Seismic Wave Propagation and Inversion with Neural Operators

Yan Yang[*†1], Angela F. Gao[†2], Jorge C. Castellanos[†1], Zachary E. Ross[1], Kamyar Azizzadenesheli[3], and Robert W. Clayton[1]

Abstract

Seismic wave propagation forms the basis for most aspects of seismological research, yet solving the wave equation is a major computational burden that inhibits the progress of research. This is exacerbated by the fact that new simulations must be performed whenever the velocity structure or source location is perturbed. Here, we explore a prototype framework for learning general solutions using a recently developed machine learning paradigm called neural operator. A trained neural operator can compute a solution in negligible time for any velocity structure or source location. We develop a scheme to train neural operators on an ensemble of simulations performed with random velocity models and source locations. As neural operators are grid free, it is possible to evaluate solutions on higher resolution velocity models than trained on, providing additional computational efficiency. We illustrate the method with the 2D acoustic wave equation and demonstrate the method's applicability to seismic tomography, using reverse-mode automatic differentiation to compute gradients of the wavefield with respect to the velocity structure. The developed procedure is nearly an order of magnitude faster than using conventional numerical methods for full waveform inversion.

Introduction

The simulation of seismic wave propagation through Earth's interior underlies most aspects of seismological research, from the simulation of strong ground shaking due to large earthquakes (Graves and Pitarka, 2016; Rodgers *et al.*, 2019), to imaging the subsurface velocity structure (Fichtner *et al.*, 2009; Tape *et al.*, 2009; Virieux and Operto, 2009; Lee *et al.*, 2014; Gebraad *et al.*, 2020), to derivation of earthquake source properties (Duputel *et al.*, 2015; Ye *et al.*, 2016; Wang and Zhan, 2020). The compute costs associated with these wavefield simulations are substantial; and, for reasons of computational efficiency, 1D models are often used, even when better 3D velocity models are available. As a result, seismic wave simulations are often the limiting factor in the pace of geophysical research.

Recently, deep learning approaches have been explored with the goal of solving various geophysical partial differential equations (Moseley, Markham, and Nissen-Meyer, 2020; Moseley, Nissen-Meyer, and Markham, 2020; Smith *et al.*, 2020; Moseley *et al.*, 2021). Beyond the goal of accelerating compute capabilities, such physics-informed neural networks may offer other advantages such as grid independence, low-memory overhead, differentiability, and on-demand solutions. These properties facilitate deep learning being used to solve geophysical inverse problems (Zhu *et al.*, 2020; Smith *et al.*, 2021; Xiao *et al.*, 2021; Zhang and Gao, 2021), as a wider

1. Seismological Laboratory, California Institute of Technology, Pasadena, California, U.S.A., https://orcid.org/0000-0002-6105-2918 (YY); https://orcid.org/0000-0002-0103-6430 (JCC); https://orcid.org/0000-0002-6343-8400 (ZER); https://orcid.org/0000-0003-3323-3508 (RWC); 2. Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, California, U.S.A., https://orcid.org/0000-0001-8574-8728 (AFG); 3. Department of Computer Science, Purdue University, West Lafayette, Indiana, U.S.A., https://orcid.org/0000-0001-8507-1868 (KA)

*Corresponding author: yanyang@caltech.edu

†These authors contributed equally to this work.

selection of algorithms and frameworks then are available for use, such as approximate Bayesian inference techniques like variational inference.

One of the major challenges associated with wave propagation is that a new simulation must be performed whenever the properties of the source or velocity structure are perturbed in some way. This alone substantially increases the necessary compute costs, making some problems prohibitively expensive even if they are mathematically or physically tractable. For the most part, these limitations have been accepted as an inevitable part of seismology, but now physics-informed machine learning approaches have started to offer some pathways for moving beyond this issue. For example, Smith et al. (2020) use a deep neural network to solve the Eikonal equation for any source–receiver pair by taking these locations as input. Then, this can be exploited for hypocenter inversion by allowing for gradients of the travel time field to be computed with respect to the source location (Smith et al., 2021). However, these models are relatively inefficient to train and even then are only able to learn approximate solution operators to the differential equations.

The aforementioned limitations may seem surprising but result from a basic attribute of neural networks that in fact makes them ill-suited for solving differential equations. Specifically, neural networks are designed for learning maps between two finite-dimensional spaces, whereas learning a general solution operator for a differential equation requires the ability to map between two infinite dimensional spaces (i.e., function spaces). A paradigm for learning maps between function spaces was recently developed (Li et al., 2020a,b, 2021) and has been termed as neural operator. The general idea behind these models is that they have shared parameters over all possible functions describing the initial conditions, which allows them to operate on functions, even when the inputs are a numerically discretized representation of them.

Here, we explore the potential of neural operators in improving seismic wave propagation and inversion. We develop a prototype framework for training neural operators on the 2D acoustic wave equation and show that this approach provides a suite of tantalizing new advantages over conventional numerical methods for seismic wave propagation. This study provides a proof of concept of this technology and its application to seismology.

