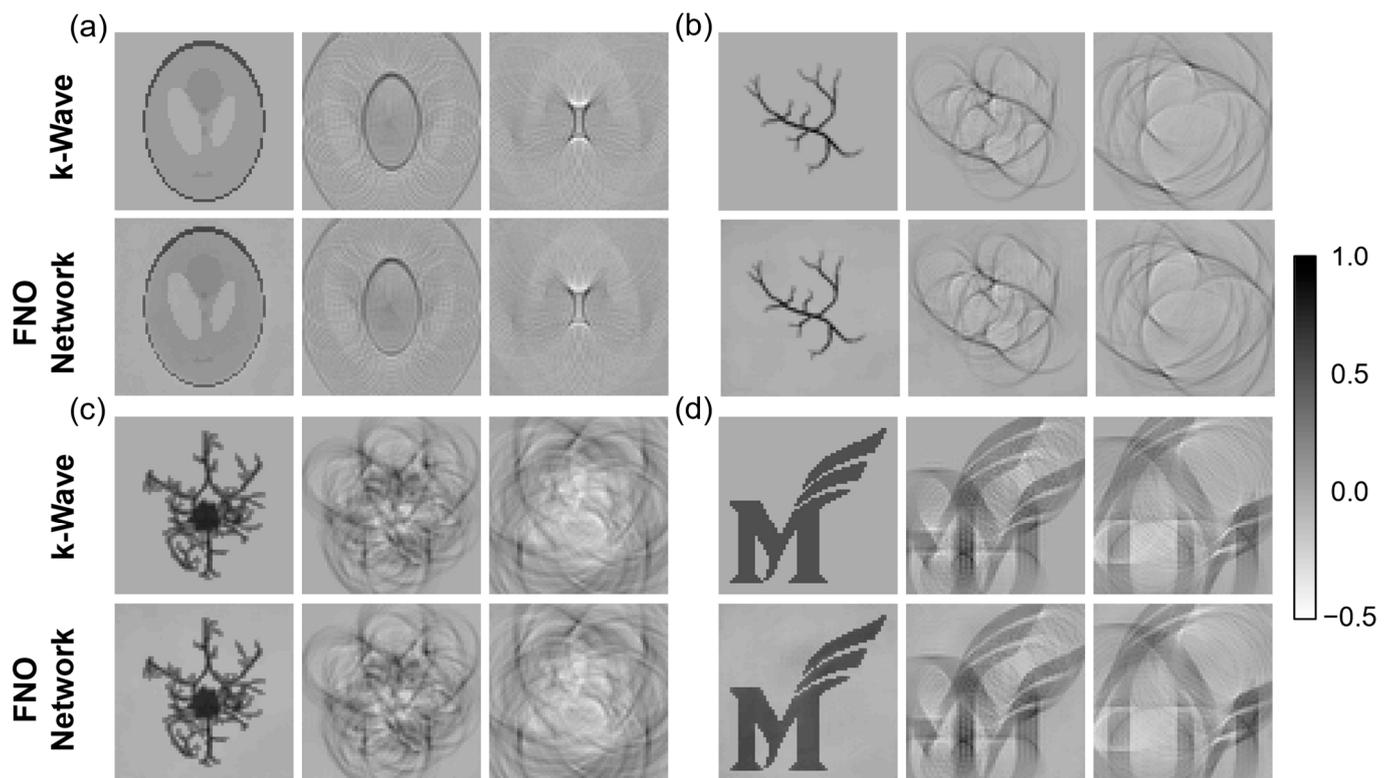
Figure 6. Comparison between FNO Network and k-Wave simulations for initial pressure sources using the (**a**) Shepp–Logan, (**b**) synthetic vasculature, (**c**) tumor and (**d**) Mason-M phantoms at t = [1, 10, 20] time steps.

3.4. Hyperparameter Optimization

A study was conducted to investigate the impact of hyperparameter selection on the FNO network's accuracy. All FNO networks were trained on the breast vasculature dataset for 200 epochs, which was sufficient for all networks to converge to a minimum loss. The number of Fourier modes had the largest impact on the FNO network because it directly affected the truncation error in the Fourier layers. FNO networks with fewer modes typically produced simulations with a blurred appearance (Figure 7). This was due to the loss of high-frequency information necessary for accurately simulating the sharp transitions of the acoustic wavefront. For a computational grid of 64×64, the FNO network with a maximum number of 64 modes produced the highest quality photoacoustic simulations.

