Implicit Seismic Full Waveform Inversion With Deep Neural Representation

Jian Sun¹, Kristopher Innanen², Tianze Zhang³, and Daniel Trad⁴

¹College of Marine Geosciences, Key Lab of Submarine Geosciences and Prospecting Techniques MOE China, Ocean University of China, Qingdao, China. ²Department of Geoscience, University of Calgary, Calgary, AB, Canada

Abstract Full waveform inversion (FWI) is arguably the current state-of-the-art amongst methodologies for imaging subsurface structures and physical parameters with seismic data; however, important challenges are faced in its implementation and use. Keys amongst these are (a) building a suitable initial model, from which a local minimum is unlikely to be reached, and (b) availability of tools for evaluation of uncertainty. An algorithm we refer to as implicit full waveform inversion (IFWI), designed using continuously and implicitly defined deep neural representations, appears in principle to address both of these issues. We observe in IFWI, with its random initialization and deep learning optimization, improved convergence relative to standard FWI model initialization and optimization. Models close to the global minimum, capturing relatively high-resolution subsurface structures, are obtained. In addition, uncertainty analysis, though not solved in IFWI, is meaningfully addressed by approximating Bayesian inference with the addition of dropout neurons. Numerical experimentation with a range of 2D geological models is suggestive that IFWI exhibits a strong capacity for generalization, and is likely well-suited for multi-scale joint geophysical inversion.

Plain Language Summary We propose implicit full waveform inversion (IFWI) by replacing the grid-based subsurface parameters with a continuous neural network representation. Compared to the conventional full waveform inversion (FWI), this simple reparameterization allows IFWI to start from a random initial model by benefiting from the frequency principle of deep learning optimization. The uncertainty of inversion results using IFWI can be easily performed by adopting the Bayesian neural network, or adding dropout neurons into the Multilayer Perceptron. In addition, one can use a single network to represent physical parameters at different scales, thus accommodating multi-scale and multi-physics problems. Synthetic examples demonstrate that IFWI is able to produce a high-resolution image of subsurface with fine structures, and has strong generalization ability and a certain degree of robustness.

1. Introduction

Imaging high-resolution heterogeneities in the Earth's subsurface is a key objective of geophysical inversion. Images are typically obtained through a process of iterative optimization, in which a function of the misfit between simulated data and observations is minimized. Full waveform inversion (FWI) is arguably the current state-of-the-art methodology (Tarantola, 1984; Virieux & Operto, 2009). Optimization in FWI is normally carried out through an adjoint state method (Q. Liu & Tromp, 2006; Plessix, 2006; Yedlin & Van Vorst, 2010). It is based on the assumption of convexity, which is inconsistent with the non-linearity of its objective function, leading to a range of well-known challenges, such as cycle-skipping. This can be reduced by involving low-frequency/high-angle measurements, and/or starting from an initial model that contains wavenumber components missing from the data (Virieux & Operto, 2009). However, the acquisition of low-frequency geophysical data remains a challenge because of the associated increase in difficulty to characterize noise. A common mitigating practice is formulation of new objective functions, including for instance regularization terms (Esser et al., 2018; Lin & Huang, 2014) or envelope-based misfit functions (Bozdag et al., 2011; Chi et al., 2014). Other adaptations include reflection waveform inversion (e.g., RWI, Chi et al., 2015; Xu et al., 2012), and adaptive waveform inversion (e.g., AWI, Guasch et al., 2019; Warner & Guasch, 2016), which make physics based assumptions that force FWI optimization to incorporate low-wavenumber model components. The other alternative solution is to build increasingly realistic initial models using other methods. For instance, to invert a smooth model, one can use traveltome tomography by matching the first-break information (Clement et al., 2001) or adopt migration velocity analysis by measuring the flatness of the common-image gathers (Symes, 2008). In addition to cycle-skipping, FWI in general is computationally demanding, which tightly constrains current FWI algorithms.
Deep learning, primarily in the form of deep neural networks (DNNs), is drawing widespread attention in scientific and engineer fields, such as image classification/recognition, shape representation, self-driving, machine translation, and natural language processing. For instance, in shape representation, coordinate-based DNNs are trained to learn continuous signed distance functions from a single or multi-view images (Chen & Zhang, 2019; Mescheder et al., 2019; Mildenhall et al., 2020; Park et al., 2019; Sitzmann et al., 2020). After training, such a DNN can be treated as an implicit representation of an image or a 3D object with respect to coordinates, which provides a new perspective for image/object reconstruction. A strong effort has been made in recent years to facilitate deep learning applications in geophysics and its practitioners, such as automatic first-arrival picking (Cano et al., 2021; H. Wu, Zhang, Li, & Liu, 2019; Yuan et al., 2020), seismic facies classification (Alaudah et al., 2019; Feng et al., 2021; M. Liu et al., 2020), and noise attenuation (Saad & Chen, 2020; H. Wu, Zhang, Lin, et al., 2019; L. Yang et al., 2021)—domains for which data are plentiful but underlying physical rules are not explicitly known. Deep learning has had significant successes in such domains, in which problems to be solved can be treated/converted into the cognitive tasks of pattern recognition. Attempts have also been made to devise a DNN architecture that directly maps seismograms to subsurface models in a fully data-driven sense (Wang et al., 2018; Y. Wu et al., 2018; F. Yang & Ma, 2019). However, due to the ill-posedness and complexity of seismic inversion, training a properly-generalized and robust DNN for inversion requires vast numbers of highly variable subsurface model/seismogram pairs. Building a dataset that contains all possible combinations of subsurface structures is in practice not possible.

In comparison with the purely data-driven approaches, which are unlikely to be truly predictive, FWI (as a large optimization problem) exhibits persistent predictability and a strong capacity to generalize, because of the guidance it has from governing physical equations (i.e., wave propagation). This is strongly suggestive of the importance of including underlying physical principles when pursuing a new generation of data-centric deep learning inversion. With this in mind, several ways of incorporating FWI with DNNs have been explored. Sun et al. (2019); Sun, Niu, et al. (2020) proposed a theory-guided seismic inversion framework in which the forward modeling is devised in a framework of recurrent neural network (RNN) and the inversion process is described as the RNN training using automatic differentiation that can be easily accelerated by deep learning platforms. The theory-guided RNN framework can also be extended to more complete multidimensional parameter estimations (Zhang et al., 2020) or electromagnetic inversion (Hu et al., 2021). Y. Wu and McMechan (2019), He and Wang (2021), and Zhu et al. (2022) parameterized velocity models with a convolutional neural network (CNN), which creates a multi-grid representation of the velocity and automatically includes regularization effects in the FWI procedure. Sun et al. (2021) developed a hybrid network design by simultaneously involving a data-driven model misfit and a physics-guided data residual during the training of the network. However, while these efforts set out important methodologies, they by and large encounter the same challenges as traditional FWI (as listed above). For instance, they generally require accurate initial models to pre-train the CNN, and do not naturally permit evaluation of uncertainty.

The goal of this paper is to further explore deep fusion of FWI and DNNs, in a way that addresses the challenges of FWI in initial model building and uncertainty analysis, while retaining the strong robustness and generalization ability of FWI. According to the universal approximation theorem, DNNs are able to represent any continuous functions. For instance, in shape representation, the surface of the target object can be interpreted as a continuous signed distance function. Analogously, we are able to build an implicit functional space using DNNs that represent various parameterizations of the subsurface model. Comparing to the CNN reparameterization, which lacks the flexibility (i.e., supports only fixed-size outputs and has the difficulty in transferring to other models) and requires initial models for pretraining, the implicit function built with DNNs can be adapted to physical parameterizations and models of arbitrary size, due to the properties of continuous functions. In addition, we also seek an implicit function using DNNs to reduce the dependency of FWI on the initial model by increasing the numbers of degree of freedoms in seismic inversion.

