IOP Publishing Smart Materials and Structures
Smart Mater. Struct. **34** (2025) 035059 (17pp) https://doi.org/10.1088/1361-665X/adc359

A deep learning-based approach to improve reconstruction of ultrasound computed tomography with full waveform inversion

Shoaib Anwar[1], Austin Yunker[2], Rajkumar Kettimuthu[2,3], Mark A Anastasio[4], Zhengchun Liu[2], Weihua Su[1,*] and Jiaze He[1,*]

[1] Department of Aerospace Engineering and Mechanics, The University of Alabama, Tuscaloosa, AL 35401, United States of America
[2] Argonne National Laboratory, Lemont, IL 60439, United States of America
[3] Consortium for Advanced Science and Engineering, University of Chicago, Chicago, IL 60637, United States of America
[4] Department of Bioengineering, University of Illinois Urbana-Champaign, Urbana, IL 61801, United States of America

E-mail: suw@eng.ua.edu and jiazehe@gmail.com

Received 6 December 2024, revised 17 February 2025
Accepted for publication 20 March 2025
Published 28 March 2025

Abstract

Enabled by multiple modalities of smart materials-based actuation and sensing, full waveform inversion (FWI) nowadays is an advanced ultrasound computed tomography technique that utilizes waveform data to generate high-resolution images of scanned regions. This technology offers promise for defect/damage detection and disease diagnosis, showing potential in nondestructive testing, structural health monitoring, and medical imaging. To reduce the lengthy computational time caused by time-domain FWI, modern AI-driven and data-driven approaches have been studied to accelerate the reconstruction process in recent years. However, most existing research focused on tuning specific neural networks for fixed-domain applications, leaving the relationship between model performance and characteristics of various training datasets underexplored. This paper presents a comprehensive investigation into the improvements achievable by integrating deep learning with the adjoint tomography theory, addressing the scientific questions of how amounts/distributions of the training data, augmentation, and loss functions influence the efficacy of this approach. The selected integration strategy involves training a U-Net neural network model using pairs of low-resolution inverted images and their corresponding high-resolution ground truth images. Once trained, the U-Net model instantaneously converts low-resolution inference images to high-resolution reconstructions. Generated using a proposed high-performance computing-based framework, multiple datasets were designed to offer a general representation of various applications while maintaining the shared characteristics across different use cases. The study also incorporates augmentation strategies to expand the size and complexity of the training dataset without significantly increasing the number of samples. Furthermore, a hyperparameter tuning framework was introduced to investigate the impact of multiple loss functions on the model

* Authors to whom any correspondence should be addressed.

performance. An additional challenge in data-driven approaches is the generalizability of the neural network model when exposed to out-of-distribution data. This study rigorously tests the model's generalizability against several out-of-distribution datasets and finds that the U-Net model maintains a degree of generalizability when trained with unscaled datasets.

Keywords: ultrasound computed tomography, full waveform inversion, deep learning, convolution neural network, reconstruction improvement

1. Introduction

Ultrasound waves exhibit high sensitivity to mechanical property variations, including alterations in wave speed within a given structure [1–3]. This inherent sensitivity has rendered ultrasound an invaluable tool for detecting, characterizing, and evaluating inhomogeneities, defects, and material properties within diverse scanned regions [4–7]. Moreover, the advantages of ultrasound testing extend beyond its effectiveness, as it offers distinct benefits of reduced time and cost when compared to experimental data acquisition [8]. Consequently, ultrasound testing has gained significant traction in the fields of structural health monitoring (SHM), nondestructive testing (NDT), and medical imaging [9–12]. To convert ultrasound signals into meaningful images various imaging techniques are employed, e.g, reverse time migration (RTM) [13], full waveform inversion (FWI) [14, 15], and total focusing method [16]. This paper focuses on the FWI technique with a modern AI-based approach to improve the quality of imaging and accelerate the process.

Conventional ultrasound imaging often lacks quantitative representation of objects because wave equations were not utilized in imaging. Unlike most other ultrasound imaging techniques, FWI stands out for its ability to fully exploit the waveform information contained within ultrasound signals [17, 18]. FWI is an inverse process designed to quantitatively determine the properties (model parameters) of a target specimen by minimizing the discrepancy (misfit) between the acquired waveform signals and the synthetic waveform signals [19]. The centerpiece of FWI is the adjoint method, which enables the calculation of cross-correlations between forward and back-propagated waveforms, facilitating the formulation of an iterative optimization problem to determine the optimal model parameters [20, 21]. As a result, FWI is capable of producing high-resolution inverted images of the specimen. FWI-based ultrasound computed tomography (USCT) has become a promising technique for material characterization in industries related to SHM and NDT [22].

Currently, the majority of FWI applications are situated in the time domain, in contrast to the frequency domain, largely due to the computational challenges associated with frequency domain FWI (FDFWI) involving large, sparse, indefinite linear systems per frequency [23]. Consequently, it can use a few frequencies while the time domain FWI (TDFWI) simultaneously simulates all relevant frequencies and hence produces high-resolution images [24, 25]. Because of this constraint, it is more challenging to implement FDFWI in a very complex and large system, e.g. a 3D case. Hence, this study focuses on TDFWI only. However, both domains have multiparameter inversion ability [26, 27], which has made FWI favorable for material characterization [28], reconstructing structures with defects [15, 29], and polycrystalline grain distribution [30]. Furthermore, FWI results with low-frequency data can be exploited with other imaging modalities e.g. high-frequency RTM for defect imaging in complex heterogenous structures [31]. Beyond the SHM-NDT industries, FWI has been explored for more advanced applications in medical imaging [32–34]. Some examples of this application can be found in the imaging of breast [32, 35], bones [36, 37], and the human brain [22].

Although FWI offers advantages in model reconstruction, it also presents certain challenges that need to be addressed. These challenges include extreme non-linearity, ill-posed problems, and the requirement for a good starting model [18, 38]. Furthermore, the mathematical complexity of solving inverse problems also makes FWI an intensive task. Consequently, FWI is computationally demanding, requiring significant computational resources and time [18]. To mitigate these challenges and enhance the efficiency of FWI, an alternative approach using Artificial Intelligence (AI) can be employed.

In parallel with the advancement of computational powers, a data-driven AI-based approach has been investigated in numerous image-based inverse problems in tomographic imaging (e.g. selection of a good starting model, high computational cost, local minima, and cycle-skipping) [39, 40]. These approaches can be broadly categorized into two groups: (1) 'end-to-end' such as convolution neural networks (CNN), recurrent neural networks (RNN), generative adversarial networks (GAN), and (2) physics-informed neural networks (PINN). A recent study on inversion with a supervised deep learning technique combined CNN to provide a good initial model and optimize the inversion results [41]. Robins *et al* [42] used 2D CNN (U-Net model) to extrapolate band-limited signals to include additional lower-frequency content to alleviate the cycle-skipping effects [42]. To mitigate the computational burden of FWI during inference, a DL model can be trained with low-resolution inverted images at early iterations to convert them into high-resolution images. Kleman *et al* [43] applied a similar technique with single-iteration

inverted images to get high-resolution images, hence accelerating the TDFWI process. Recent studies have explored the integration of GANs with seismic FWI [44] and similar data-driven inversion methods for problems like seismic impedance inversion [45]. Zhang *et al* proposed a GAN-based model for velocity estimation from raw seismic waveform data [46]. The application of RNNs in FWI for elastic and transversely isotropic media has also demonstrated significant potential [47, 48]. PINNs approximate the solution of a partial differential equation (PDE) by training on control points and enforcing constraints on the solution's derivatives to satisfy the underlying PDE [49]. In contrast, RNN-based FWI involves computing partial derivatives using finite differences and utilizing automatic differentiation solely for model parameter updates. However, the RNN method necessitates storage for the entire multi-component wavefield [50].

However, a vast majority of current DL-related advancements have not been explored for FWI-based USCT. Prior to the application of this technique to real-world data, it is imperative to establish a fundamental understanding of the concept of combining DL methodology and FWI-USCT. This paper focuses on the improvements achievable with the integration of deep learning with the adjoint tomography theory, addressing the scientific questions of how amounts/distributions of the training data, augmentation, and loss functions influence the efficacy of this approach. Given the computationally intensive nature of TDFWI, generating training datasets can be highly time-consuming. To address this challenge, a novel and efficient high-performance computing (HPC) framework is proposed to accelerate dataset creation. Multiple datasets of varying sizes (1k and 25k) were generated from different TDFWI iterations (1–10) using this framework to systematically analyze the impact of dataset size and iteration count on predictive accuracy. Additionally, various data augmentation strategies were employed to enhance material distribution complexity, expand dataset size, and assess their influence on DL-based FWI solutions. A multi-loss function strategy was implemented to further enhance the model's predictive capabilities. Inspired by Liu *et al* [51], who improved DL-based x-ray computed tomography using a combination of adversarial loss, MSE-loss, perceptual loss, and gradient difference loss, this study employed a similar approach in the U-Net model, excluding the adversarial loss function, to evaluate its impact on FWI-USCT. Furthermore, the model's generalizability was tested using 'out-of-distribution' (OOD) samples with material properties outside the training dataset range.

The remainder of the paper is arranged as follows. First, a brief theoretical background of TDFWI and the relation between TDFWI and DL techniques are drawn out in the methodology section. This is followed by the implementation, data, and model section wherein the numerical samples, HPC-enabled framework of data generation, data processing techniques, augmentation strategies, and neural networks are introduced. The results of different experiments are then detailed in the results and discussion sections, which are followed by the conclusion.

2. Methodology

FWI is a non-linear inversion method. Figure 1 illustrates the process. TDFWI entails solving a forward problem of a time-dependent wave equation to simulate wave propagation from an ultrasound signal generator (source) to signal receivers, followed by an inverse problem of solving a PDE constraint optimization problem [20]. It aims to generate a model of subsurface properties that minimize the gradient of the misfit function between observed and modeled data [15]. The first step is to create an initial model and sources. The wave propagation in this initial model is solved by using a 2D spectral element solver (i.e. SPECFEM2D) to create the synthetic data. This process is called forward simulation [15]. In numerical studies, forward simulation is also performed for the ground truth model (GTM) to create the observed data. Initially, the disparity or misfit between these two datasets is expected to be significant. To update the synthetic model, an adjoint simulation is performed. This procedure requires introducing adjoint sources as source time functions at the receiver locations, back-projecting the measured data difference, and optimizing the synthetic model. The interaction between forward and adjoint wavefields determines the model parameters (e.g. wave speed). This process is repeated until the misfit decreases below a specific value, where a high-resolution model is reconstructed. In this study, a customized version of SeisFlows (a Python-based framework) was used to implement TDFWI and the corresponding deep-learning data generation. The following subsections briefly discuss the underlying mathematical background behind TDFWI.