Increasing the number of channels improved the FNO network's accuracy with diminishing returns (Table 1). Results show that there was little benefit in having an FNO network with more than five channels. Parameterizing the FNO network with a higher number of modes or channels results in a more complex model that requires more GPU memory. Interestingly, the time for a trained network to complete a simulation remained approximately the same. There was a small increase in computation time for the larger FNO networks with 64 modes and a higher number of channels.
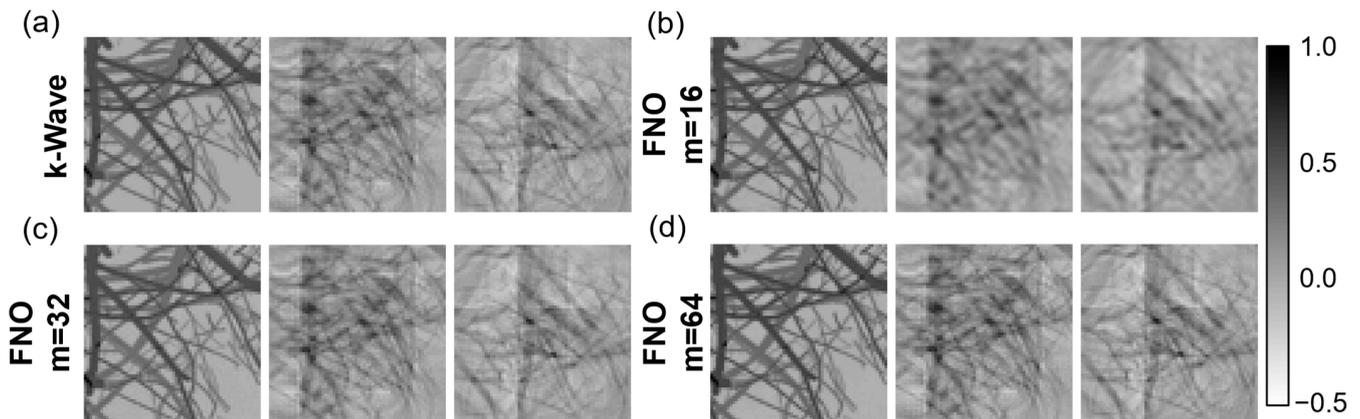
Figure 7. Comparison of FNO simulations at t = 1, 5, 10, 15 and 20 time steps. (**a**) k-Wave simulation. (**b**) The FNO network were parametrized with channels = 5 and modes = 16. (**c**) The FNO network were parametrized with channels = 5 and modes = 32. (**d**) The FNO network were parametrized with channels = 5 and modes = 64.

Table 1. Comparison of FNO network hyperparameters.

Grid Size	M	C	RMSE	Std. Dev.	Time (s)	GPU Mem. (GB)
$64 \times 64 \times 151$	16	5	2.4×10^{-2}	2.7×10^{-2}	0.016	0.85
$64 \times 64 \times 151$	32	5	8.5×10^{-3}	7.9×10^{-3}	0.021	0.94
$64 \times 64 \times 151$	64	5	4.7×10^{-3}	4.6×10^{-3}	0.061	1.68
$64 \times 64 \times 151$	64	2	1.3×10^{-2}	1.2×10^{-2}	0.048	0.84
$64 \times 64 \times 151$	64	3	8.5×10^{-3}	8.2×10^{-3}	0.051	1.05
$64 \times 64 \times 151$	64	4	5.4×10^{-3}	5.4×10^{-3}	0.055	1.33
$64 \times 64 \times 151$	64	5	4.7×10^{-3}	4.6×10^{-3}	0.061	1.68
$64 \times 64 \times 151$	64	6	5.1×10^{-3}	4.9×10^{-3}	0.066	2.09
$64 \times 64 \times 151$	64	7	4.4×10^{-3}	4.3×10^{-3}	0.074	2.57
$64 \times 64 \times 151$	64	8	5.1×10^{-3}	6.2×10^{-3}	0.082	3.12
$128 \times 128 \times 302$	64	5	6.9×10^{-3}	5.5×10^{-3}	0.110	6.45

Comparison between FNO networks with varying grid sizes, Fourier modes (M) and channels (C).