Preliminaries

For a given function A and a Green's function G, let U denotes the solution to a linear partial differential equation (PDE), that is, the solution operator,

$$U(x) = (\mathcal{L}A)(x) = \int G(x, y)A(y)dy, \tag{1}$$

in which $\mathcal{L}$ is the corresponding linear operator. For example, suppose that the PDE to be solved is the acoustic wave equation; and then A could describe the velocity structure as well as the initial conditions. Neural operator generalizes this formulation to the nonlinear setting in which a set of linear operators are sequentially applied to construct a general nonlinear solution operator. In its basic form, an ℓ-layered neural operator is constructed as follows:

$$U(x) = \mathcal{L}_\ell(\sigma(\mathcal{L}_{\ell-1}...\sigma(\mathcal{L}_1 V)...))(x), \tag{2}$$

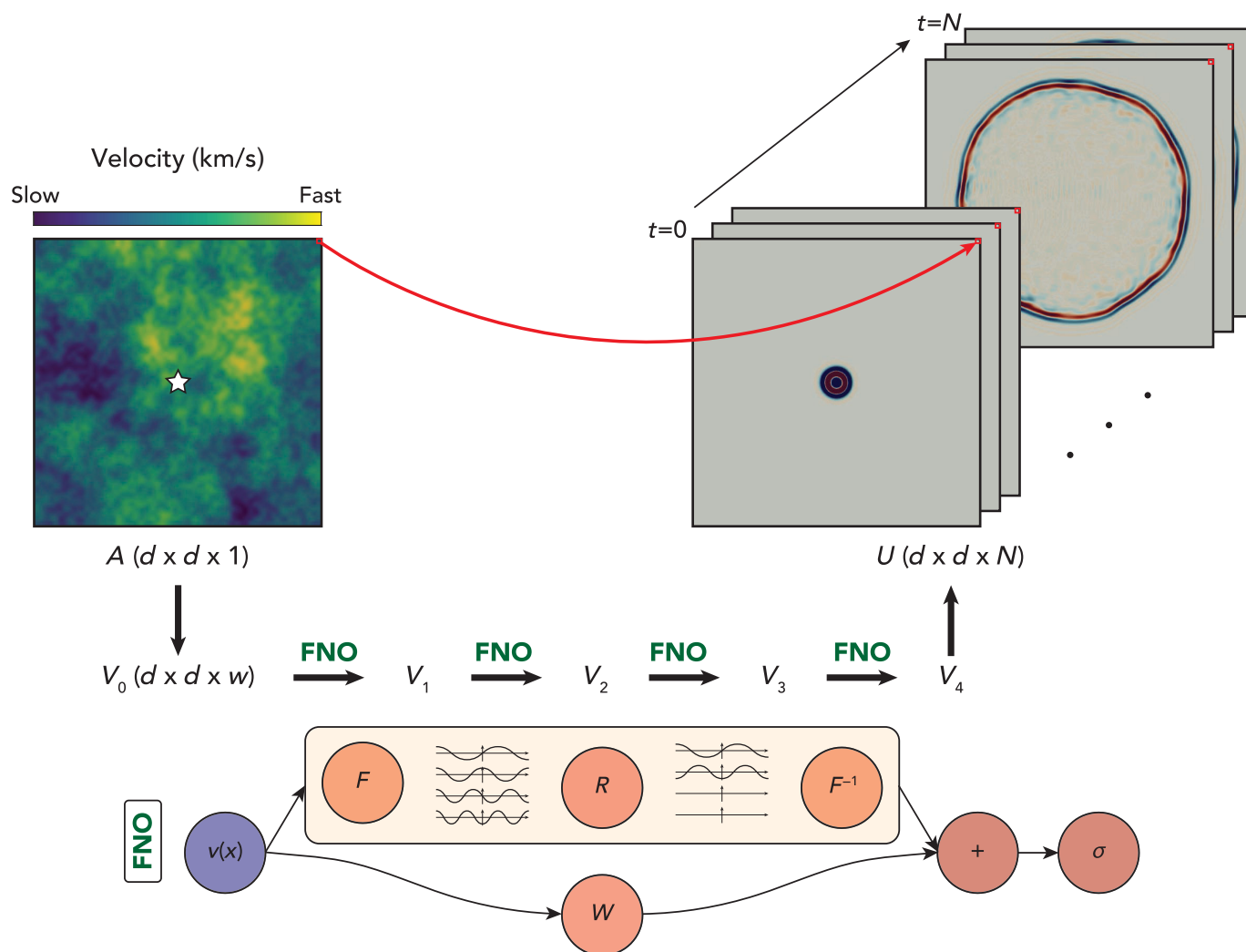in which $\mathcal{L}_i$ is such that for any function V, we have,

$$(\mathcal{L}_i V)(x) = W_i(x) + \int K_i(x, y)V(y)dy. \tag{3}$$

Under this framework, $W_i(x)$ and $K_i(x, y)$ constitute the learnable components of the neural operator and allow for a solution to be produced for any prescribed function A. In a limited sense, neural operators can be viewed as generalized Green's functions.

Methods

We designed a framework that applies neural operators to the 2D acoustic wave equation. The basic idea for this procedure is outlined schematically in Figure 1. A specific type of neural operator called a Fourier neural operator (FNO; Li et al., 2021) receives a velocity model specified on an arbitrary, possibly irregular mesh, along with the coordinates of a point source. One of the main features of FNO is that the major calculations are performed in the frequency domain that allows the convolutions in equation (3) to be rapidly computed. The output of the FNO is the complete wavefield solution, which can be queried anywhere within the medium, regardless of whether the points lie on the input mesh.

The most basic component of the FNO is a Fourier block (Fig. 1), which first transforms an input function (V) to the Fourier domain. In the first layer of the network, V is equal to the initial conditions A. Then, a kernel (K_i) is computed specifically for this function and is truncated at low order, before performing the integration via multiplication. Finally, the result is transformed back and a nonlinear activation function is applied, which concludes the Fourier block. For this study, the architecture of the FNO contains four sequential Fourier blocks and applies a Rectified Linear Unit (ReLU)

Figure 1. Our approach applying Fourier neural operator (FNO) to the 2D acoustic wave equation. The inputs to the FNO model are the velocity model with dimensions $d \times d \times 1$ and the source location, indicated by the white star. The input velocity model is lifted to a higher dimensional space with size $d \times d \times w$ using a neural network. Then, we apply four Fourier operator layers and finally project back to the target wavefield dimensions of $d \times d \times N$ using a neural network. The bottom panel shows details of the Fourier layer architecture; we define v to be the input. On top: We apply a Fourier transform F to v and then apply a linear transformation R to the lower Fourier modes, filtering out higher modes. Then, we apply an inverse Fourier transform F^{-1}. On the bottom, we apply a local linear transform W to v.

activation to the output of each (Fig. 1). We note that the truncation of the Fourier modes is performed on the function values after lifting them to a higher dimensional space, rather than the raw input function, so that this does not lead to compression. We refer the interested readers to Li *et al.* (2021) for more mathematical details about the FNO.

We constructed a training dataset of simulations to learn from by first generating random velocity models. We set up a 64×64 grid with 0.16 km spacing to define the velocity model. Then, we created 5000 random velocity models by sampling random fields having a von Karman covariance function with the following parameters: Hurst exponent $\kappa = 0.5$, correlation length $\mathbf{a} = [a_x, a_y] = [1.6 \text{ km}, 1.6 \text{ km}]$, and $\varepsilon = 0.1$, $\mu = 3 \text{ km/s}$, and $\sigma = 0.15 \text{ km/s}$. Then, for each of these velocity models, we apply a Ricker wavelet source at a random point and solve the acoustic wave equation using a spectral

element method (SEM; Afanasiev *et al.*, 2019). It should be noted that there is a source grid used, because this is a requirement of the SEM. Because Gaussian random fields (GRFs) can represent all continuous functions, the purpose of these steps is

to create a suite of simulations that span a range of possible conditions. We show later that they can even well approximate strongly discontinuous velocity models. An example velocity model and simulation is shown in Figure 1. Applying the aforementioned procedure results in a training dataset of 5000 data samples, each of which is a different simulation.

Given the simulation dataset, we can proceed to train an FNO model in a supervised capacity, in which the goal is to learn a model that can reliably output a solution to the wave equation for arbitrary input conditions. The training is performed using batch gradient descent, for which the parameters of the FNO are updated to minimize the error against the "true" spectral element solutions. A mean-square error loss function is used. We use a batch size of 30 simulations and train the model for a maximum of 300 epochs. We use all but 200 of the simulations for training and set aside the remainder for cross validation of the model. The time required to train the model from scratch is approximately 18 hr using six NVIDIA Tesla V100 graphical processing units (GPUs).

Experiments
Initial wavefield demonstration
Figure 2 shows two example wavefields corresponding to two different velocity models, each with a different source. The spectral element solution is shown alongside the wavefield predicted by the FNO for the eight different receivers (blue triangles). For these examples, the input velocity model is 64×64. The relative ℓ_2 loss of the FNO wavefields are 0.180 and 0.363. These examples are representative of the entire validation dataset, which has a loss of 0.273 relative to the spectral element solutions.

The number of simulations needed for training
Once fully trained, the FNO can evaluate a new solution in a fraction of a second, and thus the time to train the FNO will be the vast majority of the needed compute time. A primary concern about the computational demands of the FNO approach is, therefore, the number of simulations needed for training. Here, we examine how the number of training simulations influences the accuracy of the solution. We create a series of tests in which the number of training simulations is varied from 1200 at the fewest to 4800 at the most. The results are shown in Figure 3, in which we show the FNO wavefield predictions for each dataset. Even with 1200 training samples, there is no indication that there is overfitting, because the training waveform error is similar across different models (Fig. 3a,b). Training using just 1200 simulations is able to

predict the major arrival. Increasing number of training samples provides a better fit of the reflections (e.g., 3.2 s in Fig. 3c).

Generalization to arbitrary velocity models
The FNO was trained only on velocity models drawn from Gaussian random fields; and, although this family of functions is broad, it does not include some types of functions that exist in the Earth, such as discontinuous functions. This raises the question of whether the FNO can still generalize well to these cases. Figure 4a–c shows an example of a predicted wavefield for a velocity model containing a constant velocity square embedded within a homogeneous medium. Although the velocity model itself is rather simple, it is actually very far removed from the characteristics of the random fields that the FNO was trained on and represents a challenging example. We can see that the predicted wavefield does a very good job of approximating the wavefield compared to the ground truth. We believe that the small residual errors can be reduced with better hyperparameter selection.