This paper is organized as follows. First, we introduce the concept of deep neural representation (DNR), and then demonstrate how to build FWI with implicit neural representation. Second, we discuss the network selections for DNR and examine its capacity of representing subsurface geological models. After that, experimental examples

(Byrd et al., 1995; Métivier et al., 2013; Nocedal & Wright, 2006). Beyond solving for an optimal subsurface model, the critical task of evaluating uncertainty is a serious challenge in FWI problems, driven by the high dimensionality of model space.
are presented to exemplify the performance of the proposed method, including random initialization, robustness, uncertainty evaluation, and generalization ability. Finally, we discuss its further potentials as well as challenges.

2. Methodology

In this section, we introduce the theory of DNR, and present the mathematical principle of implicit full waveform inversion (IFWI) by deeply integrating DNR and FWI. Next, we discuss the network design for IFWI, and explain why IFWI has the ability to avoid convergence to local minima even with a randomly initialized starting model. To make formulas easier to follow, all bold symbols represent vectors, and all mathematical italics are functions.

2.1. Deep Neural Representation

Assume we are interested in a set of features \( \Psi \) that can be interpreted as a continuous function \( \Phi \), with respect to the input \( x \in \Omega_k \), \( k = 1, \ldots, K \), which is implicitly defined by equations of the form

\[
\mathcal{L}_k (x, \Phi, \nabla_x \Phi, \nabla^2_x \Phi, \ldots) = 0, \quad \Phi : x \to \Phi(x)
\]

where the derivation of \( \mathcal{L}_k \) can be purely data-driven or physics-deterministic and the physical meaning of the input \( x \) relies on features of interest \( \Psi \) to be represented. For example, suppose we are interested in the RGB values of an image, which is divided into \( K \) sub-regions with \( \Omega_k \) representing the input space of the \( k \)th sub-region. \( \Psi \) is the RGB values to be represented and \( \Phi \) is the coordinate-based function of RGB. \( \mathcal{L}_k \) represents the known constraint corresponding to the \( k \)th sub-region.

Based on the universal approximation theorem, there must exist a neural network that is equipped to map \( x \) to the quantity of interest \( \Psi \) while satisfying constraints shown in Equation 1. Learning a network that parameterizes an implicitly defined function \( \Phi \) is referred to as DNR. The training process of such a network denoted as \( \mathcal{N}_\theta \) can be implemented by minimizing

\[
\arg \min_{\theta} \sum_{k=1}^{K} \lambda_k \| \mathcal{L}_k \left( x, \mathcal{N}_\theta, \nabla_x \mathcal{N}_\theta, \nabla^2_x \mathcal{N}_\theta, \ldots \right) \|^2, \quad \mathcal{N}_\theta : x \to \mathcal{N}_\theta(x)
\]

where \( \theta \) are the trainable weights and biases of the neural network, \( \lambda_k \) denotes the trade-off parameter.

With proper training, the DNR-based network \( \mathcal{N}_\theta \) is approximately equivalent to the implicitly defined function \( \Phi \) for the input domains of \( x \), that is, \( \mathcal{N}_\theta(x) \approx \Phi(x), \forall x \in \Omega_k \). It indicates that we may use a continuous and differentiable function to represent features of interest for a certain input domain. Compared to the discrete parameterization, in which memory and precision are highly dependent on the grid resolution, a continuous representation \( \mathcal{N}_\theta(x)/\mathcal{N}_\theta(x) \) on the continuous domain of \( x \) can be much more efficient while preserving fine details. For instance, instead of saving large grid-based wavefields, a better alternative is saving the continuous implicit representation of wavefields constrained by wave equations, which can be commonly solved in the framework of physics-informed neural networks (PINNs). Thus, PINNs are special forms of DNRs, in which constraints \( \mathcal{L}_k \) are determined by the underlying physical principles and the input domains are usually defined as the spatial or spatio-temporal coordinates.

2.2. Implicit Full Waveform Inversion

DNRs are also well suited to a wide variety of problems in science and engineering. For instance, in geophysical inverse problems, the features of interest \( \Psi \) are physical properties of the subsurface model denoted as \( \textbf{m} \), where the constraints \( \mathcal{L} \) can be determined by wave propagation theory in FWI or by Zoeppritz equation in amplitude-versus-offset (AVO) inversion. We refer to FWI using implicit deep neural representation as IFWI. In IFWI, Equation 1 can be rewritten as

\[
\mathcal{R}(\textbf{m}, \textbf{s}, x, t) - \textbf{d} = 0, \quad \textbf{m} : x \to \textbf{m}(x)
\]

where \( \mathcal{R} \) denotes the forward modeling operator of wave propagation, \( x \) and \( t \) are spatial and temporal coordinates, respectively, \( \textbf{s} \) represents the source information, \( \mathcal{R} \) denotes the matrix of receiver layouts, and \( \textbf{d} \) is the observed data.
The optimal implicit representation are obtained by minimizing the objective function of IFWI, using
\[
\arg \min_{\mathbf{a}} \| R F (\mathcal{N}_0, \mathbf{s}, \mathbf{x}, t) - \mathbf{d} \|^2, \quad \mathcal{N}_0 : \mathbf{x} \rightarrow \mathcal{N}_0(\mathbf{x}) \approx \mathbf{m}(\mathbf{x})
\]  
(4)

Equation 4 indicates that we are seeking a continuous and implicit functional representation, instead of a grid-based solution, of subsurface parameters using IFWI. Note that IFWI allows its optimization to be performed in a mesh-free manner as long as a mesh-free forward operator \( F \) (for instance, with a PINN solver) is adopted. However, to concentrate on the validity of DNR, a time-domain grid-based forward modeling is employed using finite difference method in this paper.

Sun, Niu, et al. (2020) indicates that, by designing the time cell of the RNN as a finite difference operator with a single time step, in which the subsurface model \( \mathbf{m} \) is the trainable parameter of RNN, the theory-based RNN simulates the forward modeling of seismic waves in the time domain. Thus, training the theory-based RNN is theoretically equivalent to performing a conventional FWI. Compared to the conventional FWI, there are a few advantages of using the theory-based RNN. For instance, the theory-based RNN can be easily deployed on the deep learning platform with GPU acceleration. In addition, with the theory-based RNN, the model updates can be automatically computed using automatic differentiation algorithm, without the need to manually derive and calculate the gradient, which may provide important benefits for FWI in complex media. Therefore, the forward simulation of wave propagation in this paper is embedded into the theory-based RNN, as shown in the work of Sun, Niu, et al. (2020), to make the best use of deep learning platform.

### 2.3. Network Design

Multilayer perceptrons (MLPs) appear to have strong potential in DNR applications, such as shape representation (Chen & Zhang, 2019; Genova et al., 2019; Park et al., 2019), object reconstruction (Mescheder et al., 2019; Xie et al., 2019), and scene representation (Mildenhall et al., 2020; Sitzmann et al., 2019). Due to its continuous and memory-efficient characteristics, we select MLPs as the primary architecture for IFWI. A MLP is a type of feed-forward artificial neural network that is composed of multiple layers of interconnected artificial neurons, in which information only flows in one direction from input to output, through hidden layers of neurons. Mathematically, each neuron in the MLP performs a weighted sum of its inputs (i.e., all neurons in the previous layer), followed by a nonlinear activation function, to produce its output:

\[
a'_i = f ( \mathbf{w}_i \mathbf{a}^{l-1} + b'_i)
\]

where \(a'_i\) represents the output of the \(i\)th neuron in the \(l\)th layer, \(\mathbf{a}^{l-1}\) indicates the vector of all neurons in the \(l-1\)st layer in shape of \(n \times 1\), \(\mathbf{w}_i\) is in size of \(1 \times n\) representing the vector of weights applied to \(\mathbf{a}^{l-1}\) for summation, \(b'_i\) is the scalar bias corresponding to \(a'_i\), \(f\) indicates the nonlinear activation function.