3.5. AWA2 Simulations

FNO networks with 5 channels and 64 Fourier modes were trained using the AWA2 datasets to perform the forward and adjoint simulations on a 128×128 computational grid. The simulations by the FNO network and k-Wave were visually almost identical (Figure 8). The RMSE in the forward simulation was consistently several orders of magnitude smaller than the normalized pressure across all time steps (Figure 9). The RMSE in the adjoint simulation increased over time but remained smaller than the normalized pressure (Figure 10). An increasing error was expected because the adjoint simulations began with a zero computational grid, and the pressure waves entered the grid from the top over time. The RMSE over the entire testing dataset ($n = 500$) was similarly small for the forward ($6.9 \pm 5.5 \times 10^{-3}$) and adjoint ($5.2 \pm 5.5 \times 10^{-3}$) simulations.
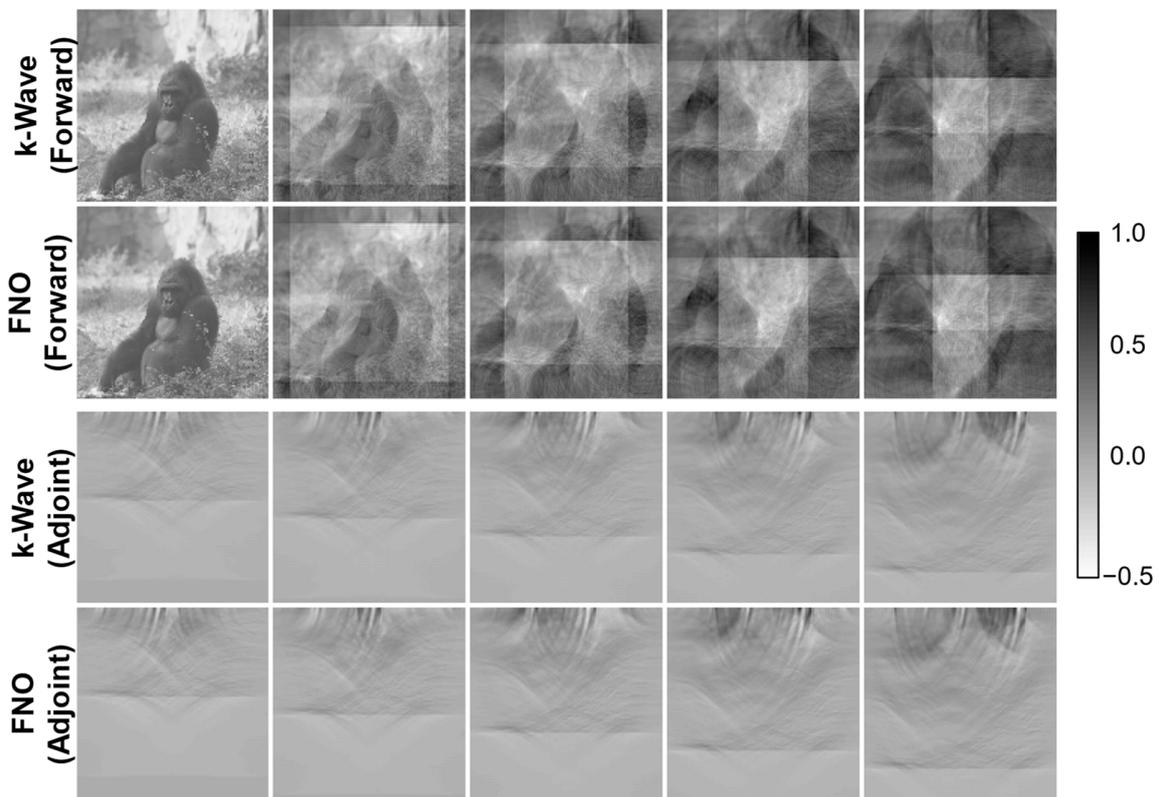
Figure 8. Simulations shown are for a 128 × 128 computation grid. (**Top**) Comparison between FNO and k-Wave forward simulations at t = [1, 40, 80, 120, 160] time steps. (**Bottom**) Comparison between FNO and k-Wave adjoint simulation t = [200, 220, 240, 260, 280] time steps. The input to both adjoint simulations was sensor data sampled with a linear array from the k-Wave forward simulation.
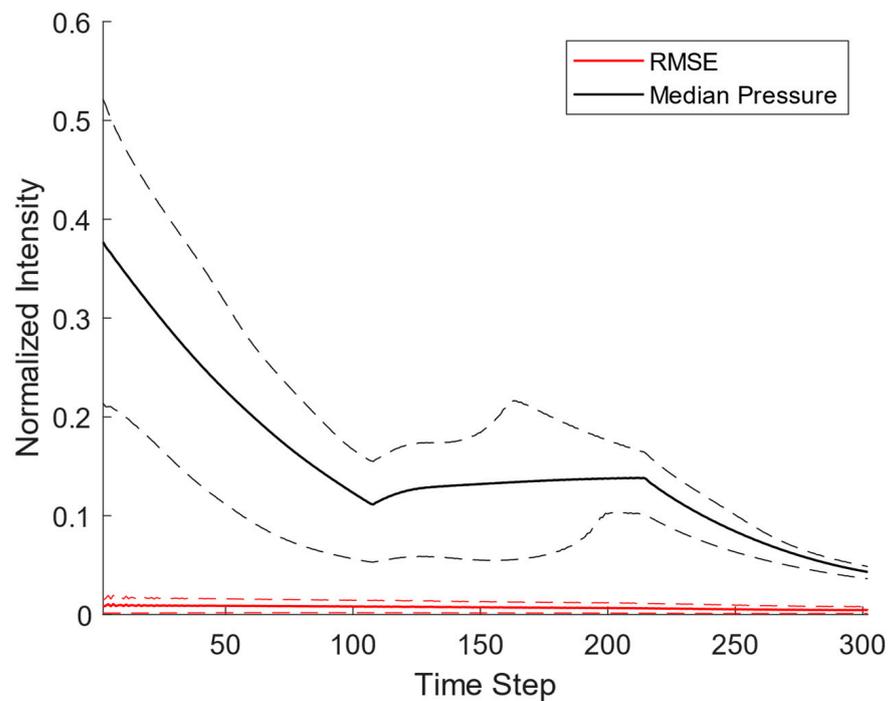