2.1. Brief description of the mathematics behind TDFWI

For the acoustic case, the displacement wavefield $\mathbf{s}(\mathbf{x},\mathrm{t})$ can be expressed as

$$\rho\partial_t^2\mathbf{s} = -\nabla p + \mathbf{f}, \tag{1}$$

where p denotes pressure, ρ is mass density, and $\mathbf{f}$ is the sources as excitation forces. Luo [52] introduced a scalar potential ϕ to define the displacement as $\mathbf{s} \equiv \frac{1}{\rho}\nabla\phi$ and a scalar f such that $\mathbf{f} = \nabla(\kappa f)$, κ is bulk modulus. The relation between p and $\mathbf{s}$ can be drawn through the continuity equation of acoustic wave propagation as

$$p = -\kappa(\nabla\cdot\mathbf{s}) = -\kappa\left(\nabla\cdot\left(\frac{1}{\rho}\nabla\phi\right)\right). \tag{2}$$

Hence, equation (1) can be written in terms of ϕ, given by

$$\frac{1}{\kappa}\partial_t^2\phi = \nabla\cdot\left(\frac{1}{\rho}\nabla\phi\right) + f. \tag{3}$$

For waveform tomography, Tarantola *et al* [17] introduced the least-squares waveform misfit function, which can be defined as

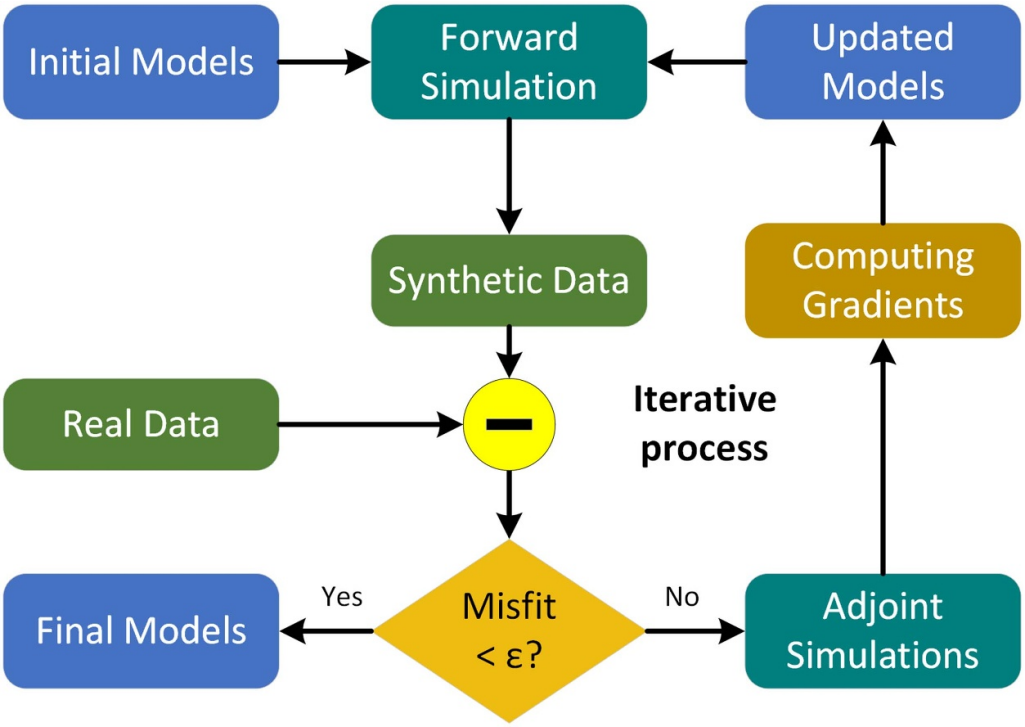

Figure 1. A schematic flow-chart of TDFWI process [15].

$$\chi(\mathbf{m}) = \frac{1}{2}\sum_{r=1}^{N}\int_0^T \left[p_{\text{syn}}(\mathbf{x}_r,t,\mathbf{m}) - p_{\text{obs}}(\mathbf{x}_r,t)\right]^2 \mathrm{d}t, \quad (4)$$

where p_{obs} and p_{syn} are the pressure signals of the observed and synthetic data at a particular receiver (r), and N is the total number of receivers at an excitation event with a constant source number. Time is defined by t and ranges between $(0,T)$. $\mathbf{m}$ is the model parameters (e.g. wave speed, attenuation, and density) to be inverted. This study focused on a frequency range where the attenuation is relatively low, so the attenuation effects have been neglected. In addition, the density was assumed to be constant throughout the process. Furthermore, this study is based on acoustic wave propagation with no shear contribution. Hence, only longitudinal wave speed ($v_{\text{p}} = \sqrt{\frac{\kappa}{\rho}}$) was considered as the key model parameter.

By introducing the Lagrange multiplier, $\lambda(\mathbf{x}_r,t)$ to enforce the PDE constraint in the TDFWI optimization, the misfit function is reformulated as follows:

$$\begin{aligned}\chi(\mathbf{m}) = {} & \frac{1}{2}\sum_{r=1}^{N}\int_0^T \left[p_{\text{syn}}(\mathbf{x}_r,t,\mathbf{m}) - p_{\text{obs}}(\mathbf{x}_r,t)\right]^2 \mathrm{d}t \\ & - \int_0^T\int_\Omega \lambda\left[\frac{1}{\kappa}\partial_t^2\phi - \nabla\cdot\left(\frac{1}{\rho}\nabla\phi\right) - f\right]\mathrm{d}\mathbf{x}\mathrm{d}t, \quad (5)\end{aligned}$$

where Ω is the model space and N is the total number of receivers. To avoid cluttering, $\Delta p_r(\mathbf{x}_r,t)$ is defined as

$$\Delta p_r(\mathbf{x}_r,t) = p_{\text{syn}}(\mathbf{x}_r,t,\mathbf{m}) - p_{\text{obs}}(\mathbf{x}_r,t). \quad (6)$$

The variation of equation (5) is

$$\begin{aligned}\delta\chi(\mathbf{m}) = {} & \int_0^T\int_\Omega \left[\sum_{r=1}^{N}\Delta p_r(\mathbf{x}_r,t)\,\delta(\mathbf{x}-\mathbf{x}_r)\right]\Delta p_r(\mathbf{x}_r,t)\,\mathrm{d}\mathbf{x}\mathrm{d}t \\ & - \int_0^T\int_\Omega \lambda\left[\delta\left(\frac{1}{\kappa}\right)\partial_t^2\phi + \frac{1}{\kappa}\delta\left(\partial_t^2\phi\right)\right. \\ & \left. -\nabla\cdot\left(\delta\frac{1}{\rho}\nabla\phi\right) - \nabla\cdot\left(\frac{1}{\rho}\nabla\delta\phi\right) - \delta f\right]\mathrm{d}\mathbf{x}\mathrm{d}t, \quad (7)\end{aligned}$$

where $\Delta p_r(\mathbf{x}_r,t)$ is the perturbation in pressure field p_{syn} due to to the model perturbation $\delta\mathbf{m}$ (e.g. $\delta\rho,\delta\kappa,\delta f$) [20, 53]. Luo [52] has shown that, when no model perturbations are present, $\lambda(\mathbf{x}_r,t)$ satisfies

$$\frac{1}{\kappa}\partial_t^2\lambda = \nabla\cdot\left(\frac{1}{\rho}\nabla\lambda\right) - \sum_{r=1}^{N}\left[\partial_t^2\Delta p_r(\mathbf{x}_r,t)\right]\delta(\mathbf{x}-\mathbf{x}_r) \quad (8)$$

with the end conditions $\lambda(\mathbf{x},T)=0$ and $\partial_t\lambda(\mathbf{x},T)=0$, and the free-surface boundary condition $\lambda=0$.

The adjoint potential wavefield can be defined using the time-reverse Lagrange multiplier by $\phi^\dagger(\mathbf{x},t)\equiv\lambda(\mathbf{x},T-t)s$ which must satisfy the following adjoint wave equation,

$$\begin{aligned}\frac{1}{\kappa}\partial_t^2\phi^\dagger - \nabla\cdot\left(\frac{1}{\rho}\nabla\phi^\dagger\right) = {} & \sum_{r=1}^{N}\left[\partial_t^2\Delta p_r(\mathbf{x}_r,T-t)\right] \\ & \times\delta(\mathbf{x}-\mathbf{x}_r) \quad (9)\end{aligned}$$

with initial conditions for adjoint potential $\phi(\mathbf{x}_r,T)=0$ and $\partial\phi(\mathbf{x}_r,T)=0$. The right-hand side of this equation can be referred to as the adjoint source.

By introducing adjoint wave equation, $\delta\chi$ can be expressed in terms of the sensitivity kernels, K_ρ and K_κ, given by

$$\delta\chi(\mathbf{m}) = \int_\Omega \left[\delta\left(\frac{1}{\kappa}\right)K_\kappa + \delta\left(\frac{1}{\rho}\right)K_\rho\right]\mathrm{d}\mathbf{x}. \tag{10}$$

K_ρ and K_κ determine how the model parameters should be updated [20]. These kernels represent the Fréchet derivatives with respect to the model parameters and are defined as

$$K_\kappa(\mathbf{x}) = -\int_0^T \phi^\dagger(\mathbf{x}, T-t)\,\partial_t^2\phi(\mathbf{x},t)\,\mathrm{d}t, \tag{11}$$

$$K_\rho(\mathbf{x}) = -\int_0^T \nabla\phi^\dagger(\mathbf{x}, T-t)\cdot\nabla\phi(\mathbf{x},t)\,\mathrm{d}t, \tag{12}$$

where the source variation perturbation is ignored, as ultrasound transducers are assumed to be well-defined in modeling. If the Fréchet derivatives can be calculated for individual misfit functions, the new model can be defined from the previous model by

$$\mathbf{m}_{\text{new}} = \mathbf{m}_{\text{old}} + \alpha\mathbf{H}_{\text{old}}^{-1}(-\mathbf{g}_{\text{old}}), \tag{13}$$

where α is the step length, $\mathbf{g}$ is the first-order derivatives of $\chi(\mathbf{m})$ (i.e. $g_k = \partial\chi/\partial m_k$), and $\mathbf{H}$ is the Hessian matrix of the second-order derivatives of $\chi(\mathbf{m})$ (i.e. $H_{kk'} = \partial^2\chi/(\partial m_k \partial m_{k'})$, k and k' are number of model parameters). However, computing $\mathbf{H}^{-1}$ for every misfit function is computationally expensive and memory-intensive. Hence, different optimization methods (e.g. L-BFGS) are used to approximate $\mathbf{H}^{-1}$ based on gradients from the previous iteration. The subscripts 'new' and 'old' represent the current and previous iterations.

2.2. Connection of TDFWI with deep learning

TDFWI can perform high-resolution damage imaging by iteratively solving the inverse problems of finding the best damage model to minimize the non-linear error/objective function. However, this iterative optimization process of TDFWI is time-consuming. Furthermore, approximation of $\mathbf{H}^{-1}$ can still be challenging as the stability of these computations highly depends on initial modeling. The neural network can be employed as a surrogate for approximating $\alpha\mathbf{H}_{\text{old}}^{-1}(-\mathbf{g}_{\text{old}})$ in equation (13).

In the proposed data-driven approach, the neural network is utilized as a post-processing approximation for TDFWI. Instead of iterating to the optimal misfit reduction, the process is halted early, yielding low-resolution inverted models referred to as early-stopped FWI inverted images (EFIs). The neural network is then trained using EFIs and their corresponding GTMs. Once trained, it can generate high-resolution reconstructions for new samples using only their EFIs.

3. Implementations, data, and model

3.1. Motive of the dataset building

As a pivotal preliminary step, this study concentrates on a water-immersed scanning setup with embedded, unknown material distributions with rectangular, possibly overlapped shapes with irregular patterns at the boundary. The design of the datasets aims to preserve the dominant challenges of nonlinear characteristics of the inverse problems that ultrasonic FWI faces in material properties reconstruction to identify different types of materials or foreign object debris. Because this research focuses on building the relationship between the aforementioned characteristic of data/model and the efficacy of the AI-empowered approach, random material disturbance from the background material (e.g. water) was selected to enable the full control of the dataset distribution for probing the inherent FWI-related physics.

3.2. Implementation of TDFWI in the model of interest

This study employs a simplified 2D acoustic model. The GTM were constructed by introducing regions with varying material properties in a water-immersed environment with a wave speed of 1479 m s^{-1}. They occupied a 12 mm $\times$ 12 mm space and were discretized into 2500 spectral elements, with 50 elements along each axis and 25 nodes within each element.

To generate the ultrasonic signals, a linear array of 100 source points is situated on one side of the domain (e.g. positioned right below the upper boundary as depicted in figure 2). These source points are excited simultaneously to produce a planar wavefield with a center frequency of 1 MHz. In order to capture this signal, an array of 400 receiver elements is distributed across the domain boundaries. Figure 2 visually represents the temporal evolution of the wave propagation at different time intervals. The initial model was purposefully constructed with different meshing configurations and background water wave speed compared to the GTMs to prevent the occurrence of 'inverse crime' in FWI [54]. The initial model was discretized into 900 spectral elements, featuring 30 elements along each axis, and assumed a homogeneous water-immersed environment with a wave speed of 1500 m s^{-1}. Figure 3 illustrates the reconstruction of the longitudinal wave speed mapping of an example model (figure 3(a)) with the Initial Model (figure 3(b)) using classical TDFWI. At the first iteration, it reconstructed a very blurry image (figure 3(c)) of the model, which gradually improved in further iterations (see the 10th iteration reconstructed image in figure 3(d). The TDFWI process was intentionally terminated before achieving optimal misfit reduction, thereby categorizing the resultant inverted images as EFI.

A spectral finite element solver called SPECFEM2D [19] was used to simulate forward and adjoint wave propagation. A Python-based framework called SeisFlows for TDFWI [15, 55] is used. The output of both forward and inverse simulations was the wave speed information of every node in the domain.

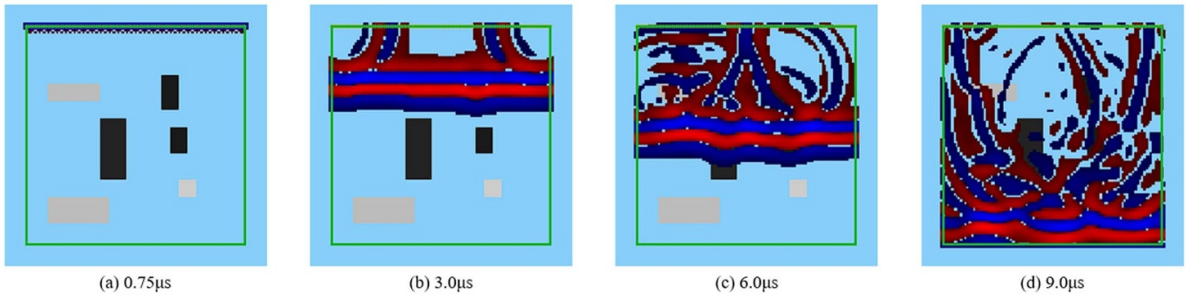

Figure 2. Illustration of the wave propagation at various times in the domain of 12 mm × 12 mm with 100 transducers near the top wall and 400 sensors (green) surrounding the domain. There are three unknown materials inside this domain.

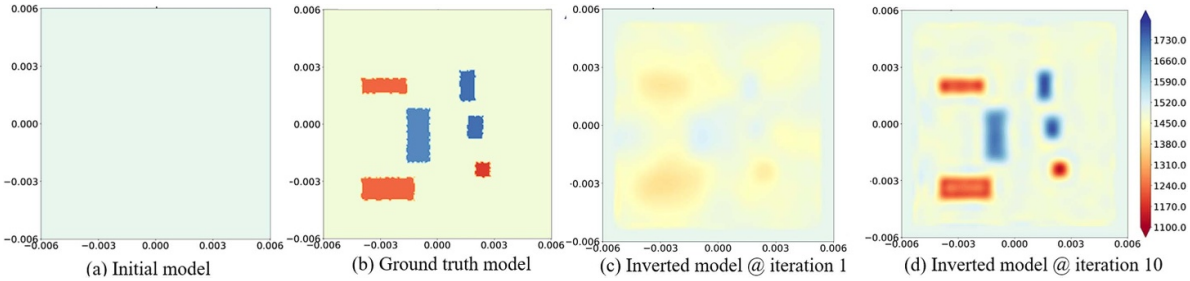

Figure 3. Longitudinal wave speed (V_p m/s) mapping of the (a) GTM, (b) initial model, and inverted images at (c) 3rd iteration and (d) 10th iteration.

3.3. Data generation with high performance computing (HPC)

Although SPECFEM2D and SeisFlows are well-developed and effectively manage HPC resources in each TDFWI simulation, generating large datasets with these tools remains highly time-consuming. For instance, a single TDFWI iteration for a simple model takes about a minute, meaning that sequentially simulating 1000 samples for 10 iterations would require approximately seven days. To overcome this bottleneck, an HPC-based framework was developed to enable rapid, parallel data generation. Figure 4 provides a schematic representation of this parallelized process.

The data generation process in this numerical study begins with developing the GTM and performing forward modeling simulations using SPECFEM2D to generate corresponding observed signals (referred to as real data in figure 1). A total of 25 000 GTMs were created by introducing randomly distributed unknown materials. To ensure dataset diversity, the number, dimensions, locations, and wave speeds of these unknown materials were randomly selected. Forward modeling simulations for all 25 000 GTMs were managed using the Simple Linux Utility for Resource Management (SLURM), following a similar approach as described for TDFWI simulations in the subsequent paragraph. (To avoid redundancy, this step is not shown in figure 4).

The next step involved performing iterative TDFWI with the observed data. A single initial model was used for every TDFWI simulation in this study, with parallelization managed by SLURM. Each SLURM job processed a batch of 10 models simultaneously using all 40 CPU cores. For this study, only 10 FWI iterations were performed, and to minimize disk space usage, only the updated models were stored after each iteration. Each SLURM job completed 10 TDFWI iterations in approximately 15 min. To generate 1000 models, 100 SLURM jobs were required, allowing the framework to produce 1000 models every 15 min across 10 datasets (one for each iteration).

SLURM was also used for post-processing, which is discussed in the following section. After processing, the dataset was saved in an HDF5 file before being fed to the neural network. The HDF5 file contained the EFI and corresponding GTM image pairs.

3.4. Post-processing

The solver's output yielded a one-dimensional wave speed array, measuring 22 500 × 1, representing every node of the spectral elements. This array inherently contained duplicate points due to the overlap of nodes between adjacent spectral elements. Furthermore, to mitigate artifacts, a mask was applied to the domain's periphery during simulation. Consequently, the removal of these data points resulted in a final one-dimensional array sized 9025 × 1. To facilitate the transformation into 2D images, the 1D array was converted into a 2D array measuring 95 × 95 using the SciPy interpolate library. Subsequently, all samples were aggregated into a unified three-dimensional NumPy array, sized 25 000 × 95 × 95, corresponding to the 25 000 samples, with each sample composed of pixel values spanning 95 × 95. To streamline max pooling and upsampling machine learning processes, zero-padding was applied to the samples along their bottom and right sides, expanding the dimensions to 96 × 96. Within

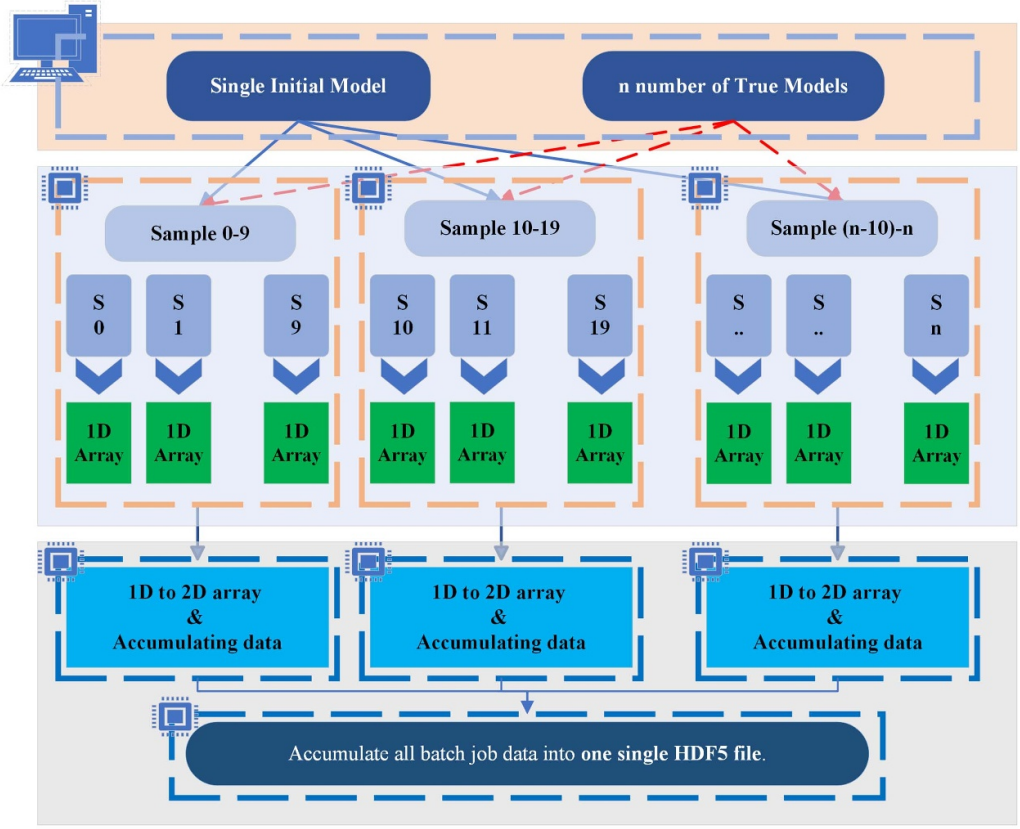

Figure 4. A schematic of HPC-based data generation technique using SLURM.

the context of the deep learning model, each sample was treated as a grayscale image, hence featuring a single channel (as opposed to the three channels used for RGB images). Consequently, the final configuration of the authors' dataset stood at 25 000 $\times$ 96 $\times$ 96 $\times$ 1.

In adherence to standard practice, the wave speeds underwent normalization, employing a linear regression technique that scaled them between 0 and 1. Consequently, this normalization resulted in the lowest wave speed (1180 m s^{-1}) in the GTM dataset being set to 0, and the highest wave speed (1730 m s^{-1}) in the GTM dataset was rescaled to 1. Similarly, each EFI dataset for different iterations was normalized between 0 and 1 with respect to the individual minimum and maximum wave speed values. Throughout most of the experiments in this study, this scaled dataset was utilized. However, when assessing the neural network model's generalizability, as described in section 4.4, the performance was suboptimal when using the scaled dataset. Therefore, for this specific portion of the study, the data normalization process was deliberately omitted during data processing.

The final step in the data preparation process is to randomize and split the dataset into respective training, validation, and testing sets. The training and validation datasets (of 20 000 and 3000 samples, respectively) are used to train the neural network and the testing dataset (of 2000 samples) is reserved to test the performance of the trained model. A smaller dataset of 1000 samples (700 for training and 300 for validation) was subset from the original training dataset (of 23 000 samples) to train the model with a smaller dataset. In addition, to study the TDFWI iterations over DL-solution in adjoint tomography, the model was separately trained on 3rd iteration EFI-GMT pair and 10th iteration EFI-GMT pair datasets consisting of 1000 samples each. Two different augmentation strategies were implemented for the smaller datasets.

3.5. *Augmentation strategies and datasets*

This study employed two augmentation strategies to increase the quantity and complexity of the training dataset without generating samples. The first one was the classical augmentation strategy, where the sample was flipped along its axes. This allowed us to increase the original dataset with three times additional training samples. The second one was the mixing augmentation strategy. Here, two samples were multiplied by coefficients α and $(1-\alpha)$, respectively, and then added together to make a new sample. The value of α randomly varied from 0 to 1.

The neural network was trained with multiple datasets. The size of the datasets is described in table 1. Here, D3-1k denotes the dataset with 3rd iteration EFI-GMT pairs of 1000 samples. The augmentation strategies were only applied to the smaller

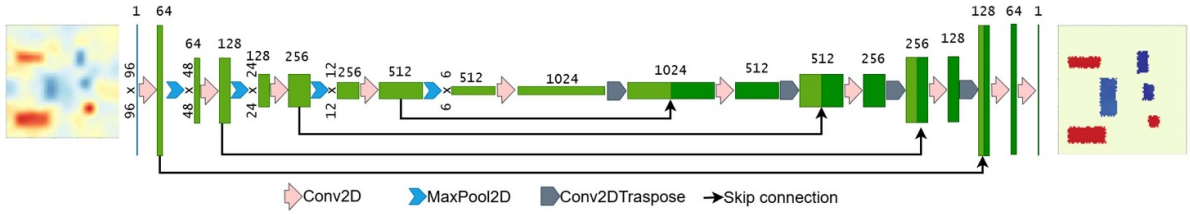

Figure 5. The architecture of the U-Net model used in this study [43].

Table 1. Size of different datasets. Size of the each dataset would be 'X'×96×96×1, where 'X' is the sample number in the datasets described in the table. D3-1k denotes the dataset with 3rd iteration EFI-GMT pairs of 1000 samples.

	Training				
Datasets	No Aug	Flipping Aug	Mixing Aug	Validation	Testing
D3-1k	700	2800	2800	300	2000
D10-1k					
D3-23k	20 000	–	–	3000	

training datasets. No augmentation was applied to validation, testing, and larger training datasets (of 20 000 samples).

3.6. Deep neural network architecture

The task of accelerating the TDFWI iteration process can be posed as translating low-iteration TDFWI images into their corresponding ground truth images. The U-Net architecture is employed for this purpose, shown in figure 5. Initially introduced for biomedical image segmentation by Shan *et al* [56], U-Net, with its encoding-decoding architecture, has exceptional capabilities to capture local and global features, producing high-resolution images with limited data. The general U-Net architecture is followed, composed of four down-sampling layers (red arrows) and four up-sampling layers (dark green arrows). In the down-sampling process, four sets of two convolutional kernels extract feature maps. Then, followed by a pooling layer (light green arrows), the feature map projections are distilled to the most essential elements by using a signal-maximizing process. By the end of the down-sampling process, the feature maps are 1/16 of the original size: 6x6 in figure 5. Successful training should result in 6 channels in this feature map retaining important features, which is verified by the model predictions in figure 7. Bi-linear interpolation is used in the up-sampling process to expand the feature maps. At each layer, high-resolution features from the down-sampling path are concatenated to the up-sampled output from the layer below to form a larger number of feature channels (white arrows). This structure allows the network to propagate context information to higher-resolution layers so that the following convolution layer can learn to assemble a more precise output based on this information.

3.7. Combined loss functions

The current predominant approach involved deploying the mean-square error (MSE) as the principal loss function, in conjunction with the Adam optimizer, as expounded in the work by Kleman *et al* [43]. Nonetheless, there are numerous loss functions available, each adept at extracting distinct features during the reconstruction process [57]. Consequently, combining a variety of loss functions can significantly enhance the fidelity of local structure reconstructions. For instance, Liu *et al* [51] on the DL-based x-ray computed tomography model, TomoGAN incorporated multiple loss functions, including adversarial loss, MSE-loss, and perceptual loss, and demonstrated improved prediction results. Based on the inspiration from this research, this study also adopted a multi-loss functions training approach.

In this study, the comprehensive loss function training comprises a weighted average loss that encompasses the following components:

Pixel-wise MSE (mean square error): To quantify the pixel-wise discrepancy between the EFI image (I^C) and the GTM image (I^E), the MSE was computed as follows:

$$\mathcal{L}_{\text{mse}} = \frac{1}{m}\sum_{i=1}^{m}\sum_{c=1}^{W}\sum_{r=1}^{H}\left(\left(I_i^C\right)_{c,r} - \left(I_i^E\right)_{c,r}\right)^2, \qquad (14)$$

where W and H denote the width and height of the GTM or EFI images, respectively, and m is the number of total training mini-batches.

Gradient difference loss (GDL): GDL applied to a GTM/EFI pair is employed to preserve edges and enhance the sharpness of the reconstructed unknown materials. It calculates the sum of the L2-norm differences between the gradients of an EFI image and its corresponding GTM as follows [58, 59]:

$$\mathcal{L}_{\text{gdl}} = \frac{1}{m}\sum_{i=1}^{m}\left(||\nabla_x I_i^C - \nabla_x I_i^E||_2 + ||\nabla_y I_i^C - \nabla_y I_i^E||_2\right). \qquad (15)$$

Perceptual loss: A VGG (Visual Geometry Group) network, pre-trained with ImageNet data, has proven to be an effective high-level feature extractor for CT images. In this study, a 16-layer pre-trained VGG network [51, 60] was utilized to compute the perceptual loss as the Euclidean distance between the

feature representations of an EFI and its corresponding GTM image:

$$\mathcal{L}_{\text{perc}} = \frac{1}{m}\sum_{i=1}^{m}\sum_{h=1}^{L}\left(E\left(I^{C}\right)_{i,h} - E\left(I^{E}\right)_{i,h}\right)^{2}. \tag{16}$$

In this equation, $E(\cdot)$ represents the pre-trained CNN encoder responsible for embedding an image into an L-dimensional latent vector [51].

To create a comprehensive loss function for model training, the individual loss functions with the corresponding hyperparameters (weights) were combined as follows:

$$\mathcal{L} = \lambda_{\text{mse}}\mathcal{L}_{\text{mse}} + \lambda_{\text{gdl}}\mathcal{L}_{\text{gdl}} + \lambda_{\text{perc}}\mathcal{L}_{\text{perc}}, \tag{17}$$

where $\lambda_{\text{mse}}, \lambda_{\text{gdl}}, \lambda_{\text{perc}}$ represent the hyperparameters (weights) assigned to each loss function.

To determine the optimal weights for these distinct loss functions, a hyperparameter tuning algorithm was developed using Ray Tune. This algorithm utilized the AsyncHyperBandScheduler in conjunction with the Optuna search algorithm [61]. This particular search algorithm has demonstrated superior performance and efficiency compared to conventional methods such as grid search or random search for this study. The maximum value for the validation structural similarity index measure (SSIM) was set as the evaluation metric for the tuning. Such an approach enables effective hyperparameter tuning. This optimization results in enhanced accuracy and fidelity in the reconstructed output.

4. Results and discussion

For comparing two images (GTM and inverted/predicted images), the SSIM index and pick-signal noise ratio (PSNR) were used. In the simplified version of the SSIM index and PSNR, two images (G and I) can be compared as

$$\text{SSIM}(G,I) = \frac{(2\mu_G\mu_I + C_1)(2\sigma_G\sigma_I + C_2)}{(\mu_G^2 + \mu_I^2 + C_1)(\sigma_G^2 + \sigma_I^2 + C_1)}, \tag{18}$$

$$\text{PSNR}(G,I) = 10\log_{10}\left(\frac{\text{MAX}_G^2}{\text{MSE}_{G,I}}\right), \tag{19}$$

$$\text{MSE}_{G,I} = \frac{\sum_{c=1}^{W},\sum_{r=1}^{H}[G_{c,r} - I_{c,r}]^2}{W \times H}, \tag{20}$$

where μ_G and μ_I are the local means of luminance, σ_G, and σ_I are the standard deviations of the contrast of the images, and C_1 and C_2 are positive constants to avoid a null denominator. C_1 and C_2 are defined as $(K_1 * R)^2$ and $(K_2 * R)^2$, where K_1 and K_2 are small constants with a value of 0.01 and 0.03, respectively, and R is the dynamic range of the pixel value of the images. MAX_G denotes the maximum possible pixel value of the image, W, H are width and height fo the images.

Before delving into the outcomes of the deep-learning approach, a comparative analysis of the inverted images obtained through classical FWI simulations at higher iterations is conducted. A randomly selected sample from the test dataset was used for this analysis, as shown in figure 6. For this specific sample, the classical FWI process could converge only up to 36 iterations. The reconstructed images from the 1st and 36th iterations are displayed in figures 6(b) and (c), respectively. The misfit reduction curve, illustrated in figure 6(d), shows that the reduction rate plateaued after the 15th iteration, with the process ultimately halting at the 36th iteration. This early termination occurred due to discrepancies between the initial model and the GTM, particularly in terms of mesh configuration and background water wave speed. Additionally, figure 6(e) illustrates the variations in the SSIM index across different FWI iterations concerning the GTM image. Notably, the SSIM index started at approximately 50% similarity with the GTM in the 1st iteration and gradually increased to around 67% by the 15th iteration. Similarly, the PSNR value, which began at 10 dB in the initial iteration, improved to 17.5 dB after 15 iterations. However, in line with the misfit reduction trend, both the SSIM index and PSNR values stagnated beyond the 15th iteration, showing no further improvements. Although the FWI reconstruction effectively identified the unknown materials and their wave speeds, it struggled to accurately capture the irregular boundary patterns of these materials at the selected frequency. It is important to note that this particular sample represents a simplified case, disregarding many uncertainties that could arise in real-world scenarios. However, the focus of this study is not to evaluate the efficacy of FWI reconstruction itself but rather to demonstrate how a U-Net model can enhance FWI reconstructions using these suboptimally reconstructed EFI images at low iterations.

In the following sections, the results of the DL-based FWI solutions are presented. The U-Net model was trained with several smaller datasets (D3-1k and D10-1k in table 1) with three variants according to the augmentation strategy: no augmentations, flipping augmentations, and mixing augmentations and a large dataset (D3-23k in table 1). The training time for all smaller datasets was about 8–10 min and 2 h 45 min for the larger dataset using a single GPU core.

4.1. Test prediction: small datasets - 3rd and 10th iteration inverted samples

Predictions of all U-Net models trained with variants of D3-1k datasets achieved SSIM values above 91%. Comparing the different augmentation strategies, the performance of the U-Net models was comparable in the mixing augmentations and no-augmentations in most of the test samples. On the other hand, the model trained with the flipping augmented dataset of D3-1k performed slightly better than other strategies in terms of average SSIM and PSNR values (table 2). The average SSIM and PSNR values of these test results were 90.95% and 25.63 dB (no-augmentations), 93.15% and 26.94 dB (flipping augmentation), and 92.87% and 26.51 dB (mixing augmentation).

Figure 7 illustrates the model performance for two specific test samples: Sample-1, where the material distribution was sparse with no overlapping regions, and Sample-2, where the material distribution was more concentrated with overlapping regions. Figures 7(a) and (f) represent the ground truth

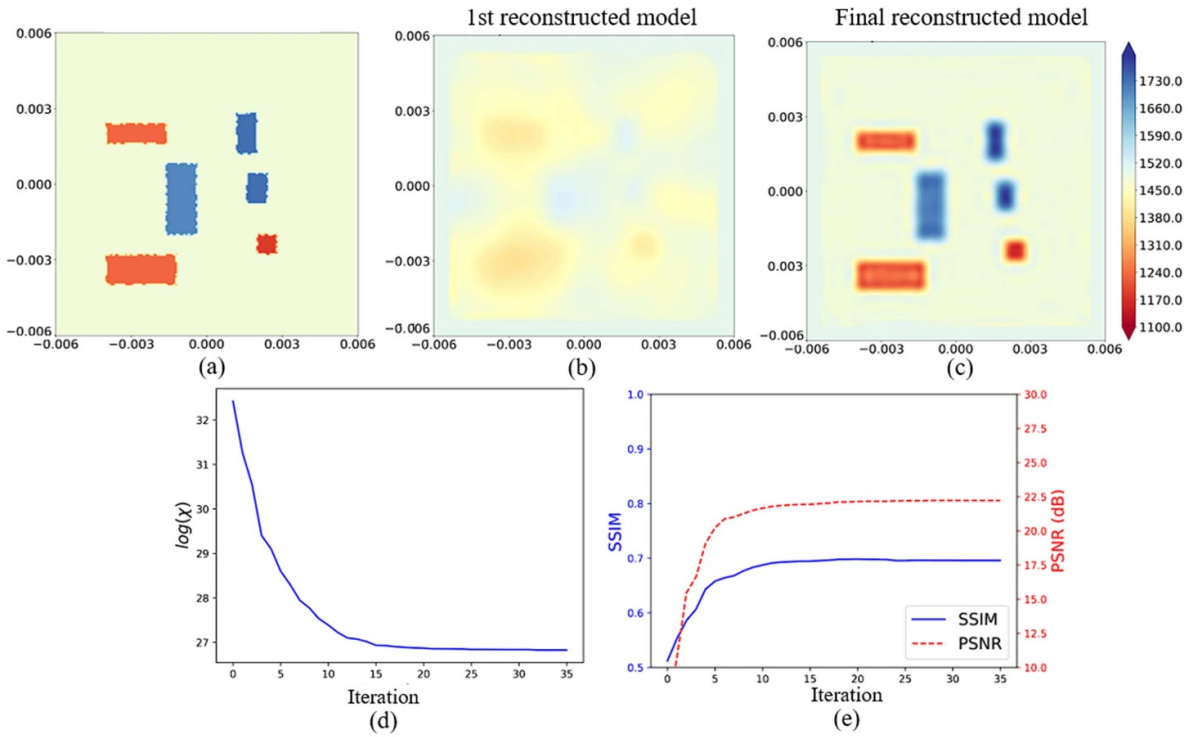

Figure 6. Comparison between inverted images at different iterations of the classical FWI with the corresponding GTM for an example sample. (a) is the GTM, (b) inverted image at the 3rd iteration, (c) final inverted images at the 36th iteration, (d) Misfit reduction over iterations, and (e) SSIM index and PSNR between GTM and inverted images at different iterations.

images of the samples, while (b) and (g) display the corresponding EFIs at the first iteration, which served as inputs to the neural network. Figures 7(c)–(e) are the predictions of sample-1 when the model was trained with the D3-1k datasets with no augmentations, flipping augmentation and mixing augmentation, respectively. Similarly, figures 7(h)–(j) are the predictions of sample-2 that depict the predictions for sample 2 under the same training conditions. For both samples, augmentations somewhat helped in the boundary definition of the material over the no-augmentation case. Overall, while the SSIM and PSNR values were comparable across all cases, the flipping augmentation strategy yielded slightly better results. However, some artifacts and anomalies persisted in the predicted wave speed reconstructions. Although the DL-based FWI solution demonstrated quantitative improvements over the classical FWI approach, the model still struggled to capture the irregular boundary patterns of the unknown materials. This indicates that the quality and quantity of the training dataset may not have been sufficient to achieve significantly enhanced predictions.

By utilizing higher iteration EFIs from the 10th iteration datasets (D10-1k), the model was able to achieve approximately 99% similarity to the GTM and 40 dB PSNR value across all augmentation strategies, as detailed in table 2. The model's performance for the same specific test samples used in the previous case is depicted in figure 8. In this case, figures 8(b) and (g) display the EFIs at the tenth iteration, which served as inputs to the neural network. Figures 8(c)–(e) and (h)–(j) present the predictions obtained using different augmentation strategies. Unlike the classical FWI solution or DL-based FWI solution shown in the previous case, this model could predict the irregular patterns of the unknown material boundaries. Notably, the neural network exhibited similar performance across all cases. The artifacts and anomalies in wave speed predictions were also reduced substantially. However, augmentations did not appear to provide additional benefits when the model was trained with higher iteration EFIs.

These results reinforce the idea that higher iteration EFIs already contained richer and cleaner features of wave speed distribution for accurate predictions, making additional augmentation redundant. Since the key significance of the augmentation strategies is to make the training dataset diverse and more generalizable, they benefit the model trained with lower iteration EFIs. Additionally, augmentations, especially mixing augmentation, may introduce certain spatial transformations that may not align with the physical properties of wave propagation, which lead to the inclusion of artifacts (figure 7(e)). On the other hand, augmentations have some benefits with the complex samples (similar to sample-2) with complicated overlaps than the sample with sparse material distribution (figures 8(h)–(j)).

4.2. Test prediction: large dataset - 3rd iteration inverted samples

Expanding the dataset size has been proven to be instrumental in enhancing the model's training. Consequently, the same U-Net architecture was trained with a dataset comprising 20 000

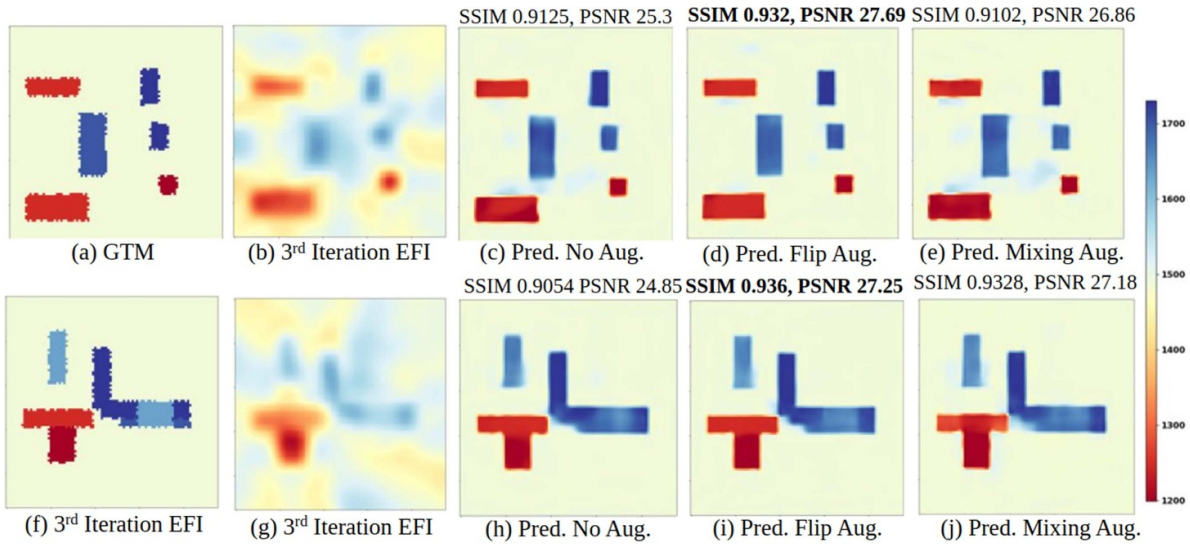

Figure 7. Prediction results of the U-Net model trained with D3-1k datasets. (a) and (f) are the GTMs and (b) and (g) are the respective EFI at 3rd iteration. (c)–(e) are the predictions of sample-1 when the model was trained with the D3-1k datasets with no augmentations, flipping augmentation, and mixing augmentation, respectively. (h)–(j) are the predictions of sample-2 when the model was trained with the D3-1k datasets with no augmentations, flipping augmentation, and mixing augmentation, respectively.

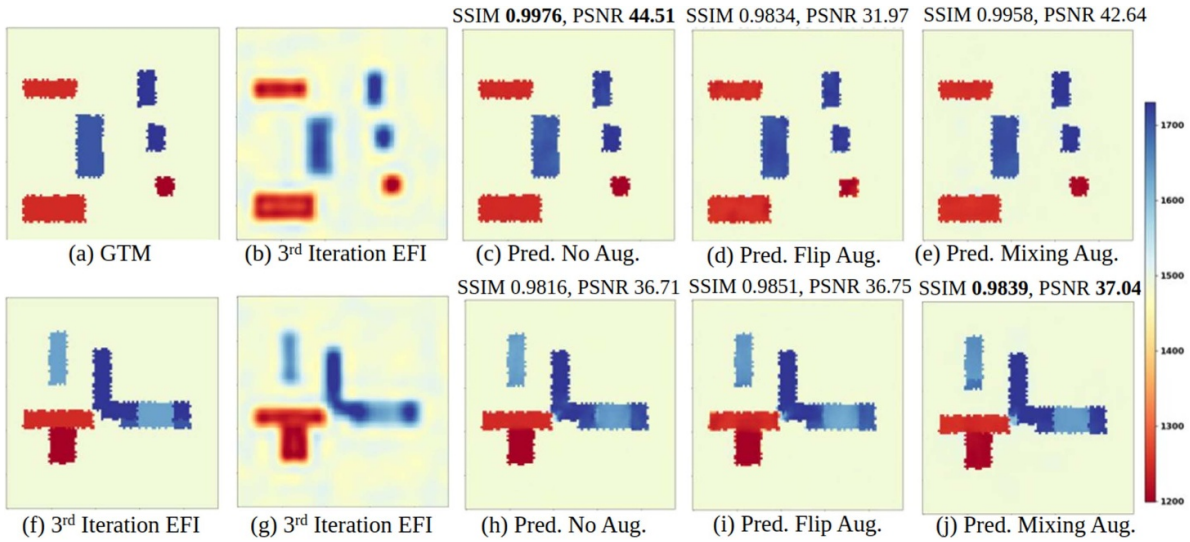

Figure 8. Prediction results of the U-Net model trained with D10-1k datasets. (a) and (f) are the GTMs and (b) and (g) are the respective EFI at the 10th iteration. (c)–(e) are the predictions of sample-1 when the model was trained with the D10-1k datasets with no augmentations, flipping augmentation, and mixing augmentation, respectively. (h)–(j) are the predictions of sample-2 when the model was trained with the D10-1k datasets with no augmentations, flipping augmentation, and mixing augmentation, respectively.

Table 2. Average SSIM (in %) and PSNR (in dB) values for testing the model trained different datasets.

Datasets	No Augmentation	Flipping Augmentations	Mixing Augmentation
D3-1k	90.95%, 25.63	93.15%, 26.94	92.87%, 26.51
D10-1k	98.24%, 39.10	97.62%, 32.55	98.13%, 38.42
D3-23k	98.0%, 36.46	×	×

samples and validated with 3000 samples (referred to as D3-23k in table 1). In this scenario, only the third iteration of TDFWI images with no augmentation was utilized.

Figure 9 showcases the neural network's performance for the same test samples featured in previous cases. While the neural network with D3-1k datasets failed to predict the irregularities boundary of the inside materials, this model was able to identify them with SSIM and PSNR boost to 98.9% and 40.5 dB for sample-1 and 98.13% and 35.78 dB for sample-2, respectively.

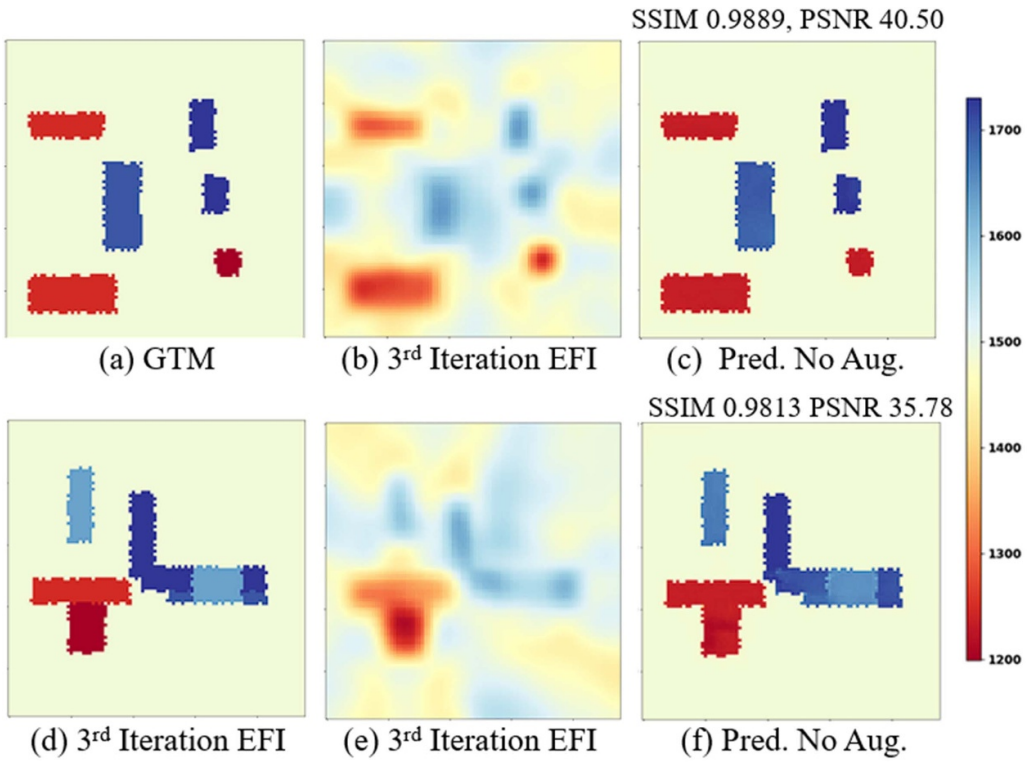

Figure 9. Prediction results of the U-Net neural network trained with D3-23k datasets. (a)–(c) are the GTM, 3rd iteration EFI, and predicted images for sample-1, and (d)–(f) are the GTM, 3rd iteration EFI, and predicted images for sample-2, respectively. No augmentation strategy has been used in the training process.

The average SSIM and PSRN values increased from 92% and 25.63 dB of D3-1k dataset with no augmentation to about 98% and 36.46 dB, as illustrated in table 2. The prediction results exhibited similarities to those obtained with the D10-1k dataset. This underscores that the current neural network can produce nearly perfect high-resolution images when trained with very low-resolution samples, provided it is trained with sufficient data.

4.3. Results of using combined loss functions

Up to this point, discussion has been made regarding enhancing prediction results through dataset quality improvements and the augmentation of data samples. In this section, the improvement of prediction results through the incorporation of multiple loss functions during training is discussed.

Among the various scenarios examined previously, it was observed that the model trained with the D3-1k dataset yielded suboptimal results. Thus, this dataset was selected for further refinement through hyperparameter tuning to boost the model's testing performance. It is important to note that the tuning process significantly relies on the initial values and the range of hyperparameters.

At first, the tuning was initiated with values of 1 for λ_{mse}, λ_{gdl}, and λ_{perc}, and explored a range of (1–10), (1–10), and (0–2) for each, respectively. The weights were iteratively updated using the best values obtained in each trial. After several iterations, the initial values for the loss function weights were set at $\lambda_{\text{mse}} = 19.48$, $\lambda_{\text{gdl}} = 23.86$, and $\lambda_{\text{perc}} = 2.0$, with an extended range of (1, 40), (1, 40), and (0, 2) for λ_{mse}, λ_{gdl}, and λ_{perc}, respectively. With this setup, the optimal weights were found as follows: $\lambda_{\text{mse}} = 22.90$, $\lambda_{\text{gdl}} = 30.37$, and $\lambda_{\text{perc}} = 1.32$. The tuning process, comprising 40 trials on 4 GPUs, was completed in approximately 20 to 25 min.

In the next experiment, the perceptual loss was dropped, and the model was tuned with only MSE and GDL with initial values of $\lambda_{\text{mse}} = 22.90$ and $\lambda_{\text{gdl}} = 30.37$ and range of (1, 40) for each. The optimal weights in this case were $\lambda_{\text{mse}} = 21.002$, $\lambda_{\text{gdl}} = 28.3987$.

Figure 10 presents the prediction results for the same samples discussed previously. Comparing these prediction results reveals a significant qualitative improvement in predictions when employing the multi-loss function training approach. The neural network trained with all three loss functions showed the best results in terms of SSIM and PSNR value as well as artifact reduction. Surprisingly, this model was able to identify the irregular patterns on the boundary of the unknown materials to some extent (figures 10(c) and (h)). On the other hand, the neural network trained with MSE and GDL loss function only (figures 10(d) and (i)) slightly improved prediction in terms of boundary definition comparing with the prediction results of the neural network with MSE loss function only (figures 10(e) and (j)).

Figure 11 illustrates a quantitative comparison between these three cases. Combining GDL with the MSE loss function slightly improved the overall performance compared to the case with only the MSE loss function. On the other hand,

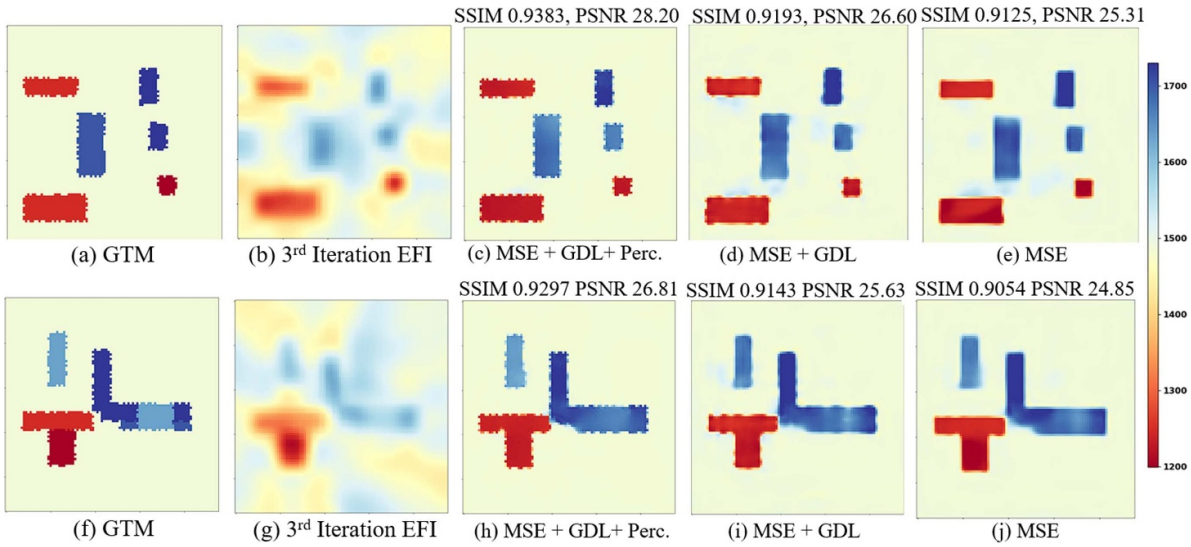

Figure 10. Prediction results for the model with multiple loss functions. The model was trained with the D3-1k dataset with no augmentation. (a) the GTM, (b) 3rd iteration EFI, (c), (d), and (e) are the predicted images of the model trained with MSE + GDL + perceptual loss functions, MSE + GDL loss function, and MSE loss function only for sample-1, respectively. Similarly, (f) the GTM, (g) 3rd iteration EFI, (h)–(j) are the predicted images of the model trained with MSE + GDL + perceptual loss functions, MSE + GDL loss function, and MSE loss function only for sample-2., respectively. ((e) and (j) are the same figures in figures 7(c) and (h), respectively.)

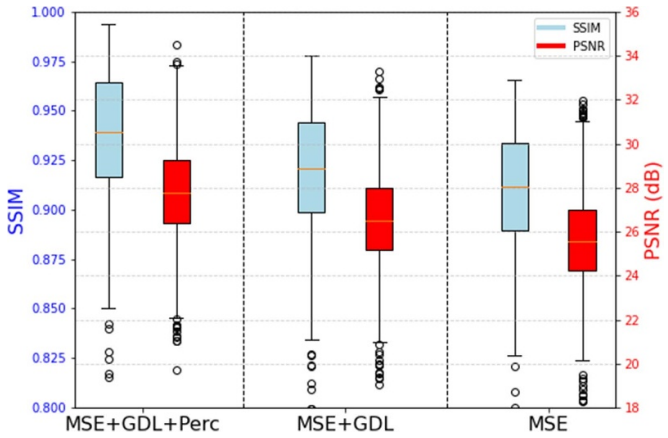

Figure 11. Quantitative comparison among predictions of neural networks trained with various loss function combinations.

the neural network with all three loss functions outperformed the other two cases with an average SSIM index of 93.83% and PSNR of 27.83 dB. This implies that the addition of GDL and perceptual loss significantly improves reconstruction quality. However, outliers in all cases suggest variability in results depending on the test case.

4.4. OOD generalization and possible solution

Generalization, within the context of neural networks, denotes the model's capability to effectively accommodate novel and previously unobserved data. Establishing the model's generalizability presents notable complexities, particularly in the context of supervised learning, as exemplified in this study [62]. Conventionally, models are trained with the presumption that both the training and test datasets adhere to the same underlying distribution. Consequently, the challenge escalates when the model encounters data points that fall outside the distribution learned during training, a scenario commonly referred to as OOD data [63, 64].

Within the original dataset, the wave speed spanned from 1180 m s^{-1} to 1730 m s^{-1}, with a background wave speed of 1479 m s^{-1}. To assess the model's generalizability to OOD data, six additional test datasets were created, each comprising 100 samples of the 3rd iteration of EFI, as detailed in table 3. In this table, the first three datasets (H1-3) exhibited higher wave speed distributions than the original dataset, while the last three datasets (L1-3) featured overlapping and lower wave speed distributions compared to the original. Notably, datasets H1 and L1 closely approximated the original dataset, while H3 and L3 represented the furthest deviations from it. To further challenge the model's adaptability, the mesh configuration was also modified. Instead of the 2500 spectral elements used in the training dataset, these six datasets were discretized into only 1600 spectral elements. Additionally, the boundaries of the unknown materials were smoothed, differing from the original training datasets. These datasets were specifically designed to evaluate the model trained with D3-23k.

Initially, all datasets underwent standard data preprocessing, scaling them to a range between 0 and 1. However, under the current model configuration, it became evident that the model struggled to predict the wave speed distribution of each test dataset. Figure 12 showcases the outcomes of these experiments. Figures (a), (d), (g), (j), (m) and (p) display example samples' GTMs from datasets H1, H2, H3, L1, L2, and L3, respectively. Moreover, images in figure 12(b), (e), and (h) demonstrate that the model's predictive capability for datasets H1-3 was restricted by the upper bounds of the wave speed range found in the training dataset, which reached

Table 3. Specifications of different datasets for generalizability testing. The maximum and minimum wave speeds include the background wave speed of a particular dataset.

	wave speed ($m\,s^{-1}$)		
Dataset	Background	Max	Min
H1	1800	2270	1800
H2	2000	2370	1850
H3	2200	2476	2020
L1	1300	1420	890
L2	1100	1220	820
L3	900	1220	700

1700 $m\,s^{-1}$ (1 after scaling). Conversely, for datasets L1-3, the model managed to extract certain features of the wave speed distribution that closely aligned with or fell within the training dataset's range. Therefore, the model's inability to forecast the wave speed distribution primarily arose from being trained on a dataset with wave speed distributions confined within the 0 to 1 range (corresponding to 1180 $m\,s^{-1}$ to 1730 $m\,s^{-1}$ before scaling).

To address this challenge, the model was trained with unscaled data, recognizing that such data might introduce a slower and potentially more unstable learning process. As depicted in figures 12(c), (f), (i), (l), (o) and (r), these images showcase the prediction outcomes achieved by the neural network model trained on unscaled data.

The model improved its ability to identify various features, including the size, shape, and spatial distribution of unknown materials within the domain. Additionally, it showed some accuracy in predicting wave speeds, particularly for datasets H1 and L1, which closely resembled the training dataset. However, the neural network failed to reconstruct the smooth boundaries of the unknown materials in every case. This outcome is expected, as all training samples had rough boundary definitions. Conversely, certain artifacts and anomalies persisted, primarily due to differences in the mesh configuration used in the FWI simulations. These findings suggest that further refinement and optimization are needed to improve the model's generalization performance.

5. Conclusion

In order to address the challenges of computational costs and extended processing times of TDFWI, this paper performed fundamental research on the improvements achievable with the integration of deep learning with the adjoint tomography theory. In this study, numerical models were developed for FWI simulation, and the SSIM was used to evaluate the inverted images against GT<s. An HPC-based framework has been built with a detailed description to accelerate the data generation process. Without the integration of deep learning, classical FWI achieved an SSIM of approximately 67% and PSNR of 22.5 dB after 36 iterations, revealing the need for improvement in structural reconstruction accuracy.

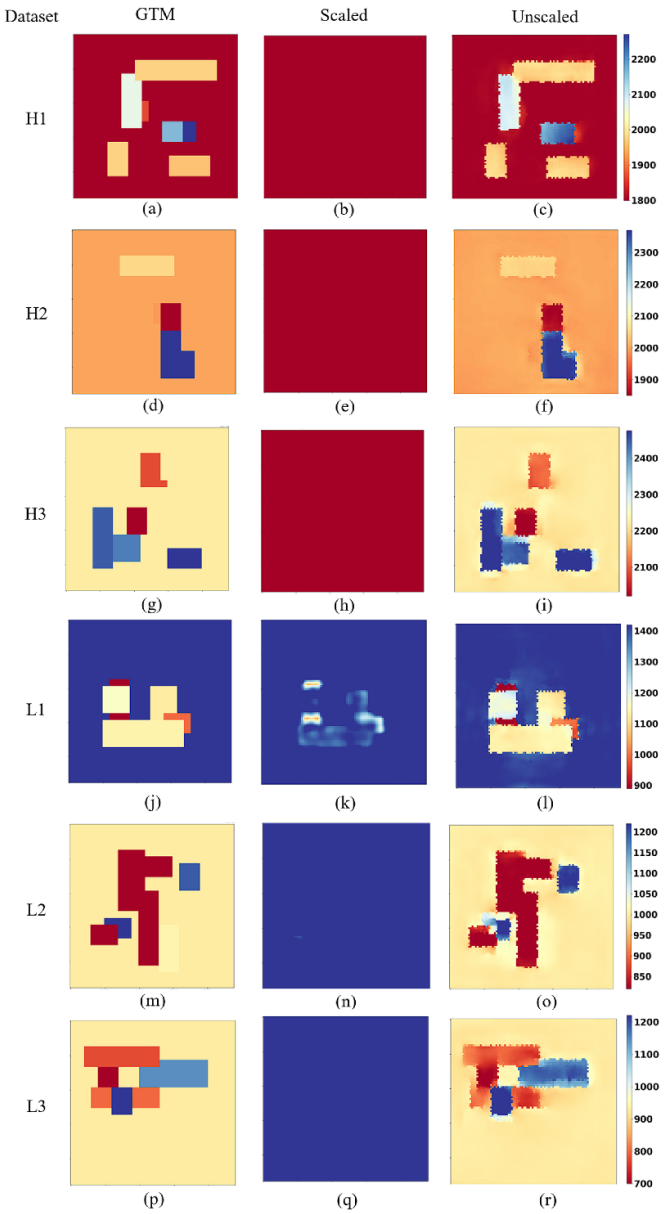

Figure 12. Prediction results of the neural network for OOD data. (a), (d), (g), (j), (m) and (p) are GTM of example samples from datasets H1, H2, H3, L1, L2, and L3, respectively. (b), (e), (h), (k), (n), and (q) are the prediction results of the respective samples, when tested with the model trained with the scaled dataset. (c), (f), (i), (l), (o), and (r) are the prediction results of the respective samples, when tested with the model trained with the scaled dataset.

A 2D CNN based on the U-Net architecture was selected to explore the scientific questions on how amounts/distributions of the training data, augmentation, and loss functions influence the efficacy of this approach.

- When using a small dataset comprising 3rd-iteration TDFWI images (700 training and 300 validation samples), the current model achieved a 91% similarity index and 26 dB. By augmenting this dataset, the SSIM and PSNR

increased to 92%–93% and 27 dB. However, artifacts persisted, and no notable improvements in boundary delineation between different wave speed distributions, highlighting the need for more extensive training data. Additionally, with such a limited and low-quality dataset, prediction accuracy may be further enhanced by incorporating multiple loss functions during training.

- When using a small dataset comprising 10th-iteration TDFWI images (700 training and 300 validation samples) with more embedded material information, the model achieved an impressive SSIM score of 98% and PSNR of 39 dB with notable improvements in boundary delineation between different wave speed distributions. Notably, in this scenario, augmentations did not confer additional benefits, reinforcing the notion that higher iteration data inherently contained ample information for precise predictions.
- When using a large dataset containing 1st-iteration TDFWI images (20 000 training and 3000 validation samples), a surprising observation is that lower iteration data can achieve near-perfect predictions when provided in large quantities.

The above conclusions are with in-distribution data. The generalizability studies indicated that the model could successfully reconstruct wave speed maps even for samples out of the wave speed ranges in the original dataset. However, anomalies and artifacts were present, suggesting that further research is needed to enhance the model's robustness in handling unseen data, especially for those with different shapes and sizes.

While this study provides insights into the efficacy of deep-learning-based FWI reconstruction, the current simplified datasets do not fully reflect real-world applications across various fields. Therefore, future work will focus on incorporating more complex and heterogeneous data and the scanning configuration aligning with the medical imaging practice, modeling frequency-dependent attenuation effects, and implementing multi-scale frequency inversion.

Data availability statement

The data cannot be made publicly available upon publication because they are not available in a format that is sufficiently accessible or reusable by other researchers. The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgments

This work is supported by the financial contributions of NSF projects #2152764 and #2152765. The authors extend gratitude for the computational resources provided by ACCESS Project #MDE220005. Special thanks to Dr Paul Rodriguez for his expert technical support in data generation and training through the XSEDE Extended Collaborative Support Service (ECSS) Grant.

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

ORCID iDs

Shoaib Anwar https://orcid.org/0009-0002-6732-4424
Mark A Anastasio https://orcid.org/0000-0002-3192-4172
Weihua Su https://orcid.org/0000-0002-4458-0524
Jiaze He https://orcid.org/0000-0001-6198-0688

References

[1] Su Z, Ye L and Lu Y 2006 Guided lamb waves for identification of damage in composite structures: a review *J. Sound Vib.* **295** 753–80
[2] Yao G, Wu D and Wang S X 2020 A review on reflection-waveform inversion *Pet. Sci.* **17** 334–51
[3] Drinkwater B W and Wilcox P D 2006 Ultrasonic arrays for non-destructive evaluation: a review *NDT & E Int.* **39** 525–41
[4] Choi H and Popovics J S 2015 Nde application of ultrasonic tomography to a full-scale concrete structure *IEEE T. Ultrason. Ferr. Freq. Control* **62** 1076–85
[5] Aggelis D G 2013 Wave propagation through engineering materials; assessment and monitoring of structures through non-destructive techniques *Mater. Struct.* **46** 519–32
[6] An Y-K, Park B and Sohn H 2013 Complete noncontact laser ultrasonic imaging for automated crack visualization in a plate *Smart Mater. Struct.* **22** 025022
[7] Huang L, Zeng L, Zhang X, Wan X and Chen F 2024 Hidden corrosion monitoring via phase shift evaluation of lamb wave modes *Smart Mater. Struct.* **33** 085022
[8] Chen C H 2007 *Ultrasonic and Advanced Methods for Nondestructive Testing and Material Characterization* (World Scientific)
[9] Rao J, Qiu H, Teng G, Al Mukaddim R, Xue J and He J 2022 Ultrasonic array imaging of highly attenuative materials with spatio-temporal singular value decomposition *Ultrasonics* **124** 106764
[10] Ramalli A, Boni E, Roux E, Liebgott H and Tortoli P 2022 Design, implementation and medical applications of 2-D ultrasound sparse arrays *IEEE Trans. Ultrason. Ferr. Freq. Control* **69** 2739–55
[11] Guo S, Ding H, Li Y, Feng H, Xiong X, Su Z and Feng W 2022 A hierarchical deep convolutional regression framework with sensor network fail-safe adaptation for acoustic-emission-based structural health monitoring *Mech. Syst. Signal Process.* **181** 109508
[12] Liu H and Zhang Y 2019 Deep learning based crack damage detection technique for thin plate structures using guided lamb wave signals *Smart Mater. Struct.* **29** 015032
[13] Beniwal S and Ganguli A 2015 Defect detection around rebars in concrete using focused ultrasound and reverse time migration *Ultrasonics* **62** 112–25
[14] He J, Rocha D C and Sava P 2020 Guided wave tomography based on least-squares reverse-time migration *Struct. Health Monit.* **19** 1237–49
[15] He J, Rao J, Fleming J D, Gharti H N, Nguyen L T and Morrison G 2021 Numerical ultrasonic full waveform inversion (FWI) for complex structures in coupled 2D solid/fluid media *Smart Mater. Struct.* **30** 085044
[16] Song H and Yang Y 2021 Noncontact super-resolution guided wave array imaging of subwavelength defects using a

multiscale deep learning approach *Struct Health Monit.* **20** 1904–23
[17] Tarantola A 1984 Inversion of seismic reflection data in the acoustic approximation *Geophysics* **49** 1259–66
[18] Virieux J and Operto S 2009 An overview of full-waveform inversion in exploration geophysics *Geophysics* **74** WCC1–WCC26
[19] Tromp J, Komatitsch D and Liu Q 2008 Spectral-element and adjoint methods in seismology *Commun. Comput. Phys.* **3** 1–32
[20] Tromp J, Tape C and Liu Q 2005 Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels *Geophys. J. Int.* **160** 195–216
[21] Virieux J, Asnaashari A, Brossier R, Métivier L, Ribodetti A and Zhou W 2017 An introduction to full waveform inversion *Encyclopedia of Exploration Geophysics* (Society of Exploration Geophysicists) pp R1–1
[22] Guasch L, Calderón Agudo O, Tang M-X, Nachev P and Warner M 2020 Full-waveform inversion imaging of the human brain *npj Digit. Med.* **3** 28
[23] Operto S *et al* 2023 Is 3d frequency-domain FWI of full-azimuth/long-offset OBN data feasible? the gorgon data FWI case study *Lead. Edge* **42** 173–83
[24] Guo G, Operto S, Gholami A and Aghamiry H S 2023 Time-domain extended-source full-waveform inversion: algorithm and practical workflow *Geophysics* **89** 1–94
[25] Xu K and McMechan G A 2014 2d frequency-domain elastic full-waveform inversion using time-domain modeling and a multistep-length gradient approach *Geophysics* **79** R41–R53
[26] Operto S, Gholami Y, Prieux V, Ribodetti A, Brossier R, Metivier L and Virieux J 2013 A guided tour of multiparameter full-waveform inversion with multicomponent data: from theory to practice *Leading Edge* **32** 1040–54
[27] Rao J, Yang J, Ratassepp M and Fan Z 2020 Multi-parameter reconstruction of velocity and density using ultrasonic tomography based on full waveform inversion *Ultrasonics* **101** 106004
[28] Aktharuzzaman M, Anwar S, Borisov D and He J 2024 Experimental full waveform inversion for elastic material characterization with accurate transducer modeling *Mech. Syst. Signal Process.* **213** 111320
[29] Sun M, Yang J, Dong L, Liu Y and Huang C 2017 Density reconstruction in multiparameter elastic full-waveform inversion *J. Geophys. Eng.* **14** 1445–62
[30] He J, Borisov D, Fleming J D and Kasemer M 2022 Subsurface polycrystalline reconstruction based on full waveform inversion-a 2D numerical study *Materialia* **24** 101482
[31] Nguyen L T and Modrak R T 2018 Ultrasonic wavefield inversion and migration in complex heterogeneous structures: 2D numerical imaging and nondestructive testing experiments *Ultrasonics* **82** 357–70
[32] Agudo O C, Guasch L, Huthwaite P and Warner M 2018 3d imaging of the breast using full-waveform inversion *Proc. International Workshop on Medical Ultrasound Tomography* pp 99–110
[33] Fincke J, Zhang X, Shin B, Ely G and Anthony B W 2021 Quantitative sound speed imaging of cortical bone and soft tissue: results from observational data sets *IEEE Trans. Med. Imaging* **41** 502–14
[34] Bader K B, Crowe M J, Raymond J L and Holland C K 2016 Effect of frequency-dependent attenuation on predicted histotripsy waveforms in tissue-mimicking phantoms *Ultrasound Med. Biol.* **42** 1701–5
[35] Li F, Villa U, Park S and Anastasio M A 2021 3-d stochastic numerical breast phantoms for enabling virtual imaging trials of ultrasound computed tomography *IEEE Trans. Ultrason. Ferroelectr. Freq. Control.* **69** 135–46
[36] Bernard S, Monteiller V, Komatitsch D and Lasaygues P 2017 Ultrasonic computed tomography based on full-waveform inversion for bone quantitative imaging *Phys. Med. Biol.* **62** 7011
[37] Zhou C and Xu K 2024 Sound speed imaging of cortical bone and soft tissue based on frequency-domain ultrasound waveform tomography *J. Phys: Conf. Ser.* **2822** 012007
[38] Kormann J, Rodríguez J E, Ferrer M, Farrés A, Gutiérrez N, de la Puente J, Hanzich M and Cela J M 2017 Acceleration strategies for elastic full waveform inversion workflows in 2D and 3d: near offset elastic full waveform inversion *Comput. Geosci.* **21** 31–45
[39] Suzuki K 2017 Overview of deep learning in medical imaging *Radiol. Phys. Technol.* **10** 257–73
[40] Song C and Alkhalifah T A 2021 Wavefield reconstruction inversion via physics-informed neural networks *IEEE Trans. Geosci. Remote Sens.* **60** 1–12
[41] Tong J, Lin M, Wang X, Li J, Ren J, Liang L and Liu Y 2022 Deep learning inversion with supervision: a rapid and cascaded imaging technique *Ultrasonics* **122** 106686
[42] Robins T, Camacho J, Agudo O C, Herraiz J L and Guasch L 2021 Deep-learning-driven full-waveform inversion for ultrasound breast imaging *Sensors* **21** 4570
[43] Kleman C, Anwar S, Liu Z, Gong J, Zhu X, Yunker A, Kettimuthu R and He J 2023 Full waveform inversion-based ultrasound computed tomography acceleration using 2d convolutional neural networks *J. Nondestruct. Eval. Diagn. Progn. Eng.* **6** 1–13
[44] Deng C, Feng S, Wang H, Zhang X, Jin P, Feng Y, Zeng Q, Chen Y and Lin Y 2021 Openfwi: large-scale multi-structural benchmark datasets for seismic full waveform inversion
[45] Wang Y-Q, Wang Q, Lu W-K, Ge Q and Yan X-F 2022 Seismic impedance inversion based on cycle-consistent generative adversarial network *Pet. Sci.* **19** 147–61
[46] Zhang Z and Lin Y 2020 Data-driven seismic waveform inversion: a study on the robustness and generalization *IEEE Trans. Geosci. Remote Sens.* **58** 6900–13
[47] Zhang T, Sun J, Innanen K and Trad D 2021 A recurrent neural network for $\ell 1$ anisotropic viscoelastic full-waveform inversion with high-order total variation regularization pp 1374–8
[48] Wang Z, Xiao J, Li D, Li B, Zhang J and Ta D 2023 Full waveform inversion guided wave tomography with a recurrent neural network *Ultrasonics* **133** 107043
[49] Raissi M, Perdikaris P and Karniadakis G E 2019 Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations *J. Comput. Phys.* **378** 686–707
[50] Wang W, McMechan G A and Ma J 2021 Elastic isotropic and anisotropic full-waveform inversions using automatic differentiation for gradient calculations in a framework of recurrent neural networks *Geophysics* **86** R795–810
[51] Liu Z, Bicer T, Kettimuthu R, Gursoy D, De Carlo F and Foster I 2020 Tomogan: low-dose synchrotron x-ray tomography with generative adversarial networks: discussion *J. Opt. Soc. Am.* A **37** 422–34
[52] Luo Y, Tromp J, Denel B and Calandra H 2013 3d coupled acoustic-elastic migration with topography and bathymetry based on spectral-element and adjoint methods *Geophysics* **78** S193–202
[53] Bozdağ E, Trampert J and Tromp J 2011 Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements *Geophys. J. Int.* **185** 845–70
[54] Pladys A, Brossier R, Li Y and Métivier L 2021 On cycle-skipping and misfit function modification for full-wave inversion: comparison of five recent approaches *Geophysics* **86** R563–87

[55] Modrak R T, Borisov D, Lefebvre M and Tromp J 2018 Seisflows-flexible waveform inversion software *Comput. Geosci.* **115** 88–95
[56] Shan H, Zhang Y, Yang Q, Kruger U, Kalra M K, Sun L, Cong W and Wang G 2018 3-D convolutional encoder-decoder network for low-dose CT via transfer learning from a 2-D trained network *IEEE Trans. Med. Imaging* **37** 1522–34
[57] Yi X, Walia E and Babyn P 2019 Generative adversarial network in medical imaging: a review *Med. Image Anal.* **58** 101552
[58] Mathieu M, Couprie C and LeCun Y 2015 Deep multi-scale video prediction beyond mean square error (arXiv:1511.05440)
[59] Sánchez I and Vilaplana V 2018 Brain mri super-resolution using 3D generative adversarial networks (arXiv:1511.05440)
[60] Simonyan K and Zisserman A 2014 Very deep convolutional networks for large-scale image recognition (arXiv:1511.05440)
[61] Akiba T, Sano S, Yanase T, Ohta T and Koyama M 2019 Optuna: a next-generation hyperparameter optimization framework *Proc. 25th ACM SIGKDD Int. Conf. on Knowledge Discovery Data Mining* pp 2623–31
[62] Zhang C, Bengio S, Hardt M, Recht B and Vinyals O 2021 Understanding deep learning (still) requires rethinking generalization *Commun. ACM* **64** 107–15
[63] Shen Z, Liu J, He Y, Zhang X, Xu R, Yu H and Cui P 2021 Towards out-of-distribution generalization: a survey (arXiv:2108.13624)
[64] Arjovsky M 2020 Out of distribution generalization in machine learning *PhD Thesis* New York University