Generalization to higher resolution grids
FNO can be viewed in some sense as a method for learning generalized Green's functions valid for arbitrary boundary conditions. Because it is intrinsically learning a mapping between function spaces, the FNO is theoretically independent of the resolution at which the functions are discretized (this is only a requirement for evaluation on a computer). One important advantage of this is that the FNO can be trained on velocity models with a certain grid spacing and then be evaluated on velocity models with a different grid spacing at inference time. Here, we are not simply talking about interpolating the wavefield after solving the PDE; but, rather, the solutions to the PDE can actually be evaluated on a higher resolution velocity model with negligible extra compute cost. To demonstrate this, Figure 4d–f shows the FNO prediction for a random velocity field with 2× higher resolution (128×128) than the models used during training, alongside the spectral element solution. The FNO solution closely approximates the spectral element solution. The velocity models with different meshes have the same roughness as the training data set. Resolving more rough structure with denser spacing can be achieved by training with many more GRFs with varying correlation length scales and variance.

Full waveform inversion with neural operators
One of the most useful applications of wavefield simulations is in inversion, to image the Earth's interior. The adjoint-state
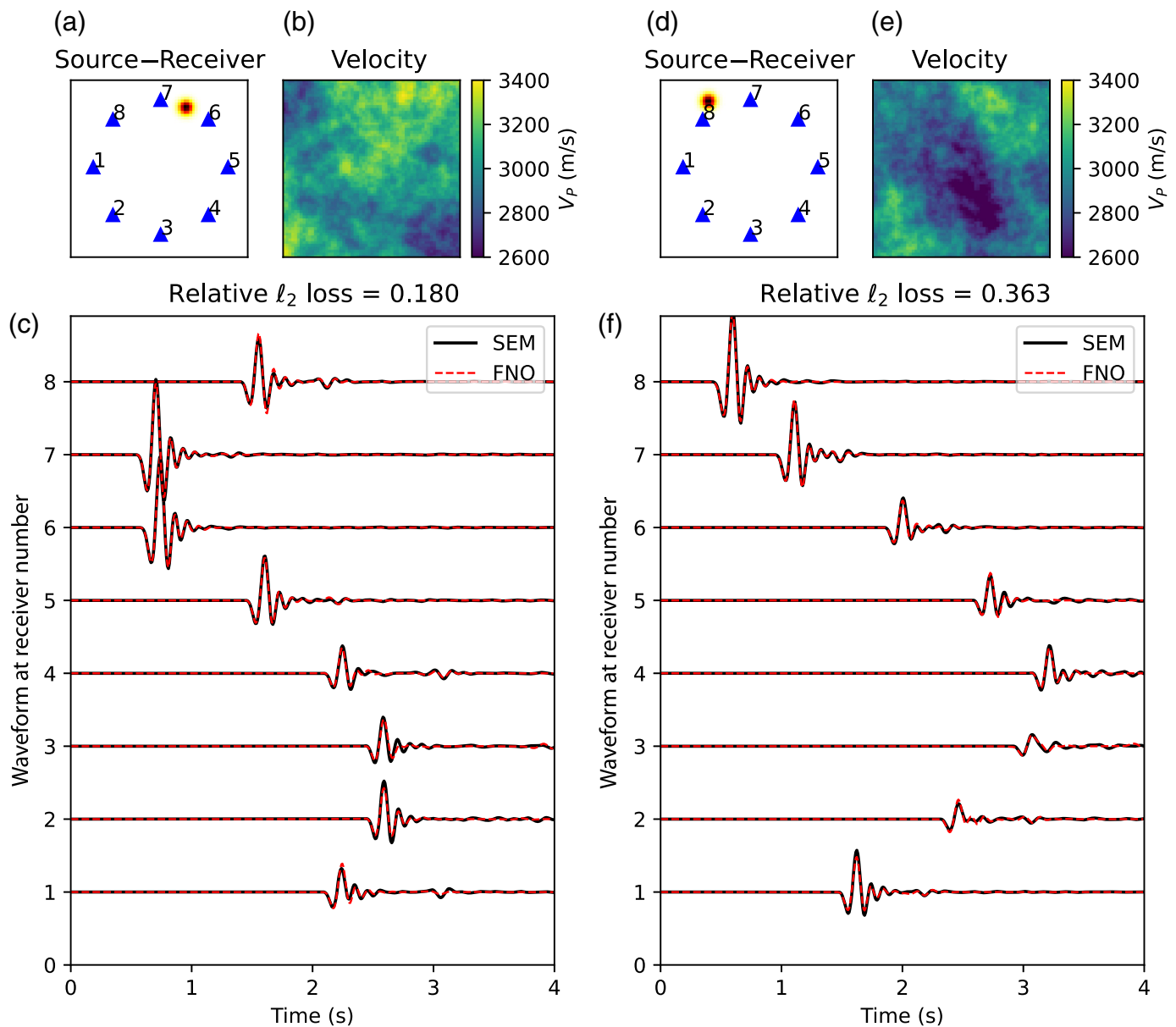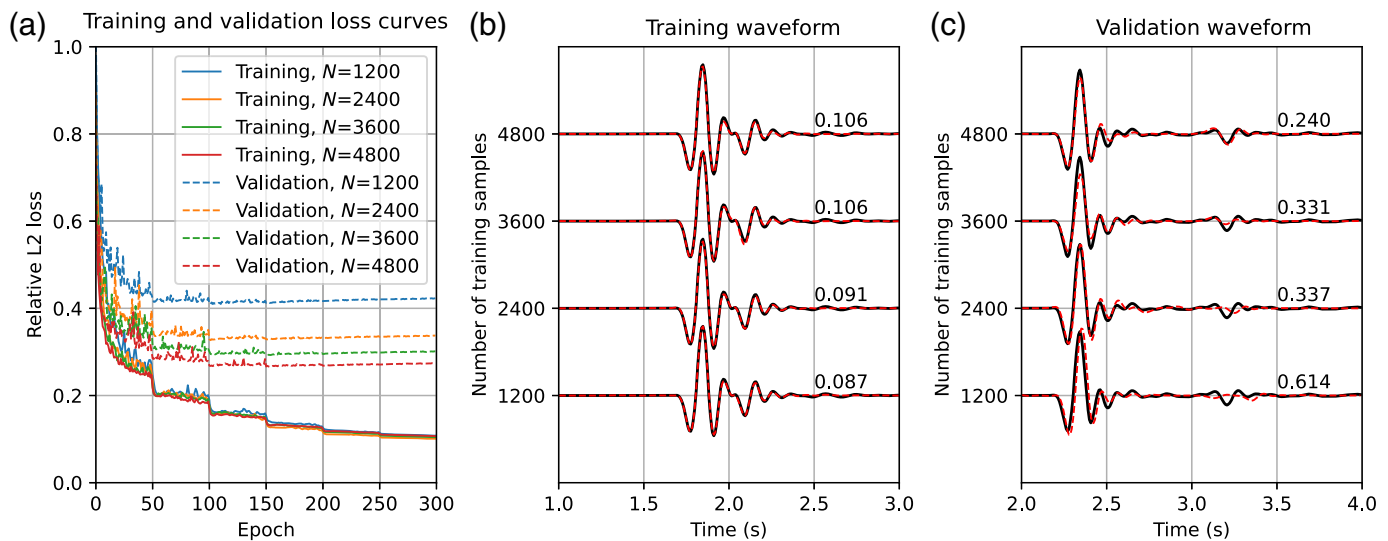
Figure 2. Examples of two validation wavefield simulations from the trained FNO model. (a) The source–receiver locations with receivers in blue and source in red, (b) the velocity structure, (c) waveforms simulated with spectral element method (SEM; black) and FNO (red), and (d–f) same as (a–c), but for a different velocity model. The relative ℓ_2 loss of the two examples are 0.180 and 0.363, respectively, which are representative of the entire validation data set with an average ℓ_2 loss of 0.273. We demonstrate that the FNO simulation results are able to capture both the major arrivals as well as some reflections.

method is an approach to efficiently compute the gradients of an objective function with respect to parameters of interest and can be used for seismic tomography (e.g., Tape *et al.*, 2009; Gebraad *et al.*, 2020). Neural operators are differentiable by design, which enables gradient computation with reverse-mode automatic differentiation. Automatic differentiation has been shown to be mathematically equivalent to the adjoint-state method (Zhu *et al.*, 2021). This allows for the gradients of the wavefield to be determined with respect to the inputs (velocity model and source location).

Figure 5 demonstrates our full waveform inversion (FWI) performance. For each case, we compute synthetic observations

using the source distribution as shown (red circles), taking every point in the 64 × 64 grid as a receiver. The observations are noise-free for this experiment. Then, we perform tomography

Figure 3. Model performance as a function of the number of training samples. (a) Training and validation loss curves as a function of different numbers of training samples. (b) Example waveform fitting of a single training example from models trained with varying number of training examples. (c) Example waveform fitting of a single validation example from models trained with varying number of training examples. The numbers to the right of each waveform shows the relative ℓ_2 misfit. This shows that the model trained on 4800 samples is able to capture the reflections, whereas the model trained on smaller number of samples does not generalize to reflections in the validation example.

by starting with a homogeneous initial velocity model and forward propagating a wavefield with the FNO for each source. We calculate the loss $L = \sum_i \sum_j [u^{\mathrm{obs}}(x_i, x_j) - u^{\mathrm{pred}}(x_i, x_j)]^2$ and compute ∇L with automatic differentiation. The velocity model is then iteratively updated with gradient descent for 1000 iterations using the Adam optimizer (Kingma and Ba, 2014) and a learning rate of 0.01. For comparison, Figure 5a,b shows the imaging result using SEM and adjoint-state method, with a relative ℓ_2 misfit between the inverted and true velocity model of 0.0289. Figure 5c,d shows the result for the same velocity structure using FNO and automatic differentiation, with a misfit of 0.0319. Figure 5e,f is designed to demonstrate sharp discontinuous changes with a short wavelength. The results demonstrate the remarkable capabilities of FNO to learn a general solution operator.

We note that our FWI approach neither requires an adjoint wavefield to be computed nor a cross correlation; the gradients can be rapidly computed with GPUs using automatic differentiation. The rapid simulation makes it substantially more efficient than adjoint methods. For these experiment, 20 sources take ~1 s for one tomographic iteration, including the costs of computing the forward model, whereas the spectral element method with adjoint methods takes ~100 s for one tomographic iteration. These time measurements are from using only a single NVIDIA Tesla V100 GPU.

Discussion

This study presents a prototype framework for applying neural operators to the 2D acoustic wave equation. We anticipate that the general framework would also be suitable for the 3D elastic-wave equation with relatively little modification. Indeed, the FNO method was applied successfully to the Navier–Stokes equations (Li *et al.*, 2021), which can be more challenging to solve than the elastic-wave equation. In our tests, we found that only a few thousand simulations were needed to train an FNO model and, from there, required negligible time to compute a new solution. Because FNO can be trained on lower resolution simulations and then generalize to higher resolution solutions once trained, this results in substantially faster computations than using traditional numerical methods at the full resolution.

One of the limitations of the approach is that the solutions are approximate, as seen in several of the figures. However, because this is a learning-based approach, the performance can be improved in the future using a better model architecture, thorough tuning of hyperparameters, improving the size of the training dataset, using a more appropriate objective function, and various other factors. In addition, as new developments within machine learning emerge in this area, they would be able to be incorporated. Thus, these performance
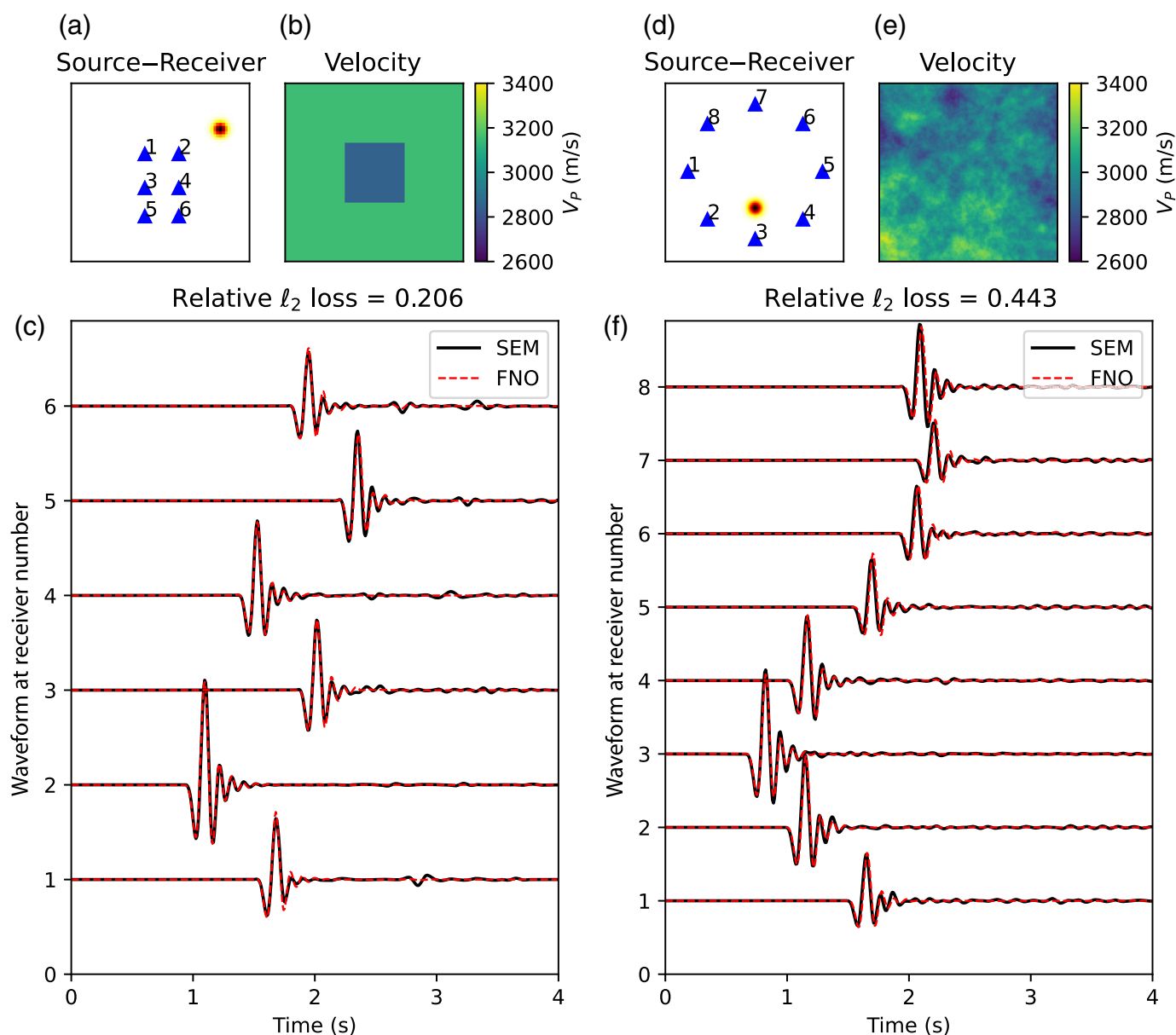
Figure 4. Model generalization experiments. (a) The source–receiver locations with receivers in blue and source in red, (b) a velocity model with a homogeneous background of 3 km/s and a 5% square anomaly, (c) waveform simulated with SEM (black) and FNO (red), and (d–f) same as (a–c), but for an input velocity model with 2× finer resolution than trained on. These experiments show that the model is not just memorizing the solutions, but is able to generalize to entirely new conditions.

metrics should only be viewed as a starting point. For some applications, the error may be enough of an issue, and traditional numerical methods may be preferable; however, for many other situations in geophysics, a reasonably accurate solution may be acceptable.

Among the most exciting benefits of our approach is that by training the FNO on random velocity models, the FNO is able to produce solutions for arbitrary velocity models. This is because FNO learns a general solution operator to the PDE and not specifically the velocity model. This means that the model does not need to be retrained for each region. Thus, the approach offers the potential for a single FNO model to be used by the entire seismology community for any region of a similar size. Although the initial cost of training an FNO and performing the training simulations may be expensive, it only needs to be done a single time for the community as a whole.
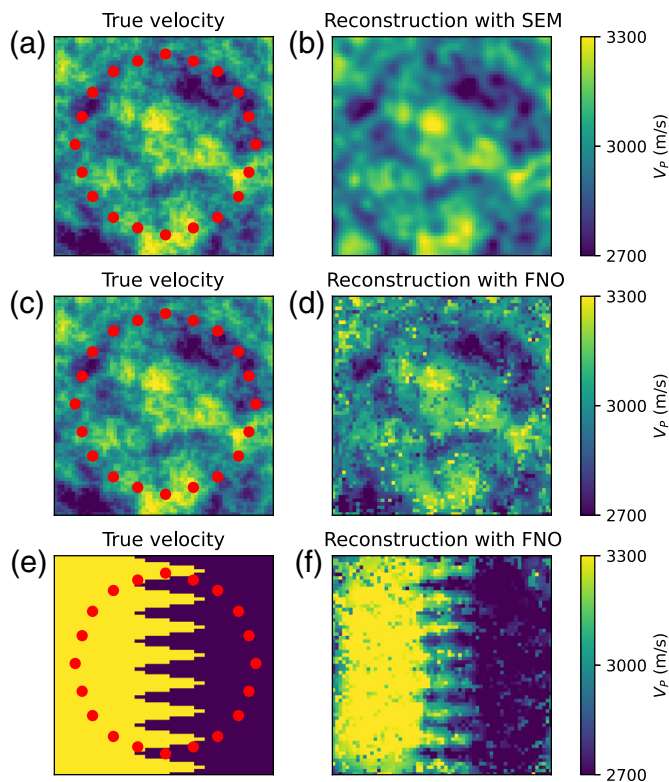
Figure 5. Example of a full waveform inversion using FNO. (a,c,e) True velocity models with source locations indicated by red circles and receivers placed at every node of the 64 × 64 grid (10 km × 10 km region). (b) Reconstruction using SEM and adjoint tomography. (d,f) Reconstruction using FNO as the forward model and automatic differentiation to compute gradients. No regularization is used for these experiments.

Data and Resources

All the data presented in this study are synthetic and available upon request. The supplemental material for this article demonstrates that the misfit between the simulations using spectral element method (SEM) and the Fourier neural operator (FNO) is minimal.

Declaration of Competing Interests

The authors declare that there are no competing interests.

References

Afanasiev, M., C. Boehm, M. van Driel, L. Krischer, M. Rietmann, D. A. May, M. G. Knepley, and A. Fichtner (2019). Modular and flexible spectral-element waveform modelling in two and three dimensions, *Geophys. J. Int.* **216,** no. 3, 1675–1692, ISSN 0956-540X, doi: 10.1093/gji/ggy469.

Duputel, Z., J. Jiang, R. Jolivet, M. Simons, L. Rivera, J.-P. Ampuero, B. Riel, S. E. Owen, A. W. Moore, S. V. Samsonov, *et al.* (2015). The Iquique earthquake sequence of April 2014: Bayesian modeling accounting for prediction uncertainty, *Geophys. Res. Lett.* **42,** no. 19, 7949–7957, ISSN 1944-8007, doi: 10.1002/2015GL065402.

Fichtner, A., B. L. N. Kennett, H. Igel, and H.-P. Bunge (2009). Full seismic waveform tomography for upper-mantle structure in the Australasian region using adjoint methods, *Geophys. J. Int.* **179,** no. 3, 1703–1725, ISSN 0956-540X, doi: 10.1111/j.1365-246X.2009.04368.x.

Gebraad, L., C. Boehm, and A. Fichtner (2020). Bayesian elastic full-waveform inversion using Hamiltonian Monte Carlo, *J. Geophys. Res.* **125,** no. 3, e2019JB018428, ISSN 2169-9356, doi: 10.1029/2019JB018428.

Graves, R., and A. Pitarka (2016). Kinematic ground-motion simulations on Rough Faults including effects of 3D stochastic velocity perturbations, *Bull. Seismol. Soc. Am.* **106,** no. 5, 2136–2153, GeoScienceWorld, ISSN 0037-1106, doi: 10.1785/0120160088.

Kingma, D. P., and J. Ba (2014). Adam: A method for stochastic otimization, available at http://arxiv.org/abs/1412.6980 (last accessed August 2021).

Lee, E.-J., P. Chen, T. H. Jordan, P. B. Maechling, M. A. M. Denolle, and G. C. Beroza (2014). Full-3-D tomography for crustal structure in Southern California based on the scattering-integral and the adjoint-wavefield methods, *J. Geophys. Res.* **119,** no. 8, 6421–6451, ISSN 2169-9356, doi: 10.1002/2014JB011346.

Li, Z., N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, and A. Anandkumar (2020a). Multipole graph neural operator for parametric partial differential equations, available at http://arxiv.org/abs/2006.09535 (last accessed August 2021).

Li, Z., N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, and A. Anandkumar (2020b). Neural operator: Graph kernel network for partial differential equations, available at http://arxiv.org/abs/2003.03485 (last accessed August 2021).

Li, Z., N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, and A. Anandkumar (2021). Fourier neural operator for parametric partial differential equations, available at http://arxiv.org/abs/2010.08895 (last accessed August 2021).

Moseley, B., A. Markham, and T. Nissen-Meyer (2020). Solving the wave equation with physics-informed deep learning, available at http://arxiv.org/abs/2006.11894 (last accessed August 2021).

Moseley, B., A. Markham, and T. Nissen-Meyer (2021). Finite Basis Physics-Informed Neural Networks (FBPINNs): A scalable domain decomposition approach for solving differential equations, available at http://arxiv.org/abs/2107.07871 (last accessed August 2021).

Moseley, B., T. Nissen-Meyer, and A. Markham (2020). Deep learning for fast simulation of seismic waves in complex media, *Solid Earth* **11,** no. 4, 1527–1549, ISSN 1869-9510, Copernicus GmbH, doi: 10.5194/se-11-1527-2020.

Rodgers, A. J., N. Anders Petersson, A. Pitarka, D. B. McCallen, B. Sjogreen, and N. Abrahamson (2019). Broadband (0–5 Hz)

fully deterministic 3D ground-motion simulations of a magnitude 7.0 Hayward fault earthquake: Comparison with empirical ground-motion models and 3D path and site effects from source normalized intensities, *Seismol. Res. Lett.* **90,** no. 3, 1268–1284, GeoScienceWorld, ISSN 0895-0695, doi: 10.1785/0220180261.

Smith, J. D., K. Azizzadenesheli, and Z. E. Ross (2020). EikoNet: Solving the Eikonal equation with deep neural networks, *IEEE Trans. Geosci. Remote Sens.* 1–12, ISSN 1558-0644, doi: 10.1109/TGRS.2020.3039165.

Smith, J. D., Z. E. Ross, K. Azizzadenesheli, and J. B. Muir (2021). HypoSVI: Hypocenter inversion with Stein variational inference and Physics Informed Neural Networks, available at http://arxiv.org/abs/2101.03271 (last accessed August 2021).

Tape, C., Q. Liu, A. Maggi, and J. Tromp (2009). Adjoint tomography of the Southern California Crust, *Science* **325,** no. 5943, 988–992, ISSN 0036-8075, 1095-9203, doi: 10.1126/science.1175298.

Virieux, J., and S. Operto (2009). An overview of full-waveform inversion in exploration geophysics, *Geophysics* **74,** no. 6, WCC1–WCC26, Society of Exploration Geophysicists, ISSN 0016-8033, doi: 10.1190/1.3238367.

Wang, X., and Z. Zhan (2020). Moving from 1-D to 3-D velocity model: Automated waveform-based earthquake moment tensor inversion in the Los Angeles region, *Geophys. J. Int.* **220,** no. 1, 218–234, ISSN 0956-540X, doi: 10.1093/gji/ggz435.

Xiao, C., Y. Deng, and G. Wang (2021). Deep-learning-based adjoint state method: Methodology and preliminary application to inverse modeling, *Water Resour. Res.* **57,** no. 2, e2020WR027400, ISSN 1944-7973, doi: 10.1029/2020WR027400.

Ye, L., T. Lay, H. Kanamori, and L. Rivera (2016). Rupture characteristics of major and great mw ≥ 7.0 megathrust earthquakes from 1990 to 2015: 2. Depth dependence, *J. Geophys. Res.* **121,** no. 2, 2015JB012427, ISSN 2169-9356, doi: 10.1002/2015JB012427.

Zhang, W., and J. Gao (2021). Deep-learning full-waveform inversion using seismic migration images, *IEEE Trans. Geosci. Remote Sens.* 1–18, ISSN 1558-0644, doi: 10.1109/TGRS.2021.3062688.

Zhu, W., K. Xu, E. Darve, and G. C. Beroza (2021). A general approach to seismic inversion with automatic differentiation, *Comput. Geosci.* **151,** 104,751.

Zhu, W, K. Xu, E. Darve, B. Biondi, and G. C. Beroza (2020). Integrating deep neural networks with full-waveform inversion: Reparametrization, regularization, and uncertainty quantification, available at http://arxiv.org/abs/2012.11149 (last accessed August 2021).