Figure 9. RMSE and standard deviation bands between the k-Wave and FNO forward simulations on the AWA2 test dataset. Distribution of pressures (25th, 50th and 75th percentiles) are provided as a frame of reference.
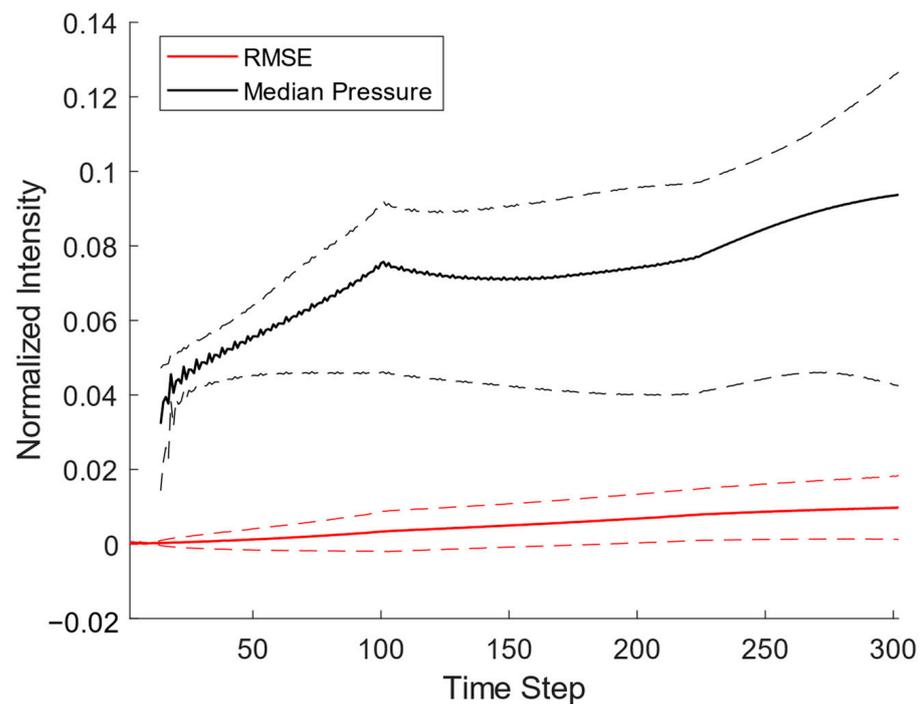
Figure 10. RMSE and standard deviation bands between the k-Wave and FNO adjoint simulations for the AWA2 test dataset. Distribution of pressures (25th, 50th and 75th percentiles) is provided as a frame of reference.

3.6. Zero-shot Super Resolution

The FNO was trained on simulations with a $128 \times 128 \times 302$ computational grid and was then used for zero-shot super resolution to simulate wave propagation on larger-sized grids. This was completed on an animal image from the testing dataset and a vasculature image to evaluate generalizability. The FNO simulations for both images strongly resembled their respective higher-resolution simulations using k-Wave (Figure 11). Errors could be seen in the background. The FNO simulations also had a more blurred appearance, as seen in the zoomed regions. This was likely due to the truncation of Fourier modes leading to the loss of high-frequency information that is useful for retaining sharper details in an image. Truncation errors became more prevalent for larger grid sizes using the FNO's super resolution feature (Table 2).

For a computational grid of $128 \times 128 \times 302$, k-Wave required on average 1.66 s to complete a simulation on a GPU. The FNO network on average required 0.11 s on the same GPU. This was approximately a $15\times$ reduction in computation time.

Table 2. FNO super resolution RMSE.

Grid Size	Animal	Vasculature
$128 \times 128 \times 302$	9.8×10^{-3}	2.7×10^{-3}
$152 \times 152 \times 302$	5.2×10^{-2}	3.4×10^{-2}
$176 \times 176 \times 302$	7.1×10^{-2}	4.5×10^{-2}
$200 \times 200 \times 302$	8.6×10^{-2}	5.5×10^{-2}
$224 \times 224 \times 302$	1.0×10^{-1}	6.3×10^{-2}