The neural network to represent the subsurface model takes the spatial coordinates (or spatio-temporal coordinates in the time-lapse IFWI) as input, and outputs the physical parameters. To learn a continuous and implicit function of complex distributed subsurface parameters, MLPs must be equipped with nonlinear activation functions. Commonly activation functions are sigmoid, hyperbolic tangent (tanh), rectified linear unit (ReLU), and its variants. Of these, ReLU is widely used, as the simplicity of its derivatives prevents the gradient from vanishing during backpropagation. However, Sitzmann et al. (2020) demonstrated that MLPs with these non-periodic activation functions have difficulties in learning high-frequency components and high-order derivatives of images or scenes to be represented. Inspired by the discrete cosine transform (Klocek et al., 2019), Sitzmann et al. (2020) proposed MLPs with periodic activation functions, especially MLPs with sinusoidal functions (also known as sinusoidal representation network, or SIREN), and demonstrated that the periodic characteristic of activation functions offer benefits in learning high-frequency and high-order derivatives information, comparing favorable to non-periodic activation functions.

To re-exemplify the merits of using periodic activation functions in subsurface representation, we build a SIREN to reconstruct the physical parameterizations of the 2D Marmousi model from the spatial coordinates in the horizontal and depth directions. To enlarge the output range of the proposed SIREN, which is limited by the sinusoidal activation function, a linear output layer is selected, that is, no activation function is applied to the output layer. Furthermore, instead of directly exporting the compressional wave velocity \(V_p\), the SIREN outputs its normalization. Here, the mean and standard deviation of compressional wave velocity of the Marmousi model are utilized to perform the normalization.
MLPs with four hidden layers, each of which contains 128 neurons, are selected for subsurface parameterizations of the acoustic Marmousi model. Proposed MLPs are equipped with sinusoidal activation functions and ReLUs, respectively, where weights of layers in the ReLU-based MLP are initialized using a uniform distribution $\alpha_i \sim U(-\sqrt{1/n}, \sqrt{1/n})$ and weights of layers in the SIREN are also uniformly initialized but with an alternative bound $\alpha_i \sim U(-\sqrt{6/n}, \sqrt{6/n})$ by following the principled scheme in the work of Sitzmann et al. (2020). Here, $n$ indicates the number of input features fed into the layer being initialized. Besides that, to further accelerate the frequency learning throughout the SIREN, a fixed tuning weight $\omega_0 = 30$ is added into the sinusoidal activation functions $\sin(\omega_0 \cdot Wx + b)$ of the first hidden layer. Refer to Sitzmann et al. (2020) for more details on initialization of the SIREN. A variant of the gradient decent algorithm, adaptive momentum (Adam) optimizer, is selected to train both MLPs using their respective optimal learning rates discovered through extensive trial and error runs (in this case, $1 \times 10^{-3}$ for the ReLU-based MLP and $1 \times 10^{-4}$ for the SIREN).

In the subsurface parameter representation, the SIREN and ReLU-based MLP take approximately 5,000 and 50,000 training epochs, respectively, until convergence occurs. Since they are almost equally efficient in forward and backward propagations, the SIREN exhibits a convergence rate 10 times faster than the ReLU-based MLP. The final representations using properly-trained networks are shown in Figure 1. In the first row of Figure 1, we compute the normalization, the gradients, and the Laplacian of $V_p$ using a grid-based Marmousi model. The second and third rows of Figure 1 show implicit representations of Marmousi and their gradients and Laplacian computed using the SIREN and the ReLU-based MLP, respectively. By comparing the first column of Figure 1, we conclude that both networks have the ability to reconstruct primary structures of the subsurface. However, the SIREN outperforms the ReLU-based MLP in the representation of fine structures, as shown in the yellow box. A similar performance is observed in the second column of Figure 1 that the SIREN is able to reconstruct the gradients of $V_p$ in finer detail, while the ReLU-based MLP can only provide a fraction of them. Moreover, in the representation of high-order derivatives shown in the third column of Figure 1, the SIREN provides the Laplacian of $V_p$ with a high resolution, where the ReLU-based MLP fails.

In Figure 2, we show the enlarged view of the selected region indicated by the yellow box in Figure 1, which further confirms the capability of the SIREN in fine structure representation. In addition, the serrated effects can

---

**Figure 1.** Comparison of implicit neural representation using different activation functions to fit a 2D acoustic Marmousi model. Columns from left to right are normalized $V_p$, gradients of $V_p$, and the Laplacians of $V_p$; rows from top to bottom correspond to ground truth, representations using a SIREN, and representations using a ReLU-based MLP.

**Figure 2.** An enlarged view of the selected area indicated by the yellow box in Figure 1.
be observed in the discrete grid-based parameterizations (the left panel of Figure 2) due to the grid limited resolution. Figure 2 indicates that implicit representations using MLPs are able to exhibit a certain degree of continuity, although slight discontinuities can be observed in the SIREN representation (the middle panel of Figure 2), which are likely caused by the overfitting and will be discussed later.

In the above experiments, with the same number of hidden layers and neurons in each layer, we observe that the SIREN outperforms the ReLU-based MLP in representing subsurface parameterizations, including a faster convergence rate and the ability to outline fine structures and high-order derivatives. Therefore, the SIREN with the same architecture (i.e., containing four hidden layers with 128 neurons in each) is applied to the rest of the work in this paper. The input and output layers are determined by the number of input and output features, respectively. For instance, the input layer contains two neurons if the two-dimensional spatial coordinates are used as inputs; the output layer contains three neurons if we are reconstructing the parameterizations of density, compressional- and shear-wave velocities. The network architecture chosen here has not been established as optimal, and better performance may be achieved using other classic architectures, such as CNNs, vision Transformers, and graph neural networks (GNNs). This is a matter for future research.

2.4. Why IFWI Works

The IFWI workflow is schematically shown in Figure 3. First, we feed the spatial coordinates (x and z) into the randomly initialized SIREN, and then perform the anti-normalization using a mean and a standard deviation. The mean and standard deviation can be calculated using well-log information. Without any prior knowledge, one can use a global mean (μ = 3.0 km/s) and a global standard deviation (σ = 1.0 km/s). From experience we believe that the global mean and standard deviation are good enough for most subsurface models, as long as they cover the minimum and maximum values of the model to be represented, since they are only used to shift and scale the distribution of the output features. In the following, we use the global mean and standard deviation for all Marmousi-related experiments. Second, the parameterizations constructed by the SIREN are used as input to forward modeling for simulated data generation. Third, we compute the discrepancy between simulated data and observation and backpropagate it to update weights of SIREN. Next, repeat all above steps until convergence.

Geophysical inversion can be understood as the problem of selecting a particular model which fits data acceptably. The fact that we observe bandlimited data makes geophysical inversion a non-convex optimization problem, which generally suffers from non-uniqueness and is prey to local minima. Thus, good initial models or low-frequency components are required. IFWI addresses this with randomly initialized models, without pretraining the network. Furthermore, the same scheme can be easily generalized in other kinds of geophysical inversion. In the following, we demonstrate this both theoretically and numerically.

Theoretically, the reason for this is that a critical point, where the gradient is zero, is unlikely to be a local minimum in deep learning optimization. For a critical point x ∈ R^n in n dimensions to be a local minimum, the Hessian matrix H(x) must be positive semi-definite. Due to its symmetric characteristics, the Hessian at such a critical point can be diagonalized:

\[
\begin{pmatrix}
\lambda_1 \\
\vdots \\
\lambda_n
\end{pmatrix}
\]

where all elements of H(x) are non-negative, that is, \(\lambda_i \geq 0\) for 0 \(\leq i \leq n\).

Considering the high dimensional non-linearity of the Hessian, we assume that the signs of the elements in H(x) are independent and the probability of each of them being non-negative is 1/κ with \(\kappa > 1\). Thus, the probability of a given critical point being a local minimum can be expressed as

\[
P(x_{\text{min}}) = P(\lambda_1 \geq 0, \lambda_2 \geq 0, \ldots, \lambda_n \geq 0) = \prod_{i=1}^{n} P(\lambda_i \geq 0) = \frac{1}{\kappa^n}
\]

Figure 3. The workflow of implicit full waveform inversion with deep neural representation.
Equation 7 states that the probability of a critical point being a local minimum decreases exponentially with increasing dimensionality. The DNN usually consists of more than million of trainable parameters, which define an ultra-high dimensional space (i.e., $n > 10^6$). In other words, a critical point in the deep learning optimization is most likely a saddle point, instead of a local minimum, that is, $\lambda \to 0$. We emphasize that Equation 7 is not a rigorous proof and certain types of saddle points may also pose challenges for deep learning optimization. More theoretical analysis about convergence of deep learning are investigated based on landscape conjecture (Bhojanapalli et al., 2016; Choromanska et al., 2015; Ge et al., 2015, 2016) or trajectory (Arora et al., 2018; Brutzkus et al., 2017; Brutzkus & Globerson, 2017; Li & Yuan, 2017; Tian, 2017). However, it is fair to say that IFWI may greatly reduce the non-uniqueness of inversion results and the risk of falling into local minima during the optimization.

A simple example can serve to illustrate the nature of the neural network learning process. Suppose we want a neural network to learn a continuous signal, such as an oscillating sinusoidal function, a 1D seismic trace, and a 1D velocity profile. A SIREN is adopted and its weights are optimized by minimizing the misfits between mimic signals and the ground truth. We record the learning process of the neural network in tasks of representing the three signals mentioned above, in which the oscillated sinusoidal signal is specifically designed as a combination of a low-frequency sinusoidal signal and a high-frequency sinusoidal signal.

In Figure 4 the learning process of the network representing these three signals, and a comparison between learned and ground-truth signals at three different training epochs, are illustrated. During the training of neural networks, we observe that low-frequency information are always learned first compared to high-frequency information. This phenomenon is known as spectral/frequency bias (Basri et al., 2020; Rahaman et al., 2019). For simplicity, we refer to it as frequency bias in the following contexts. It indicates that frequency bias may ensure IFWI firstly updates the low-wavenumber components, rather than the high-wavenumber components, of subsurface models. Thus, both theoretical and numerical results indicate that IFWI has the ability to converge using a random initialization.

3. Numerical Examples

In this section, we further examine the effectiveness and potentials of IFWI using a 2D acoustic Marmousi model with a grid size $94 \times 288$ and 15 m cells, shown in Figure 5. The 13 synthetic shot gathers are collected with a shot interval of 300 m at a depth of 30 m using a Ricker wavelet with 8 Hz dominant frequency, where shot locations are indicated as the golden stars in Figure 5. Receivers are placed on the ground surface, occupying each
A free surface at the top and a perfectly matched layer (PML) absorbing condition at three other boundaries are applied during the synthetic seismic record collection. The time sample interval during wave propagation is 1.9 ms, and 1,000 time samples are recorded. Figure 6 shows six representative examples of 13 shot gathers collected using Marmousi model. To ensure the fairness of all comparisons, both FWI and IFWI are performed with a theory-designed RNN framework using their respective optimal learning rates. All experiments in this paper are carried out using PyTorch on a Centos system with a GPU RTX 3090. For general information on the computational overhead for all experiments in this paper, FWI and IFWI require approximately 5.4 and 5.6 min, respectively, to complete 100 epochs of the Marmousi-related tests.
3.1. Random Initialization

For a fair comparison with FWI, we first perform both FWI and IFWI with a smooth initial model, which is obtained by applying a Gaussian filter on the ground truth model shown in Figure 5. The computed smooth initial model is plotted in Figure 7a. For the implementation of FWI with a smooth initial model (refer to as FWI-Smooth), an optimal learning rate of 0.01 is applied based on the instructive analysis of Sun, Niu, et al. (2020), and the final inversion result of FWI is plotted in Figure 7b. For IFWI implementation, the smooth initial model shown in Figure 7a is employed to pretrain the SIREN, which is then embedded into the IFWI procedure. With trial-and-error experiments, a same learning rate of 0.0001 is adopted for both pretraining and IFWI processes. The final inversion result of IFWI using a SIREN pretrained by a smooth initial model (refer to as IFWI-Pretrain) is plotted in Figure 7c. As indicated in the first column of Figure 7, given a good initial model, both IFWI and FWI have the ability to converge to the global minimum and produce satisfactory results. In addition, another experiment is designed to examine the performance of IFWI with randomly initialized starting model. The initial model shown in Figure 7d is randomly generated using a normal distribution and then is anti-normalized with the global mean and standard deviation. Both FWI and IFWI with this random initialization (refer to as FWI-Random and IFWI-Random, respectively) are performed using their respective optimal learning rates (0.001 and 0.0001, respectively) and their final results are plotted in Figures 7e and 7f, respectively. Figure 7e indicates that, without a good initial model, FWI can easily fall into local minima during the optimization process and cannot predict the adequate velocity model. Compared to FWI, Figure 7f shows that IFWI is able to reconstruct satisfactory results containing fine detailed structures even when starting from a random initial model without any pretraining process, which further validates our inference about IFWI. By further comparing Figures 7b and 7f, we learn that IFWI with a random initialization can produce comparable results to FWI with a good initial model.

Figure 8 shows the relative computational costs of four implementations mentioned previously, including FWI-Smooth, FWI-Random, IFWI-Pretrain,
and IFWI-Random. We observe that, when a good initial model is given, both FWI-Smooth and IFWI-Pretrain are able to converge quickly and with comparable efficiency, as delineated in blue and orange lines in Figure 8. When lacking a good initial model, it is difficult for FWI to escape from the trap of local minima, however, IFWI can gradually converge to the global minimum. Not surprisingly, IFWI-Random takes longer to converge, as it starts from a random initial model. Better convergence rates may be possible with an alternative neural network architecture; these are currently being investigated.

To analyze the inversion process of IFWI more intuitively, we plot inversion results at 500th, 1000th, 1500th, 2000th training epochs in Figure 9. The low-wavenumber information of the Marmousi model are evidently recovered after 500 epochs. With 1,000 epochs, IFWI is able to recover primary structures with robust layer information in shallow zones. After 1,500 epochs, the stratigraphy of shallow zones are refined with detailed information, and some of deep structures also appear. IFWI is able to further improve the precision of subsurface parameters and layers in deep zones after 2,000 epochs. The inversion process of IFWI on the 2D Marmousi model is also consistent with our conclusions and experiments on frequency bias in the previous section. This set of experiments shows that it is not only feasible for IFWI to represent the subsurface parameters by a continuous implicit neural network, but it also allows for a stochastic initial model due to the increased degrees of freedom.

3.2. Robustness

Next, we exemplify the robustness of IFWI algorithm by adding different levels of random noise into the observed data. Gaussian white noises are generated using standard deviations $\sigma = 2\sigma_0$ and $\sigma = 4\sigma_0$, respectively, where $\sigma_0$ represents for the standard deviation of the noise-free observed data. In Figure 10, three representative shot gathers with random noise of standard deviation $\sigma = 2\sigma_0$ (refer to as noisy-data $\sigma = 2\sigma_0$) are plotted in the top row, and data with random noise of standard deviation $\sigma = 4\sigma_0$ (refer to as noisy-data $\sigma = 4\sigma_0$) are plotted in the bottom row. For comparison purposes, both FWI and IFWI are performed using noisy observations as the ground truth signals in the calculation of data discrepancy, where the smooth initial model (Figure 7a) is employed during the implementation of FWI and IFWI utilizes the random initialization (Figure 7d) as the starting model.

Comparisons of final inversion results are shown in Figure 11, where FWI and IFWI are plotted from left to right by column, and the noise levels are differentiated in a row perspective. Specifically, inversion results using noise-free data, noisy-data $\sigma = 2\sigma_0$ and noisy-data $\sigma = 4\sigma_0$ are plotted in the first, second, and third rows in Figure 11. Figures 11b and 11c indicate that, given a proper initial model, FWI exhibits great tolerance to noise, although some small perturbations are observed in the final results. Compared to FWI, IFWI shows considerable robustness capability even with a random initial model, which are confirmed in Figures 11e and 11f. In this robustness experiments, both FWI and IFWI are implemented in a regularization-free manner, therefore, better results may be achieved if some form of regularization is added.

In order to analyze the convergence process from a macroscopic point of view, we plot the monitored objective losses with noisy observations for FWI and IFWI in Figure 12. Due to the dominant energy of the noise, only small changes of losses in percentage are observed. The top panel of Figure 12 shows data discrepancies with noisy-data $\sigma = 2\sigma_0$ vs. training epoch, in which we observe slightly faster convergence rates for both FWI and IFWI than their corresponding noise-free implementations, that is, FWI-Smooth and IFWI-Random delineated as blue and red lines in Figure 8, respectively. The objective losses of data discrepancies with
Figure 10. Representative examples of noisy shot gathers. Top row: 3 representative shots with random noise of standard deviation $\sigma = 2\sigma_0$, bottom row: 3 representative shots with random noise of standard deviation $\sigma = 4\sigma_0$.

Figure 11. Comparisons of inversion results for FWI and IFWI using noise-free and noisy observations. (a) FWI with noise-free data, (b) FWI with noisy-data $\sigma = 2\sigma_0$, (c) FWI with noisy-data $\sigma = 4\sigma_0$, (d) IFWI with noise-free data, (e) IFWI with noisy-data $\sigma = 2\sigma_0$, (f) IFWI with noisy-data $\sigma = 4\sigma_0$. 
3.3. Uncertainty

One of the most challenging tasks in FWI is the uncertainty assessment. This is due to the complexity and prohibitive computational cost of calculating the Hessian or posterior covariance matrix in the presence of very large numbers of subsurface parameters. Though the supervised deep learning aims to establish an end-to-end relationship from the input to output, which is generally deterministic, there are still several approaches to approximate the Bayesian inference. For instance, deep ensemble learning can produce collective predictions that approach Bayesian predictive distribution by retaining the same neural network multiple times using the respective data sets (Lakshminarayanan et al., 2017). Deep ensembles can be applied to IFWI framework if abundant observations are available. However, independently carrying out IFWI multiple times is expensive. Another alternative is to use a Bayesian neural network (i.e., BNN, Jospin et al., 2022; Kononenko, 1989) to represent subsurface parameters, in which weights and biases are randomly generated from a probability distribution trained by IFWI. Thus, instead of a deterministic implicit representation of subsurface parameters, the BNN-based IFWI can produce the posterior distribution of subsurface model with given datasets. By evaluating the trained BNN multiple times, we can obtain collective inversion results of subsurface models, which allows us to estimate uncertainty in predictions.

Figure 12. The variations of data discrepancies during FWI and IFWI implementations with noisy data. Top: losses using noisy data with $\sigma = 2\sigma_0$, bottom: losses using noisy data with $\sigma = 4\sigma_0$.
In order to concentrate the concept of IFWI and to keep this paper concise, the theory of BNN and BNN-based IFWI are not further discussed in this paper. Readers interested in BNN-based IFWI may refer to our preliminary results in elastic media (Zhang et al., 2022).

Besides the deep ensembles and BNN, the third, and simplest, way to gain Bayesian posterior distribution of predictions is to include dropout neurons in both training and validating procedures. Gal and Ghahramani (2016) demonstrate that training the neural network with dropouts activated is theoretically equivalent to training a large ensemble of networks, which approach to a Bayesian statistical model. Though dropout operation is initially introduced to prevent networks from overfitting as a regularization method during training process (Srivastava et al., 2014), and is usually deactivated for the deterministic prediction in validation, the predictive realizations can also be obtained by validating the trained neural network multiple times while activating dropouts (Gal & Ghahramani, 2016). Sun et al. (2021) firstly introduced the dropout method to evaluate the predictive uncertainty of seismic inversion using a physics-guided framework, which can also be extended to 3D inversion (Rusmanugroho et al., 2022). Taking advantage of its simplicity, we give the example of IFWI in predictive uncertainty analysis by easily adding the dropout neurons to the representation network.

In every training epoch, we randomly selected 80% of neurons of hidden layers using a Bernoulli distribution to form a new neural network, where the remaining 20% (i.e., dropout ratio \( p = 0.2 \)) of neurons are reset to zero. In validation, we first carry out the prediction using the trained network without activating dropout neurons, which are commonly considered the optimal generative performance of the network. The final prediction and its absolute misfits are plotted in Figures 13b and 13e, respectively. Compared to IFWI without dropouts (refer to as non-dropout IFWI, shown in Figure 9d) and Figure 13b indicates that training IFWI with dropouts may produce smoother inversion results, which is equivalent to adding certain forms of regularization into the non-dropout IFWI during optimization. However, we observe the presence of missing inversions in some areas, especially in the deep corners, which is likely caused by the difficulty of training networks with dropouts. This can usually be compensated for by either increasing the length of training time or increasing the complexity of the network.

To evaluate the predictive uncertainty of inversion results, we run 1,000 realizations using the trained network by IFWI with dropouts activated. The mean and standard deviation of 1,000 realizations are calculated and plotted in Figures 13c and 13d, respectively. The averaged inversion result shown in Figure 13c exhibits an analogous depiction of subsurface structures, but with moderate accuracy of velocity values, where the absolute misfits between the averaged result and the ground-truth are delineated in Figure 13f. Note that the measured uncertainty

**Figure 13.** The uncertainty measurements of IFWI with dropout ratio \( p = 0.2 \). (a) the ground-truth Marmousi model, (b) the inversion result without activated dropout in validation, (c) the average map of 1,000 realizations, (d) the standard deviation of 1,000 realizations, (e) the absolute error map between (a) and (b), (f) the absolute error map between (a) and (c).
map in Figure 13d is not equivalent to the absolute error map shown in Figure 13f, but shows a similar pattern. This is reasonable because the standard deviation represents for statistical biases of multiple realizations, where the absolute error map is calculated from a single realization.

3.4. Generalization

In order to examine the ability of IFWI to be applied in more general circumstances, we use a 2D slice extracted from the 3D Overthrust volume. The 2D Overthrust model is shown in Figure 14c with the grid cell of 20 m. 10 shot gathers are acquired with 800 m intervals at a depth of 20 m, and their relative locations are indicated by the golden stars in Figure 14c. A Ricker wavelet with 8 Hz dominant frequency is applied as the source and all shot gathers are sampled in 2 ms with a length of 1,500 samples. The free-surface at the top and a PML absorbing boundary condition at three other boundaries are applied during synthetic shot simulation. In this experiment, instead of using the global mean and standard deviation, we obtain a mean ($\mu = 4.412$ km/s) and a standard deviation ($\sigma = 1.116$ km/s) using a known well log indicated as the golden dashed line in Figure 14c.

The implementation of IFWI on the 2D Overthrust model is also performed with a random initial model, shown in Figure 14a, using the same hyperparameters for optimization. Figure 14b shows the final inversion result of the 2D Overthrust model using IFWI. Precise fault structures and thin layers of the 2D Overthrust model are correctly inverted, which further confirm a strong generalization ability of IFWI algorithm.

4. Discussion

We introduce the fundamentals of IFWI with implicit neural representations and how it iteratively predicts the accurate subsurface structure from a randomly initialized model without any prior information, with the help of the frequency-bias phenomena during deep learning optimization. In addition to the advantages of being insensitive to the initial model, our analysis confirms that IFWI also has a certain degree of robustness, and allows intuitive uncertainty analysis of the inversion results, all of which can be applied to more general cases. In spite of this, IFWI will likely benefit from further exploration.
Unlike the conventional geophysical inversion and imaging approaches, which are usually grid-based and most likely limited by the grid resolution, IFWI allows a coordinate-based and continuous representation of the subsurface model to be generated. This offers an opportunity to implement multiscale joint inversion, target-oriented inversion, scalable subsurface model inversion, and even mesh-free inversion if a mesh-free forward solver is adopted. Specifically, with a coordinate-based neural network, one can build a single DNR of subsurface models at different scales. For instance, it is possible to perform IFWI on a coarse grid and then reconstruct the subsurface model in a fine grid, or perform IFWI on several discrete models and then generate a continuous subsurface model. In addition, one can implement the inversion in a fully mesh-free and target-oriented manner by means of suitable mesh-free forward modeling algorithms, such as the PINN solver (Huang & Alkhalifah, 2022).

In real scenarios of geophysical data acquisition, the observation geometries are likely to be different, especially for data acquired by different institutions at different time periods. This makes it more difficult for these data to be used simultaneously, and largely limits their potentials. The continuity feature of subsurface model implicitly defined through IFWI can be adapted to any arbitrary observation settings, which may make the best use of these multi-modal geophysical data collected using different observation settings and greatly boost the multi-scale geophysical inversion. For example, by defining a neural representation based on the spherical coordinates to represent the Earth, it provides an opportunity to perform a global inversion of the Earth using seismograms collected by world-wide distributed seismometers and/or magnetotelluric data collected by different platforms at different locations, although some additional challenges may first have to be addressed. Furthermore, by incorporating the time coordinate into the representation network, the spatiotemporal evolution of the Earth's subsurface can be obtained with the constraints of data acquired at different time periods.

Many problems in understanding the solid Earth usually involve some degree of coupling between different physical fields, such as seismology, electromagnetism, geodesy, geomorphology, geology, petrology, and so on. One of the key to solving such multi-physics problems is to understand the interactions between parameters of different physical domains, which are often yet to be discovered. In the field of geophysical inversion, the joint inversion mechanism using complementary information contained in different types of geophysical measurements have been rapidly developed to reduce the risk of nonuniqueness and to improve the accuracy of results. However, in existing joint inversion approaches, parameters' relationships in different physical domains are usually depicted by empirical formulas or coupled generalization functions. With the framework of IFWI, we can construct a single neural network to represent physical parameters of interest in multiple domains, the multi-physics joint inversion can be performed by optimizing this single DNR with complementary constraints from different kinds of geophysical measurements, such as seismic, well-log (Jeong et al., 2020), gravity (Montesinos et al., 2022), electromagnetic (Blanco-Montenegro et al., 2008; Yao et al., 2022), ground penetrating radar (Meles et al., 2011), and remote sensing (Sun, Wauthier, et al., 2020). By analyzing this single DNR, further insight into the intrinsic relationships between parameters in different domains may be discovered.

However, there are challenges in undertaking such implementations of IFWI. For example, we observe that ReLU-based MLPs can easily guarantee the continuity of the represented implicit functions, but have difficulties in learning their high-frequency components. MLPs with periodic activation functions (Sitzmann et al., 2020) or Fourier features (Tancik et al., 2020) are able to learn high-frequency information quickly, but they are also prone to overfitting, which can destroy the continuity of the represented spaces. Thus, a better network architecture is needed to balance its representation capacities in terms of continuity and high-frequency. Another challenge of IFWI is its expensive computational cost. It is understandable that the implementation of IFWI is more computationally demanding than that of FWI (as shown in Figures 8 and 12), since IFWI starts from a random model. The proposed IFWI is built with PyTorch (Paszke et al., 2019) thus it can be easily accelerated by parallel computing and graphics processing units (GPUs). Besides, there are some shortcuts to reduce the computational cost of IFWI. For instance, pretraining the representation network with known prior information allows IFWI to achieve the same level of computational efficiency as FWI (see the comparison between FWI-Smooth and IFWI-Pretrain in Figure 8). Moreover, one can carry out IFWI with a coarse grid, which may significantly reduce the computational overhead of forward modeling, and then reconstruct the zone of interest in a fine grid using the trained DNR. As stated in subsection 2.3, the representation network in this paper is not determined to be optimal, thus it is possible to achieve higher computational efficiency with other architectures.
5. Conclusions

Geophysical inversion is a critical set of methods for estimating maps of subsurface structures, and estimating petrophysical properties. Amongst these, FWI is a key state-of-the-art technique. However, FWI tends to produce models associated with local minima, due to its strong non-linearity and is computationally demanding for uncertainty analysis. To address these issues, we propose the IFWI algorithm that produces a continuous and differentiable functional representation of subsurface parameters, instead of a grid-based solution. In contrast to the discrete parameterization, in which memory and precision are highly dependent on the grid solution, an implicit and continuous representation using DNNs can be much more efficient while preserving fine details. Both our theoretical and empirical analyses illustrate that IFWI is capable of automatic inversion from low-to-high-frequencies through the frequency bias of deep learning, which may significantly reduce the reliance on the initial model. This is further confirmed by numerical examples using the 2D Marmousi model. Our experimentation indicates that, given a random initial model, IFWI can gradually converge to the global minimum and produce an impressive representation of subsurface parameters with fine structures, while FWI falls into a local minimum. In addition, uncertainty analysis can be easily performed during IFWI optimization by approximating Bayesian inference with various deep learning approaches, such as dropout neurons, Bayesian neural networks, and deep ensembles. Moreover, the robustness and generalization ability of IFWI are also exemplified by adding different levels of noise and using a range of geological models, respectively.

Data Availability Statement

All FWI and IFWI implementations in this paper are carried out based on the theory-guided RNN framework, and all associated codes and data are available through the following link: https://doi.org/10.5281/zenodo.7262564.

Acknowledgments

This research was supported by the fellowship of China Postdoctoral Science Foundation (2022M722970), Shandong Provincial Nature Science Foundation, China (ZR2022QD036), and the Young-Talents-Project Start-up Foundation of Ocean University of China (300-00005422212018).

References

Alaudah, Y., Michalowicz, P., Alfarraj, M., & AIRegb, G. (2019). A machine-learning benchmark for facies classification. Interpretation, 7(3), SE175–SE187. https://doi.org/10.1190/interpret-2018-0249.1

Arror, S., Cohen, N., Golowich, N., & Hu, W. (2018). A convergence analysis of gradient descent for deep linear neural networks. arXiv preprint arXiv:1810.02281

Basri, R., Galun, M., Geifman, A., Jacobs, D., Kasten, Y., & Kritchman, S. (2020). Frequency bias in neural networks for input of non-uniform density. In International conference on machine learning (pp. 685–694).

Bhojanapalli, S., Neysabur, B., & Srebro, N. (2016). Global optimality of local search for low rank matrix recovery. Advances in Neural Information Processing Systems, 29.

Blanco-Montenegro, I., Nicolosi, I., Pignatelli, A., & Chiappini, M. (2008). Magnetic imaging of the feeding system of oceanic volcanic islands: El hierro (Canary Islands). Geophysical Journal International, 172(1), 539–535. https://doi.org/10.1111/j.1365-246X.2008.03723.x

Bozdağ, E., Trampert, J., & Tromp, J. (2011). Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements. Geophysical Journal International, 185(2), 845–870. https://doi.org/10.1111/j.1365-246X.2011.05970.x

Brutzkus, A., & Globerson, A. (2017). Globally optimal gradient descent for a ConvNet with Gaussian inputs. In International conference on machine learning (pp. 605–614).

Brutzkus, A., Globerson, A., Malach, E., & Shalev-Shwartz, S. (2017). Sgd learns over-parameterized networks that provably generalize on linearly separable data. arXiv preprint arXiv:1710.10174.

Byrd, R. H., Lu, P., Nocedal, J., & Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. SIAM Journal on Scientific Computing, 16(5), 1190–1208. https://doi.org/10.1137/0916069

Cano, E. V., Akram, J., & Peter, D. B. (2021). Automatic seismic phase picking based on unsupervised machine-learning classification and content information analysis. Geophysics, 86(4), V299–V315. https://doi.org/10.1190/geo2020-0308.1

Chen, Z., & Zhang, H. (2019). Learning implicit fields for generative shape modeling. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 5939–5948. https://doi.org/10.1109/cvpr.2019.00609

Chi, B., Dong, L., & Liu, Y. (2014). Full waveform inversion method using envelope objective function without low frequency data. Journal of Applied Geophysics, 109, 36–46. https://doi.org/10.1016/j.jappgeo.2014.07.010

Chi, B., Dong, L., & Liu, Y. (2015). Correlation-based reflection full-waveform inversion. Geophysics, 80(4), R189–R202. https://doi.org/10.1190/geo2014-0345.1

Choromanska, A., Henaff, M., Mathieu, M., Arous, G. B., & LeCun, Y. (2015). The loss surfaces of multilayer networks. Advances in Neural Information Processing Systems, 29.

Clement, F., Chavent, G., & Gómez, S. (2001). Migration-based traveltime waveform inversion of 2-d simple structures: A synthetic example. Geophysics, 66(3), 845–860. https://doi.org/10.1190/1.144974

Elasser, E., Gusikhin, L., Van Leeuwen, T., Azevkin, A. Y., & Herrmann, F. J. (2018). Total variation regularization strategies in full-waveform inversion. SIAM Journal on Imaging Sciences, 11(1), 376–406. https://doi.org/10.1137/17m111328x

Feng, R., Balling, N., Grana, D., Drasmich, J. S., & Hansen, T. M. (2021). Bayesian convolutional neural networks for seismic facies classification. IEEE Transactions on Geoscience and Remote Sensing, 59(10), 8933–8940. https://doi.org/10.1109/tgrs.2020.3049012

Gal, Y., & Ghahramani, Z. (2016). Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. International Conference on Machine Learning, 1050–1059.

Ge, R., Huang, F., Jin, C., & Yuan, Y. (2015). Escaping from saddle points—Online stochastic gradient for tensor decomposition. Conference on Learning Theory, 797–824.
Ge, R., Lee, J. D., & Ma, T. (2016). Matrix completion has no spurious local minimum. *Advances in Neural Information Processing Systems*, 29. Genova, K., Cole, F., Vlaic, D., Sarna, A., Freeman, W. T., & Funkhouser, T. (2019). Learning shape templates with structured implicit functions. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 7154–7164. https://doi.org/10.1109/ICCV.2019.00725

Guasch, L., Warner, M., & Ravaut, C. (2019). Adaptive waveform inversion: Practice. *Geophysics*, 84(3), R447–R461. https://doi.org/10.1190/geo2018-0377.1

He, Q., & Wang, Y. (2021). Reparameterized full-waveform inversion using deep neural networks. *Geophysics*, 86(1), V1–V13. https://doi.org/10.1190/geo2019-0382.1

Hu, Y., Jin, Y., Wu, X., & Chen, J. (2021). A theory-guided deep neural network for time domain electromagnetic simulation and inversion using a differentiable programming platform. *IEEE Transactions on Antennas and Propagation*, 70(1), 767–772. https://doi.org/10.1109/TAP.2021.3089585

Huang, X., & Alkalilah, T. (2022). Pinnup: Robust neural network wavefield solutions using frequency upscaling and neuron splitting. *Journal of Geophysical Research: Solid Earth*, 127(6). https://doi.org/10.1029/2021JB023703

Jeong, J., Park, E., Emelyanova, I., Pervukhina, M., Esteban, L., & Yun, S.-T. (2020). Interpreting the subsurface lithofacies at high lithological resolution by integrating information from well-log data and rock-core digital images. *Journal of Geophysical Research: Solid Earth*, 125(2), e2019JB018204. https://doi.org/10.1029/2019jb018204

Jospin, L. V., Laga, H., Bouraud, F., Buntine, W., & Bennamoun, M. (2022). Hands-on Bayesian neural networks—A tutorial for deep learning users. *IEEE Computational Intelligence Magazine*, 17(2), 29–48. https://doi.org/10.1109/MCI.2021.3155527

Klocek, S., Maziazi, L., Wolczyk, M., Tabor, J., Nowak, J., & Śmieja, M. (2019). Hypernetwork functional image representation. *International Conference on Artificial Neural Networks*, 896–910. https://doi.org/10.1007/978-3-030-30493-5_48

Kronenkonj, I. (1989). Bayesian neural networks. *Biological Cybernetics*, 61(5), 361–370. https://doi.org/10.1007/bf00200801

Lakshminarayanan, B., Pritzel, A., & Blundell, C. (2017). Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in Neural Information Processing Systems*, 30.

Li, Y., & Yuan, Y. (2017). Convergence analysis of two-layer neural networks with relu activation. *Advances in Neural Information Processing Systems*, 30.

Lin, Y., & Huang, L. (2014). Acoustic-and elastic-waveform inversion using a modified total-variation regularization scheme. *Geophysical Journal International*, 200(1), 489–502. https://doi.org/10.1093/gji/ggu393

Liu, M., Jervis, M., Li, W., & Nivlet, P. (2020). Seismic facies classification using supervised convolutional neural networks and semisupervised generative adversarial networks. *Geophysics, 85*(4), O47–O58. https://doi.org/10.1190/geo2019-0627.1

Liu, Q., & Tromp, J. (2006). Finite-frequency kernels based on adjoint methods. *Bulletin of the Seismological Society of America*, 96(6), 2383–2397. https://doi.org/10.1785/0120060041

Moles, G. A., Greenhalgh, S. A., Green, A. G., Maurer, H., & Van der Kruk, J. (2011). GPR full-waveform sensitivity and resolution analysis using an FDTD adjoint method. *IEEE Transactions on Geoscience and Remote Sensing*, 50(5), 1881–1896. https://doi.org/10.1109/igawi.2011.5963863

Mescheder, L., Oschsle, M., Niemeyer, M., Nowozin, S., & Geiger, A. (2019). Occupancy networks: Learning 3d reconstruction in function space. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4460–4470. https://doi.org/10.1109/CVPR.2019.00459

Métrier, L., Brossier, R., Virieux, J., & Operto, S. (2013). Full waveform inversion and the truncated Newton method. *SIAM Journal on Scientific Computing*, 35(2), B401–B437. https://doi.org/10.1137/120877854

Mildenhall, B., Srivastava, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., & Ng, R. (2020). NeRF: Representing scenes as neural radiance fields for view synthesis. In *European conference on computer vision* (pp. 405–421). https://doi.org/10.1007/978-3-030-58452-8_24

Montesinos, F., Aronso, J., Gómez-Ortiz, D., Benavent, M., Blanco-Montenegro, I., Vélez, E., et al. (2022). Imaging the volcanic structures beneath gran Canaria Island using new gravity data. *Journal of Geophysical Research: Solid Earth*, 127(11), e2022JB024863. https://doi.org/10.1029/2022JB024863

Nocedal, J., & Wright, S. J. (2006). Conjugate gradient methods. Numerical optimization (pp. 101–134).

Park, J. I., Florence, P., Straub, J., Newcombe, R., & Lovegrove, S. (2019). Deeppdf: Learning continuous signed distance functions for shape representation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 165–174. https://doi.org/10.1109/CVPR.2019.00025

Pastuze, A., Gross, S., Massa, F., Lerner, A., Bradbury, J., Chanan, G., et al. (2019). Pytorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems*, 32, 8024–8035. Retrieved from http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

Pliess, R.-E. (2006). A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophysical Journal International*, 167(2), 495–503. https://doi.org/10.1111/j.1365-246x.2006.02978.x

Rahaman, N., Baratin, A., Arpit, D., Draxler, F., Lin, M., Hamprecht, F., et al. (2019). On the spectral bias of neural networks. *International Conference on Machine Learning*, 5301–5310.

Rusmanugroho, H., Li, J., Davis Muhammed, M. D., & Sun, J. (2022). 3d velocity model building based upon hybrid neural network. *Second International Meeting for Applied Geoscience & Energy*, 1634–1638. https://doi.org/10.1109/IMAGE2022-3745433.1

Saad, O. M., & Chen, Y. (2020). Deep denoising autoencoder for seismic random noise attenuation. *Geophysics*, 85(4), V367–V376. https://doi.org/10.1190/geo2019-0468.1

Sitzmann, V., Martel, J., Bergman, A., Lindell, D., & Wetzstein, G. (2020). Implicit neural representations with periodic activation functions. *Advances in Neural Information Processing Systems*, 33, 7462–7473.

Sitzmann, V., Zollhöfer, M., & Wetzstein, G. (2019). Scene representation networks: Continuous 3d-structure-aware neural scene representations. *Advances in Neural Information Processing Systems*, 32.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(1), 1929–1958.

Sun, J., Innanen, K. A., & Huang, C. (2021). Physics-guided deep learning for seismic inversion with hybrid training and uncertainty analysis. *Geophysics*, 86(3), R303–R317. https://doi.org/10.1190/geo2020-0312.1

Sun, J., Niu, Z., Innanen, K. A., Li, J., & Trad, D. O. (2019). A theory-guided deep learning formulation of seismic waveform inversion. In *Seg technical program expanded abstracts 2019* (pp. 2343–2347). Society of Exploration Geophysicists. https://doi.org/10.1190/segam2019-3215762.1

Sun, J., Niu, Z., Innanen, K. A., Li, J., & Trad, D. O. (2020). A theory-guided deep-learning formulation and optimization of seismic waveform inversion. *Geophysics*, 85(2), R87–R89. https://doi.org/10.1190/geo2019-0338.1

Sun, J., Wauthier, C., Stephens, K., Gervais, M., Cervone, G., LaFemina, P., & Higgins, M. (2020). Automatic detection of volcanic surface deformation using deep learning. *Journal of Geophysical Research: Solid Earth*, 125(9), e2020JB019840. https://doi.org/10.1029/2020jb019840
Symes, W. W. (2008). Migration velocity analysis and waveform inversion. Geophysical Prospecting, 56(6), 765–790. https://doi.org/10.1111/j.1365-2478.2008.00698.x
Tancik, M., Srinivasan, P., Mildenhall, B., Fridovich-Keil, S., Raghavan, N., Singhal, U., et al. (2020). Fourier features let networks learn high frequency functions in low dimensional domains. Advances in Neural Information Processing Systems, 33, 7537–7547.
Tarantola, A. (1984). Inversion of seismic reflection data in the acoustic approximation. Geophysics, 49(8), 1259–1266. https://doi.org/10.1190/1.1441754
Tian, Y. (2017). An analytical formula of population gradient for two-layered relu network and its applications in convergence and critical point analysis. International Conference on Machine Learning, 3404–3413.
Virieux, J., & Operto, S. (2009). An overview of full-waveform inversion in exploration geophysics. Geophysics, 74(6), WCC1–WCC26. https://doi.org/10.1190/1.3238367
Wang, W., Yang, F., & Ma, J. (2018). Velocity model building with a modified fully convolutional network. In SEG technical program expanded abstracts (pp. 2086–2090). https://doi.org/10.1190/segam2018-2997566.1
Warner, M., & Guasch, L. (2016). Adaptive waveform inversion: Theory. Geophysics, 81(6), R429–R445. https://doi.org/10.1190/geo2015-0387.1
Wu, H., Zhang, B., Li, F., & Liu, N. (2019). Semiautomatic first-arrival picking of microseismic events by using the pixel-wise convolutional image segmentation method. Geophysics, 84(3), V143–V155. https://doi.org/10.1190/geo2018-0389.1
Wu, H., Zhang, B., Lin, T., Li, F., & Liu, N. (2019). White noise attenuation of seismic trace by integrating variational mode decomposition with convolutional neural network. Geophysics, 84(5), V307–V317. https://doi.org/10.1190/geo2018-0635.1
Wu, Y., Lin, Y., & Zhou, Z. (2018). InversionNet: Accurate and efficient seismic waveform inversion with convolutional neural networks. SEG Technical Program Expanded Abstracts, 2096–2100.
Wu, Y., & McMechan, G. A. (2019). Parametric convolutional neural network-domain full-waveform inversion. Geophysics, 84(6), R881–R896. https://doi.org/10.1190/geo2018-0224.1
Xie, H., Yao, H., Sun, X., Zhou, S., & Zhang, S. (2019). Pix2vox: Context-aware 3d reconstruction from single and multi-view images. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2690–2698. https://doi.org/10.1109/iccv.2019.00278
Xu, S., Wang, D., Chen, F., Zhang, Y., & Lambare, G. (2012). Full waveform inversion for reflected seismic data. 74th EAGE Conference and Exhibition Incorporating Europec 2012, 293. https://doi.org/10.3997/2214-4609.20148725
Yang, F., & Ma, J. (2019). Deep-learning inversion: A next-generation seismic velocity model building method. Geophysics, 84(4), R583–R599. https://doi.org/10.1190/geo2018-0249.1
Yang, L., Chen, W., Wang, H., & Chen, Y. (2021). Deep learning seismic random noise attenuation via improved residual convolutional neural network. IEEE Transactions on Geoscience and Remote Sensing, 59(9), 7968–7981. https://doi.org/10.1109/tgrs.2021.3053399
Yao, H., Ren, Z., Tang, J., & Zhang, K. (2022). A multi-resolution finite-element approach for global electromagnetic induction modeling with application to southeast China coastal geomagnetic observatory studies. Journal of Geophysical Research: Solid Earth, 127(8), e2022JB024659. https://doi.org/10.1029/2022jb024659
Yedlin, M. J., & Van Vorst, D. (2010). Tutorial on the continuous and discrete adjoint state method and basic implementation. CREWES Research Report, 22, 1–20.
Yuan, P., Wang, S., Hu, W., Wu, X., Chen, J., & Van Nguyen, H. (2020). A robust first-arrival picking workflow using convolutional and recurrent neural networks. Geophysics, 85(5), U109–U119. https://doi.org/10.1190/geophysics.2020-0437.1
Zhang, T., Innenan, K. A., Sun, J., & Trad, D. O. (2020). Numerical analysis of a deep learning formulation of multi-parameter elastic full waveform inversion. SEG Technical Program Expanded Abstracts, 1531–1535. https://doi.org/10.1190/segam2020-3426826.1
Zhang, T., Sun, J., Trad, D. O., & Innenan, K. A. (2022). Multilayer perceptron and Bayesian neural network based implicit elastic full-waveform inversion. Second International Meeting for Applied Geoscience & Energy, 1639–1643. https://doi.org/10.1109/image2022-3746334.1
Zhu, W., Xu, K., Darve, E., Biondi, B., & Beroza, G. C. (2022). Integrating deep neural networks with full-waveform inversion: Reparameterization, regularization, and uncertainty quantification. Geophysics, 87(1), R93–R109. https://doi.org/10.1190/geophysics-2020-0933.1