RMSE errors between k-Wave simulations and FNO zero-shot super resolution simulations for varying computational grid sizes.
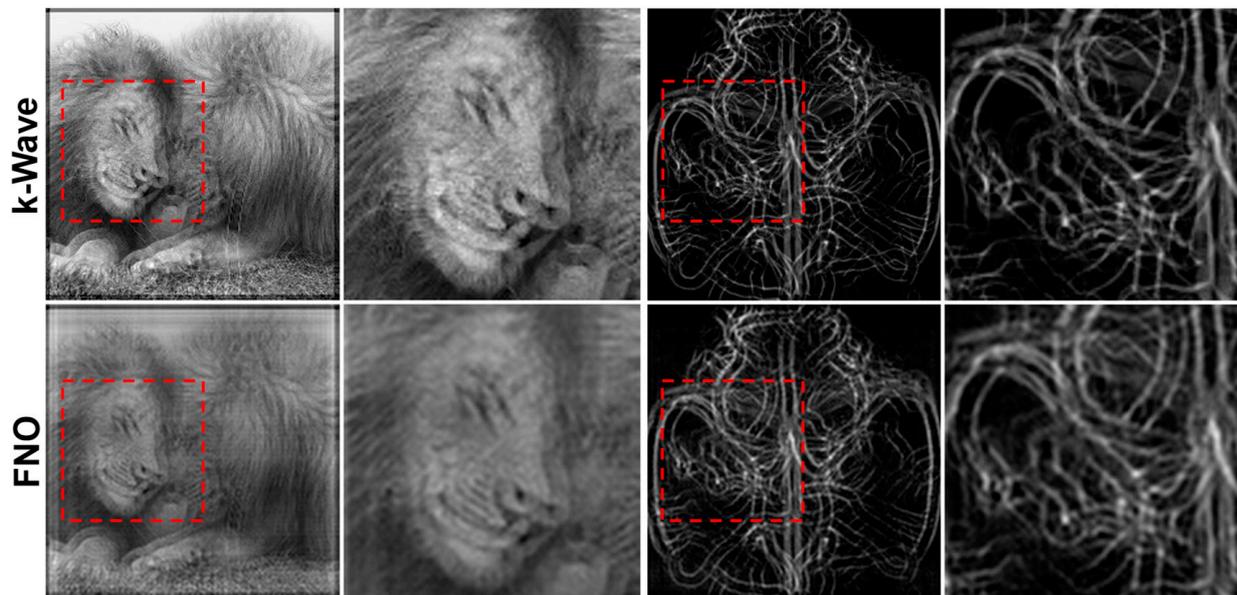
Figure 11. Zero-shot super resolution using an FNO trained on a $128 \times 128 \times 302$ computational grid to perform simulations on a $224 \times 224 \times 302$ grid. Images from the FNO and k-Wave simulations at t = 5 and zoomed regions are shown. (**Left**) Animal image of a lion from the testing dataset. (**Right**) Vasculature image to evaluate FNO generalizability.

4. Discussion

A key motivation for an FNO network is that a trained network can quickly produce accurate solutions for both forward and adjoint simulations. For a comparable simulation, the FNO network was ~26× faster on a 64×64 grid and ~15× faster on a 128×128 grid than k-Wave. This reduction in computation time with minimal loss in accuracy is ideal for applications requiring repeated evaluations of the photoacoustic operator. For example, advanced image reconstruction techniques, such as iterative methods, yield state-of-the-art results but are computationally expensive due to the repeated evaluation of the forward and adjoint operators. Therefore, the FNO network can greatly accelerate these methods by replacing traditional solvers for those operators.

The FNO network is parameterized by the number of Fourier modes and channels. Increasing either parameter typically improves model performance but at the cost of increased memory and computation requirements. PDEs with generally smoother solutions require fewer modes to achieve a satisfactory solution. However, the photoacoustic operator contains broadband frequency information, which means that a higher number of modes is needed for an accurate solution. The optimal number of Fourier modes retained is related to the spatial resolution of the initial photoacoustic source and not necessarily the size of the computational grid, meaning that not all simulations with a computational grid of 64×64 require the maximum 64 Fourier modes for the FNO network to produce an accurate solution. Hyperparameter optimization is especially important for simulations with large computational grids where limited GPU memory can become a problem.

Optimizing the FNO for larger computational grids can be partially addressed via hyperparameter selection. There are other approaches that can be employed to further reduce GPU memory limitations. In this work, a 3D CNN architecture was used, but there are other CNN architectures, such as the recurrent 2D network, that are more memory-efficient [31]. Instead of solving for the full temporal solution in a single step, the FNO can be used iteratively to solve the wave equation for *n* time steps in each forward pass over the full time interval. Conventional solvers typically require sufficiently small time steps for solution stability, but the FNO is likely not limited by this requirement. Thus, it is possible that the FNO can use larger time steps and obtain an accurate solution for downstream image reconstruction and processing tasks.

In this work, we chose to simulate the full spatio-temporal wave field to decouple the tasks of simulation from sampling. This allowed the same trained FNO to be used for any sensor configuration. The FNO could be configured and trained to directly output the sampled sensor data. This would reduce memory requirements but would also need to be retrained for each sensor geometry and configuration.

A practical limitation in data-driven PDE solvers, such as the FNO network, is the need for high-quality training data. Traditional solvers are often used to create arbitrarily large datasets to train the network. Depending on the size of the computational grid, this can be computationally formidable, such as the case of 3D photoacoustic simulations. To create a large dataset in these scenarios, a high-performance computing environment would be needed to generate the training data in a reasonable timeframe. Transfer learning could also be employed to partially address this challenge after training the initial FNO model. Rather than training from scratch, a pre-trained model can be fine-tuned with smaller datasets for specific tasks or simulation environments. This would make the model more accessible for other downstream users.

Zero-shot super resolution is a unique feature to the FNO and has been previously demonstrated to work well for the Navier–Stokes equation. Given the difficulty of zero-shot tasks, it is highly encouraging that the FNO could generate larger-sized simulations resembling the k-Wave simulations. However, the truncation of Fourier modes led to blurring and errors for larger computation grids. For example, the computational grid of 224×224 produced 224 Fourier modes, but only 64 were retained. The information loss in that scenario was greater than that in the case of 128×128 computational grid.

5. Conclusions

Solving the 2D photoacoustic wave equation with traditional methods typically requires a fine discretization of the computational grid and can be time-consuming to complete. Deep learning methods directly learn from data to solve PDEs and can be orders of magnitude faster with minimal losses in accuracy. In this work, we applied the FNO network as a fast data-driven PDE solver for the 2D photoacoustic wave equation in a homogeneous medium. The photoacoustic simulations generated by a traditional solver with the k-Wave toolbox and the FNO network for the breast vasculature testing dataset were remarkably similar, both visually and quantitatively. The RMSE between the k-Wave and FNO simulations was several orders of magnitude smaller than the pressure intensities.

Model generalizability is a highly desirable property for an FNO network. If the trained FNO network is to be a reliable alternative to traditional PDE solvers, then it needs to be capable of solving the photoacoustic operator for any arbitrary initial pressure source within an acceptable degree of error. Having a generalizable network minimizes the need to retrain the network for instances not observed in the training dataset. The generalizability of the trained FNO network was evaluated using phantoms not in the training data. Shepp–Logan, synthetic vasculature, tumor and Mason-M phantoms were used for 64×64 simulations, and the vasculature phantom was used for the 128×128 simulation. In general, the FNO network and k-Wave simulations were visually similar. The RMSE was relatively small for these phantoms but was larger than those in the testing dataset. This indicated that the FNO was overfitting the training data. These results provide evidence for the FNO network being generalizable and can be used for simulations with any arbitrary initial pressure source. Moreover, the FNO network was learning the photoacoustic operator and not memorizing specific solutions related to the training data.

In this work, the FNO network was trained for solving the 2D acoustic wave equation in a homogeneous medium. Simulations with homogeneous media are widely used in many applications because the spatial distribution of heterogeneities is often unknown. Nevertheless, an FNO network can be used for simulations with heterogeneous media by providing the spatial distribution of medium properties as additional inputs to the FNO network. During the training process, the FNO network can leverage these inputs and learn from training examples generated with varying heterogeneous media to perform

simulations with heterogeneous media. The operator for a heterogeneous medium is more complex than that of the homogeneous case. Thus, a larger and more diverse dataset is likely needed to adequately train the FNO network.

The significance of our work is that it provides a comprehensive study on solving the 2D photoacoustic wave equation with an FNO, demonstrating its application potential in downstream reconstruction and image-processing tasks. The FNO performs orders of magnitude faster than traditional solvers and can generalize to unseen out-of-domain datasets. However, for simulations in a larger computational grid, a higher Fourier mode is necessarily required to learn the broadband frequency residing in the whole time series data. Consequently, it increases GPU memory and can be an intractable problem while simulating in a larger domain, which usually requires the output of more time steps. Furthermore, the current framework is limited to simulating the 2D photoacoustic wave equation. Simulating the 3D photoacoustic source to output entire 4D time series data is impossible to handle with modern GPUs. In addition, zero-shot super resolution using an FNO tends to output blurry images due to the insufficient number of Fourier modes used to train the network to support broadband frequency components in a larger computational grid. In the future, research addressing these issues will benefit the use of FNOs in more complex settings.

Author Contributions: Conceptualization, S.G.; methodology, S.G.; validation, S.G. and K.-T.H.; writing—original draft preparation, S.G.; writing—review and editing, S.G., K.-T.H. and P.V.C.; visualization, S.G.; supervision, P.V.C. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Animals with attributes 2 (AWA2) dataset is available at https://cvml.ista.ac.at/AwA2/. Software for generating breast phantom dataset is available at https://breastphantom.readthedocs.io/en/latest/.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Xia, J.; Yao, J.; Wang, L.V. Photoacoustic tomography: Principles and advances. *Electromagn. Waves Camb. Mass* **2014**, *147*, 1–22. [CrossRef] [PubMed]
2. Xia, J.; Wang, L.V. Small-animal whole-body photoacoustic tomography: A review. *IEEE Trans. Biomed. Eng.* **2014**, *61*, 1380–1389. [CrossRef] [PubMed]
3. Nyayapathi, N.; Xia, J. Photoacoustic imaging of breast cancer: A mini review of system design and image features. *J. Biomed. Opt.* **2019**, *24*, 121911. [CrossRef]
4. Bungart, B.L.; Lan, L.; Wang, P.; Li, R.; Koch, M.O.; Cheng, L.; Masterson, T.A.; Dundar, M.; Cheng, J.-X. Photoacoustic tomography of intact human prostates and vascular texture analysis identify prostate cancer biopsy targets. *Photoacoustics* **2018**, *11*, 46–55. [CrossRef] [PubMed]
5. Moore, C.; Jokerst, J.V. Strategies for Image-Guided Therapy, Surgery, and Drug Delivery Using Photoacoustic Imaging. *Theranostics* **2019**, *9*, 1550–1571. [CrossRef]
6. Rao, A.P.; Bokde, N.; Sinha, S. Photoacoustic Imaging for Management of Breast Cancer: A Literature Review and Future Perspectives. *Appl. Sci.* **2020**, *10*, 767. [CrossRef]
7. Lin, L.; Hu, P.; Shi, J.; Appleton, C.M.; Maslov, K.; Li, L.; Zhang, R.; Wang, L.V. Single-breath-hold photoacoustic computed tomography of the breast. *Nat. Commun.* **2018**, *9*, 2352. [CrossRef]
8. Li, M.; Tang, Y.; Yao, J. Photoacoustic tomography of blood oxygenation: A mini review. *Photoacoustics* **2018**, *10*, 65–73. [CrossRef]

9. Goh, Y.; Balasundaram, G.; Tan, H.M.; Putti, T.C.; Tang, S.W.; Ng, C.W.Q.; Buhari, S.A.; Fang, E.; Moothanchery, M.; Bi, R.; et al. Biochemical "decoding" of breast ultrasound images with optoacoustic tomography fusion: First-in-human display of lipid and collagen signals on breast ultrasound. *Photoacoustics* **2022**, *27*, 100377. [CrossRef]
10. Wang, L.V. Prospects of photoacoustic tomography. *Med. Phys.* **2008**, *35*, 5758–5767. [CrossRef]
11. Li, S.; Montcel, B.; Liu, W.; Vray, D. Analytical model of optical fluence inside multiple cylindrical inhomogeneities embedded in an otherwise homogeneous turbid medium for quantitative photoacoustic imaging. *Opt. Express* **2014**, *22*, 20500–20514. [CrossRef]
12. Xu, M.; Wang, L.V. Universal back-projection algorithm for photoacoustic computed tomography. *Phys. Rev. E* **2005**, *71*, 016706. [CrossRef]
13. Hristova, Y.; Kuchment, P.; Nguyen, L. Reconstruction and time reversal in thermoacoustic tomography in acoustically homogeneous and inhomogeneous media. *Inverse Probl.* **2008**, *24*, 055006. [CrossRef]
14. Huang, C.; Wang, K.; Schoonover, R.W.; Wang, L.V.; Anastasio, M.A. Joint Reconstruction of Absorbed Optical Energy Density and Sound Speed Distributions in Photoacoustic Computed Tomography: A Numerical Investigation. *IEEE Trans. Comput. Imaging* **2016**, *2*, 136–149. [CrossRef]
15. Cox, B.T.; Treeby, B.E. Artifact Trapping During Time Reversal Photoacoustic Imaging for Acoustically Heterogeneous Media. *IEEE Trans. Med. Imaging* **2010**, *29*, 387–396. [CrossRef] [PubMed]
16. Treeby, B.E.; Cox, B.T. k-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields. *J. Biomed. Opt.* **2010**, *15*, 021314. [CrossRef] [PubMed]
17. Xu, M.; Xu, Y.; Wang, L.V. Time-domain reconstruction algorithms and numerical simulations for thermoacoustic tomography in various geometries. *IEEE Trans. Biomed. Eng.* **2003**, *50*, 1086–1099. [CrossRef] [PubMed]
18. Paltauf, G.; Viator, J.A.; Prahl, S.A.; Jacques, S.L. Iterative reconstruction algorithm for optoacoustic imaging. *J. Acoust. Soc. Am.* **2002**, *112*, 1536–1544. [CrossRef]
19. Guan, S.; Khan, A.; Sikdar, S.; Chitnis, P. Fully Dense UNet for 2D Sparse Photoacoustic Tomography Artifact Removal. *IEEE J. Biomed. Health Inform.* **2019**, *24*, 568–576. [CrossRef] [PubMed]
20. Guan, S.; Khan, A.A.; Sikdar, S.; Chitnis, P.V. Limited-View and Sparse Photoacoustic Tomography for Neuroimaging with Deep Learning. *Sci. Rep.* **2020**, *10*, 8510. [CrossRef]
21. Antholzer, S.; Haltmeier, M.; Schwab, J. Deep learning for photoacoustic tomography from sparse data. *Inverse Probl. Sci. Eng.* **2019**, *27*, 987–1005. [CrossRef]
22. Hauptmann, A.; Lucka, F.; Betcke, M.M.; Huynh, N.; Adler, J.; Cox, B.T.; Beard, P.C.; Ourselin, S.; Arridge, S.R. Model-Based Learning for Accelerated, Limited-View 3-D Photoacoustic Tomography. *IEEE Trans. Med. Imaging* **2018**, *37*, 1382–1393. [CrossRef]
23. Baumann, B.; Wolff, M.; Kost, B.; Groninga, H. Finite element calculation of photoacoustic signals. *Appl. Opt.* **2007**, *46*, 1120–1125. [CrossRef] [PubMed]
24. Xia, W.; Piras, D.; van Hespen, J.C.G.; van Veldhoven, S.; Prins, C.; van Leeuwen, T.G.; Steenbergen, W.; Manohar, S. An optimized ultrasound detector for photoacoustic breast tomography. *Med. Phys.* **2013**, *40*, 032901. [CrossRef]
25. Raissi, M.; Perdikaris, P.; Karniadakis, G.E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **2019**, *378*, 686–707. [CrossRef]
26. Greenfeld, D.; Galun, M.; Basri, R.; Yavneh, I.; Kimmel, R. Learning to Optimize Multigrid PDE Solvers, in International Conference on Machine Learning, May 2019, pp. 2415–2423. Available online: http://proceedings.mlr.press/v97/greenfeld19a.html (accessed on 22 May 2021).
27. Khoo, Y.; Lu, J.; Ying, L. Solving parametric PDE problems with artificial neural networks. *Eur. J. Appl. Math.* **2021**, *32*, 421–435. [CrossRef]
28. Adler, J.; Öktem, O. Solving ill-posed inverse problems using iterative deep neural networks. *Inverse Probl.* **2017**, *33*, 124007. [CrossRef]
29. E, W.; Yu, B. The Deep Ritz Method: A Deep Learning-Based Numerical Algorithm for Solving Variational Problems, ArXiv171000211 Cs Stat, September 2017. Available online: http://arxiv.org/abs/1710.00211 (accessed on 22 May 2021).
30. Lu, L.; Jin, P.; Karniadakis, G.E. DeepONet: Learning Nonlinear Operators for Identifying Differential Equations Based on the Universal Approximation Theorem of Operators, ArXiv191003193 Cs Stat, April 2020. Available online: http://arxiv.org/abs/1910.03193 (accessed on 22 May 2021).
31. Li, Z.; Kovachki, N.; Azizzadenesheli, K.; Liu, B.; Bhattacharya, K.; Stuart, A.; Anandkumar, A. Fourier Neural Operator for Parametric Partial Differential Equations, ArXiv201008895 Cs Math, October 2020. Available online: http://arxiv.org/abs/2010.08895 (accessed on 29 December 2020).
32. Beard, P. Biomedical photoacoustic imaging. *Interface Focus* **2011**, *1*, 602–631. [CrossRef] [PubMed]
33. Xu, M.; Wang, L.V. Universal back-projection algorithm for photoacoustic computed tomography. *Int. Soc. Opt. Photonics* **2005**, *5697*, 251–255. [CrossRef]
34. Treeby, B.E.; Zhang, E.Z.; Cox, B.T. Photoacoustic tomography in absorbing acoustic media using time reversal. *Inverse Probl.* **2010**, *26*, 115003. [CrossRef]
35. Piaggio, H.T.H. The Mathematical Theory of Huygens' Principle. *Nature* **1940**, *145*, 531–532. [CrossRef]
36. Tadmor, E. A review of numerical methods for nonlinear partial differential equations. *Bull. Am. Math. Soc.* **2012**, *49*, 507–554. [CrossRef]

37. Treeby, B.E.; Pan, J. A practical examination of the errors arising in the direct collocation boundary element method for acoustic scattering. *Eng. Anal. Bound. Elem.* **2009**, *33*, 1302–1315. [CrossRef]

38. Mast, T.D.; Souriau, L.P.; Liu, D.-L.D.; Tabei, M.; Nachman, A.I.; Waag, R.C. A k-space method for large-scale models of wave propagation in tissue. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **2001**, *48*, 341–354. [CrossRef] [PubMed]

39. Treeby, B.E.; Jaros, J.; Rendell, A.P.; Cox, B.T. Modeling nonlinear ultrasound propagation in heterogeneous media with power law absorption using a k-space pseudospectral method. *J. Acoust. Soc. Am.* **2012**, *131*, 4324–4336. [CrossRef] [PubMed]

40. Badano, A.; Graff, C.G.; Badal, A.; Sharma, D.; Zeng, R.; Samuelson, F.W.; Glick, S.J.; Myers, K.J. Evaluation of Digital Breast Tomosynthesis as Replacement of Full-Field Digital Mammography Using an In Silico Imaging Trial. *JAMA Netw. Open* **2018**, *1*, e185474. [CrossRef] [PubMed]

41. Arridge, S.; Beard, P.; Betcke, M.; Cox, B.; Huynh, N.; Lucka, F.; Ogunlade, O.; Zhang, E. Accelerated high-resolution photoacoustic tomography via compressed sensing. *Phys. Med. Biol.* **2016**, *61*, 8908. [CrossRef]

42. Xian, Y.; Lampert, C.H.; Schiele, B.; Akata, Z. Zero-Shot Learning—A Comprehensive Evaluation of the Good, the Bad and the Ugly. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *41*, 2251–2265. [CrossRef]

43. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. *arXiv* **2019**, arXiv:1912.01703.

44. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef]