Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Does Full Waveform Inversion Benefit from Big Data?

Peng Jin

pqj5125@psu.edu

Los Alamos National Laboratory

Yinan Feng

Los Alamos National Laboratory

Shihang Feng

Los Alamos National Laboratory

Hanchen Wang

Los Alamos National Laboratory

Yinpeng Chen

Microsoft

Benjamin Consolvo

Intel Corporation

Zicheng Liu

Microsoft

Youzuo Lin

Los Alamos National Laboratory

Article

Keywords:

Posted Date: December 28th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3782476/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Scientific Reports on August 28th, 2024. See the published version at https://doi.org/10.1038/s41598-024-68573-7.

Does Full Waveform Inversion Benefit from Big Data?

Peng Jin[1,2,*], Yinan Feng[1], Shihang Feng[1], Hanchen Wang[1], Yinpeng Chen[3], Benjamin Consolvo[4], Zicheng Liu[3], and Youzuo Lin[1,*]

[1]Earth and Environmental Sciences Division, Los Alamos National Laboratory
[2]College of Information Sciences and Technology, The Pennsylvania State University
[3]Microsoft
[4]Intel Corporation
[*]Corresponding Authors: pqj5125@psu.edu, ylin@lanl.gov

ABSTRACT

This paper investigates the impact of big data on deep learning models to help solve the full waveform inversion (FWI) problem. While it is well known that big data can boost the performance of deep learning models in many tasks, its effectiveness has not been validated for FWI. To address this gap, we present an empirical study that investigates how deep learning models in FWI behave when trained on OpenFWI, a collection of large-scale, multi-structural, synthetic datasets published recently. In particular, we train and evaluate the FWI models on a combination of 10 2D subsets in OpenFWI that contain 470K pairs of seismic data and velocity maps in total. Our experiments demonstrate that larger datasets lead to better performance and generalization of deep learning models for FWI. We further demonstrate that model capacity needs to scale in accordance with data size for optimal improvement.

Introduction

The recent advancements of deep learning in natural language processing and computer vision have proven that big data is one of the key ingredients for obtaining good performance[1–4]. Similarly, in the context of science, deep learning models such as AlphaFold[5] have achieved significant breakthroughs with the help of large-scale datasets. However, unlike these tasks, large-scale public datasets are not always available for many other scientific problems due to issues such as high data acquisition costs, labeling costs, intellectual property concerns, or security concerns. Due to limited dataset sizes and variation, deep learning models in scientific applications are often limited in their ability to generalize well to out-of-sample datasets.

Full waveform inversion (FWI) is a technique used to image the subsurface that has the potential to benefit from deep learning and large training datasets. Specifically, FWI aims to reconstruct subsurface velocity maps v from seismic measurements p as depicted in Figure 1. Conventional FWI methods[6–16] leverage the forward operator f governed by a partial differential equation (PDE) and perform iterative optimization per sample, which is computationally expensive and yields poor scalability. To mitigate this issue, deep learning techniques have been recently introduced to FWI and achieved promising performance[17–22]. A good summary of deep learning techniques for solving FWI problems can be found in Lin et al.[23]. In this paper, we follow the previous studies[23–25] and refer network-based FWI methods as data-driven methods. Inspired by the image-to-image translation task in computer vision, these data-driven methods directly learn an inverse mapping f^{-1} from seismic data directly to velocity maps. Nevertheless, due to the issue of lacking large-scale public datasets, the models in prior works were all developed on relatively small datasets (i.e. 130 to 67K data pairs)[24–26]. Thus, the question remains open: *does full waveform inversion benefit from big data?* Thankfully, the recently published large-scale datasets OpenFWI[27] provide us an opportunity to start to answer this question.

In this paper, we present an empirical study that attempts to answer the question of whether FWI benefits from large-scale and multi-structural training datasets from three perspectives: model performance, the relationship between the model size and data size, and model generalization. OpenFWI is a collection of large-scale, multi-structural datasets that cover different domain interests, including interfaces, geological faults, and field data. We employ 10 2D synthetic datasets from OpenFWI, and 408K/62K pairs of seismic data and velocity maps are used to train and evaluate the deep learning models, respectively. We adopt one of the OpenFWI benchmark models InversionNet[24] to serve as the baseline, and we compare the inversion results of the baselines trained on relatively small-scale individual datasets and the models trained on large-scale datasets that are composed of multiple datasets. We name the latter models *BigFWI*. Our findings are summarized as follows:

- **Big data can boost the performance of deep learning models in FWI.** BigFWI outperforms the baselines on almost every dataset in terms of all the evaluation metrics.
- **Larger data requires larger models.** When more training samples are introduced, larger model architectures are

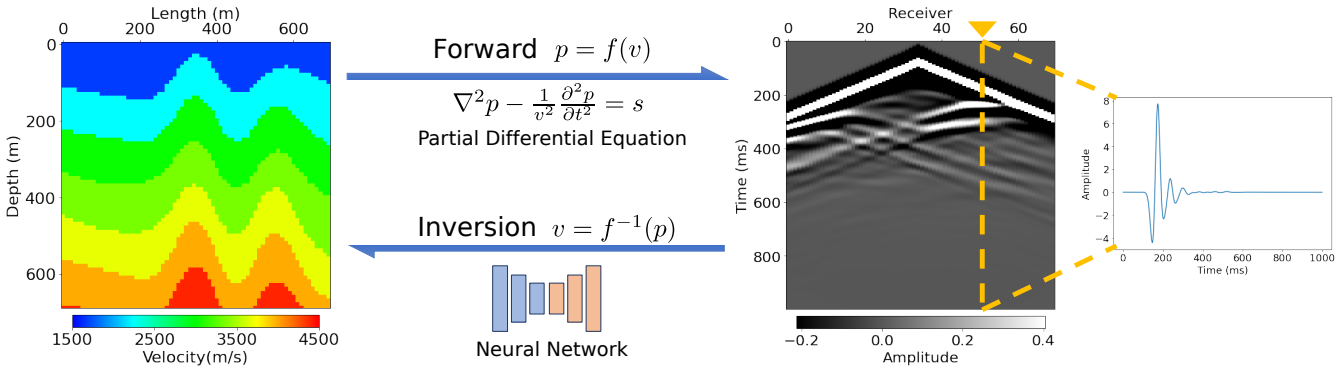


Figure 1. Schematic illustration of data-driven FWI and forward modeling. The forward modeling process computes the simulated seismic data from velocity maps, governed by a partial differential equation. Neural networks are employed in data-driven FWI methods to reconstruct velocity maps from seismic measurements.

required in BigFWI to achieve further improvement compared to the baselines.

- **Big data can improve the generalization of deep learning models in FWI.** Given a dataset that is unseen during training, BigFWI yields better performance than any baselines trained on single datasets.

Results

OpenFWI Datasets

We here briefly describe the OpenFWI datasets which are used in all the experiments. Unlike many existing synthetic datasets for FWI, OpenFWI is publicly available and offers a rich collection of large-scale multi-structural benchmark datasets. The datasets in OpenFWI are divided into four groups: "*Vel Family*", "*Fault Family*", "*Style Family*" and "*Kimberlina Family*". We exclude the "*Kimberlina Family*" in our experiments because the dimensions of both velocity maps and seismic data in "*Kimberlina Family*" are different from the other three families. This allows us to combine the data samples from different datasets to train BigFWI models. In terms of the complexity of subsurface structures, each of the three families consists of an easy version (*-A*) and a hard version (*-B*). In addition, the datasets in "*Vel Family*" and "*Fault Family*" are further divided into a flat version (*Flat-*) and a curved version (*Curve-*) in accordance with the shape of rock layers. The 10 datasets employed in our experiments are: *FlatVel-A/B*, *CurveVel-A/B*, *FlatFault-A/B*, *CurveFault-A/B*, and *Style-A/B*. We use dataset abbreviations such as FVA for FlatVel-A in the rest of the paper to simplify plots.

Each dataset in "*Vel Family*", "*Fault Family*", "*Style Family*" is split into 24K/6K, 48K/6K, and 60K/7K pairs of seismic data and velocity maps for training and testing, respectively. We follow this splitting through our experiments. Figure 1 shows an example of a velocity map and seismic data pair. Each velocity map has dimensions of 70×70 (depth $\times$ length in grids) with a grid spacing of 10 meters in both directions. The dimensions of the seismic data are $5 \times 1000 \times 70$ (# of sources $\times$ # of timesteps $\times$ # of receivers). Five sources are evenly distributed on the top surface, each of which is a Ricker wavelet with a central frequency of 15 Hz. The interval between timesteps is 1 millisecond, and the receivers are also placed with an interval of 10 meters. For more details about the forward model algorithm and simulation, please refer to the original paper of OpenFWI.

Evaluation Metrics

We follow the benchmarking guidelines in OpenFWI and compute three metrics between the ground truth and the prediction of velocity maps to evaluate the performance of a model: mean absolute error (MAE), root mean squared error (RMSE), and structural similarity (SSIM)[28]. Both MAE and RMSE are commonly used to measure pixel-wise errors, while SSIM aligns better with human vision and measures the perceptual similarity that is more related to structural information. When calculating MAE and RMSE, we keep the velocity maps in the normalized scale $[-1, 1]$. During the calculation of SSIM, we rescale the velocity maps to $[0, 1]$ as required by the algorithm.

Big data benefits FWI

We first design an experiment to explore if the performance of the data-driven FWI models can be improved by enlarging the training set. Specifically, we train a single model (i.e., BigFWI) on a combination of the datasets in OpenFWI and compare performance with the baseline InversionNet models trained on each split dataset. Both the BigFWI and InversionNet models share the same training hyperparameters and network architecture. Instead of using all the training samples in OpenFWI, we randomly select 12K samples from each of the four datasets in Vel Family, 24K from each of the four in the Fault Family, and

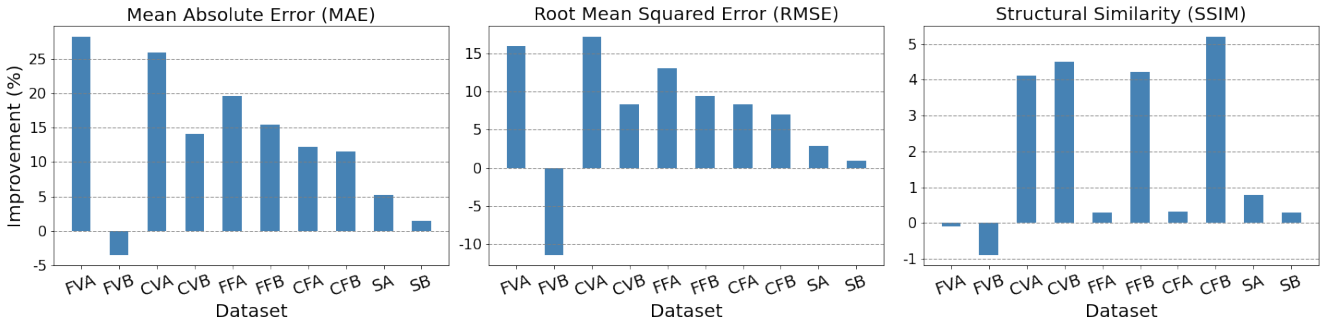


Figure 2. Performance improvement of BigFWI trained on OpenFWI-204K over InversionNet in terms of MAE, RMSE and SSIM. BigFWI trained on a large-scale dataset (OpenFWI-204K) yields better performance on almost every dataset, compared to InversionNet[24], which was trained on relatively small-scale datasets.

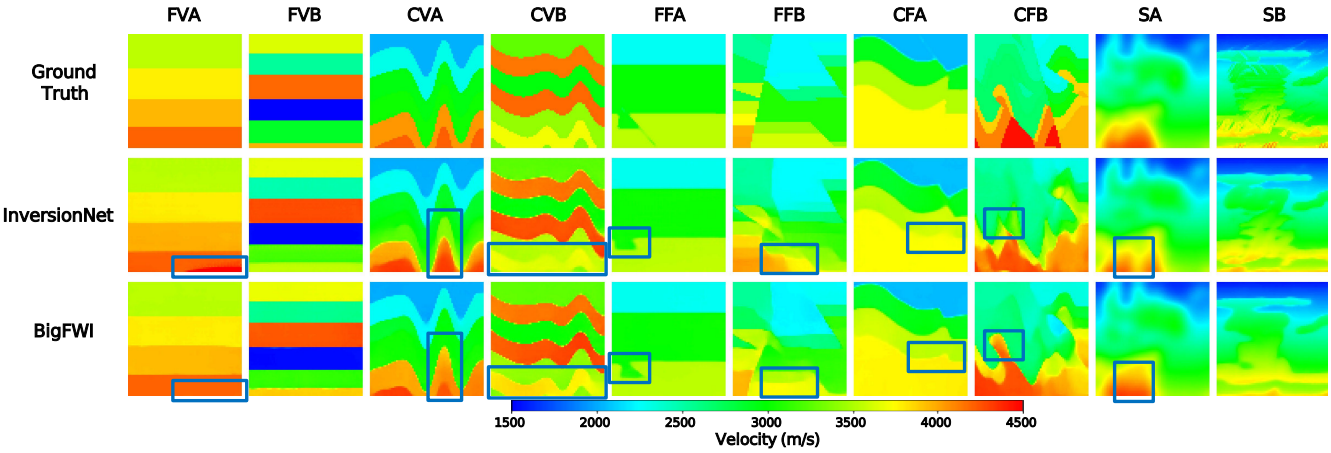


Figure 3. Comparison of ground truth (top) and predicted velocity maps generated by InversionNet (middle) and BigFWI (bottom). In general, BigFWI yields clearer layer boundaries, more accurate fault locations, and fewer artifacts compared to InversionNet as highlighted in squares.

30K from each of the two in the Style Family. The rest of the samples in the training sets are reserved for the experiment in the next section, where we further enlarge the datasets. The combined large-scale training set consists of 204K samples in total, and we name it OpenFWI-204K. The test sets are directly adopted from OpenFWI.

We plot the performance improvement of BigFWI compared to the InversionNet on each dataset in Figure 2. The quantitative results are provided in Supplementary Table S2. We observe that BigFWI shows a clear improvement for all the datasets except for datasets FVA and FVB, which are comprised of flat layers only. One potential reason for the model's degraded performance on FVA and FVB is that the network focuses more on curved layers which exist in most of the other datasets, and thus has a negative impact on the prediction of flat layers. We also observe that BigFWI exhibits significant improvement in MAE and RMSE for A datasets compared to B datasets across all families. However, the comparison of SSIM demonstrates the opposite trend, with the B datasets exhibiting better SSIM improvements compared to the A datasets in the same family. This variation in performance could be attributed to the greater complexity of the B datasets. The discrepancies in the baseline structures may not impact statistical misfits such as MAE and RMSE, but they may influence the SSIM. The simpler A datasets tend to benefit slightly more from the larger data volume than the more intricate B datasets.

Figure 3 shows a comparison of velocities maps between ground truth, InversionNet, and BigFWI. We observe that InversionNet predicts the velocity maps with various errors, such as extra bottom layer anomalies (FVA), inaccurate layer values (CVA, CVB), and inaccurate structures (FVB, FFB, CFA, CFB). BigFWI models generally yield better performance in predicting the structure and values of the velocity maps than InversionNet. We see that the improved results of BigFWI are due to the knowledge learned from the large-scale training dataset that consists of a variety of velocity map distributions. Here, we define the velocity map distributions as the different geological subsurface structures in OpenFWI: i.e., flat layers vs. curved layers, faults vs. non-fault, and smooth vs. sharp. However, in addition to benefits, the variety of the velocity map distributions

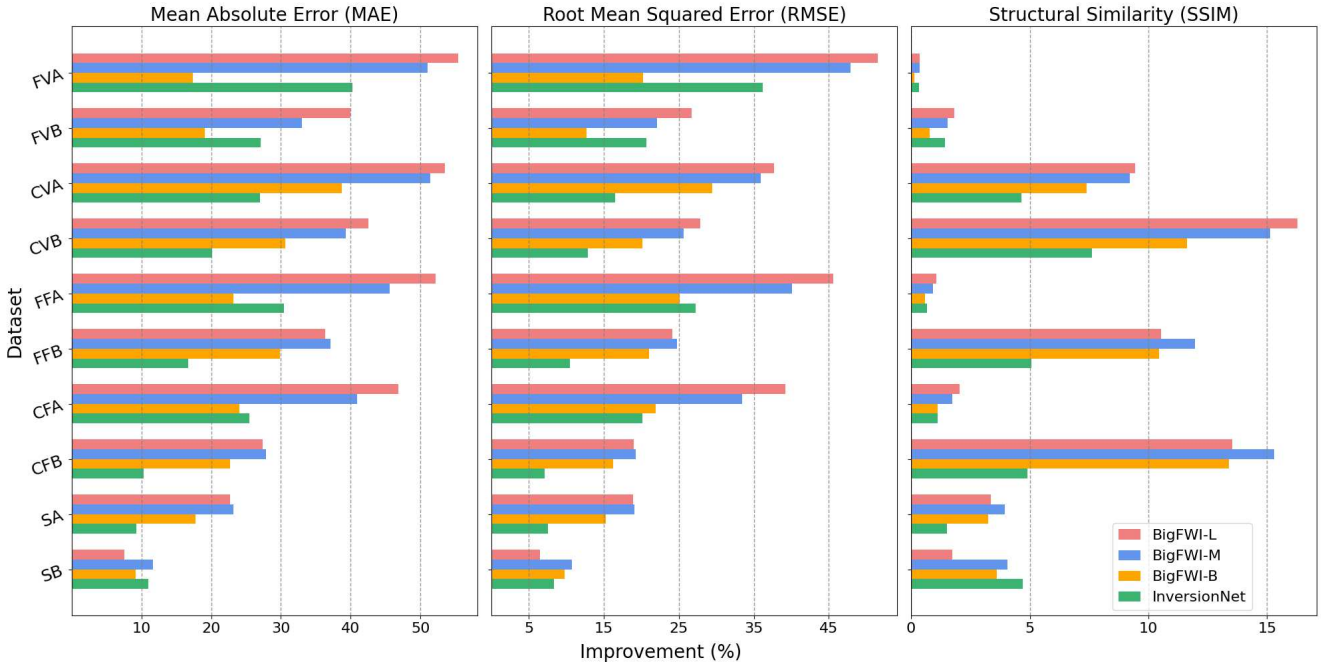


Figure 4. Comparison of the performance improvement of different methods trained on the further enlarged dataset (i.e. OpenFWI-408K) in terms of MAE, RMSE, and SSIM. Note that the improvement percentages are computed based on the InversionNet trained on OpenFWI-204K. For most of the datasets, the BigFWI-L, which has the largest model size, yields the best performance.

may also bring some negative effects such as inaccurate layer boundaries. For instance, we observe non-flat interfaces in the predictions of FVA/FVB, which are obviously affected by other velocity map distributions.

Big data in FWI requires larger models

To explore the relationship between the size of the training set and the size of the data-driven FWI models, we conduct an experiment that is similar to the previous one but employs the full training set provided by OpenFWI. Hence, the training set now contains 408K samples, and we name it OpenFWI-408K for brevity. Regarding model size, we introduce three variants of BigFWI as described in Supplementary Table S1, including the "**B**ase" configuration in the previous experiment, the "**M**iddle" model with additional layers in both encoder and decoder, and the "**L**arge" one that is both deeper and wider. The details of the network architecture are in the Method section. We keep the baselines which are InversionNet trained on OpenFWI-204K. Additionally, we train the InversionNet on each split datasets of OpenFWI-408K for comparison, and all these models share the same network architecture with BigFWI-B.

In Figure 4, we show the statistical performance improvement of BigFWI over InversionNet (trained on the split components of OpenFWI-204K). More detailed quantitative results are provided in Supplementary Table S3. Overall, all the models trained on OpenFWI-408K yield better performance compared to InversionNet trained on OpenFWI-204K, and the BigFWI models (coral, blue and orange bars) outperform InversionNet (green) for almost every dataset, which again verifies that larger training set brings better performance. We further observe that among three BigFWI variants, larger models yield better performance in general. In particular, BigFWI-L (coral) and BigFWI-M (blue) outperform BigFWI-B (orange) by a large amount in all three metrics for relatively simple datasets such as FVA, FVB, FFA and CFA. For relatively complicated datasets such as CFB and SA, the gap is narrower. For dataset SB, BigFWI-B even outperforms BigFWI-L. This infers that larger models are preferred for most big data scenarios, but additional efforts such as more advanced network architectures are still required for some complicated cases.

Figure 5 shows the ground truth and predictions of velocity maps InversionNet, BigFWI-B, BigFWI-M, and BigFWI-L. Though the performance of InversionNet has improved statistically when trained on larger datasets, errors in prediction such as extra bottom layer anomalies (FVA), inaccurate layer values (CVA, CVB), and inaccurate structures (FVB, FFB, CFA, CFB) still exist. In contrast, BigFWI generally offers enhanced accuracy in layer location and velocity values. Comparing the performance of the BigFWI models, BigFWI-L and BigFWI-M outperforms BigFWI-B in many aspects. For instance, the flat interfaces in FVA and FVB are more flat and sharp in the results of BigFWI-L and BigFWI-M than the ones of

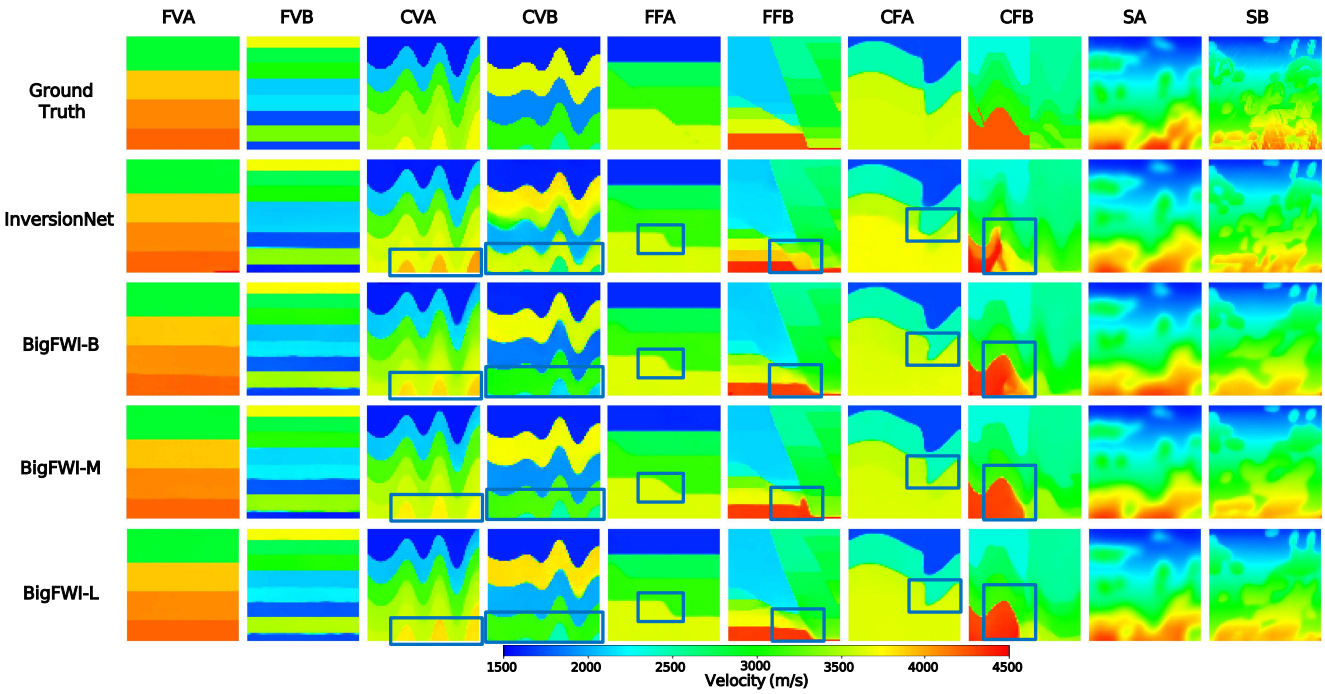


Figure 5. Comparison of ground truth (first row) and predicted velocity maps generated by InversionNet (second row) and BigFWIs (from third row to fifth row). BigFWI-L yields the most accurate results, especially in deep regions, compared to other BigFWI models and the baselines as highlighted in squares.

BigFWI-B. BigFWI-M also predicts more accurate fault slopes in FFA and FFB. A similar observation can be obtained from the Style Family results, in which BigFWI-L and BigFWI-M predict more accurate kinematic information than InversionNet and BigFWI-B. Though InversionNet predicts more high-frequency components, the scatters are inaccurate in shape, which introduces even larger data misfit.

Big data leads to better generalization

Leave-one-out Generalization Test: To verify whether large-scale training data also leads to better generalization, we design the experiment where the BigFWI models are trained under leave-one-out settings. Specifically, given a target dataset for testing (e.g., FVA), we train the BigFWI model on the combination of the training samples from all the other datasets in OpenFWI (e.g., FVB, CVA/B, FFA/B, CFA/B, and SA/B). We then compare the performance of this BigFWI model on the test samples of the target dataset (e.g., FVA) with InversionNet, which are trained on split datasets other than the target one.

In Figure 6, we present the statistical performance improvement in the percentage of the best generalization performance of InversionNet models. BigFWI shows superior performance across all the datasets, especially in terms of MAE and RMSE. This yields that big data leads to better generalization. Notably, utilizing datasets A as the target set results in greater improvements in terms of MAE and RMSE, while datasets B show greater improvements in terms of SSIM. The detailed quantitative results are provided in Supplementary Table S4.

Figure 7 compares the generalization results of different methods to the ground truth. We observe that InversionNet produces inaccurate layer structures for out-of-distribution (OOD) data. In FFA, FFB, CVA, FFA, and SA, InversionNet's generalization outputs have errors of blurred borders, wrong layer positions, and inaccurate velocity values, especially in deeper parts. Moreover, the results clearly have incorrect patterns from other datasets in more complex datasets (i.e., CVB, FFB, CFA and CFB). Meanwhile, these explain why we could find higher SSIM improvement in these four datasets in Fig. 6. Conversely, our BigFWI benefits from its large-scale cross-domain training set and can effectively capture more essential features of different datasets. Thus, BigFWI has more accurate predictions on OOD data than InversionNet.

Generalization Test on Marmousi & Overthrust: We further conduct generalization experiments on two more challenging standard test synthetic datasets Marmousi[29,30] and Overthrust[31,32]. Both velocity maps contain more practical subsurface structures and have been widely adopted for the evaluation of full waveform inversion methods[27,33–35]. Furthermore, the Marmousi velocity map was used as the style image to generate the Style Family in OpenFWI, which was specifically created for the simulation of real-world velocity maps.

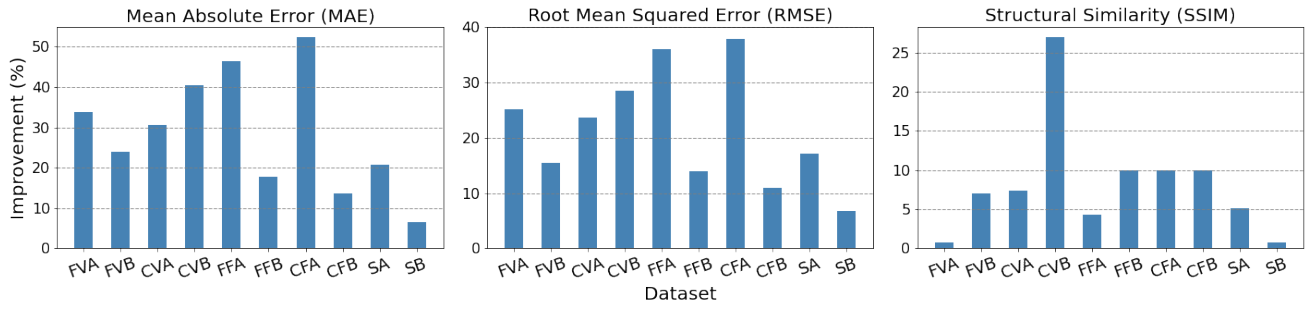


Figure 6. Generalization improvement of BigFWI models trained using leave-one-out settings in terms of MAE, RMSE and SSIM. For each target dataset, our BigFWI yields better generalization performance than all the InversionNet trained on the split datasets other than the target one.

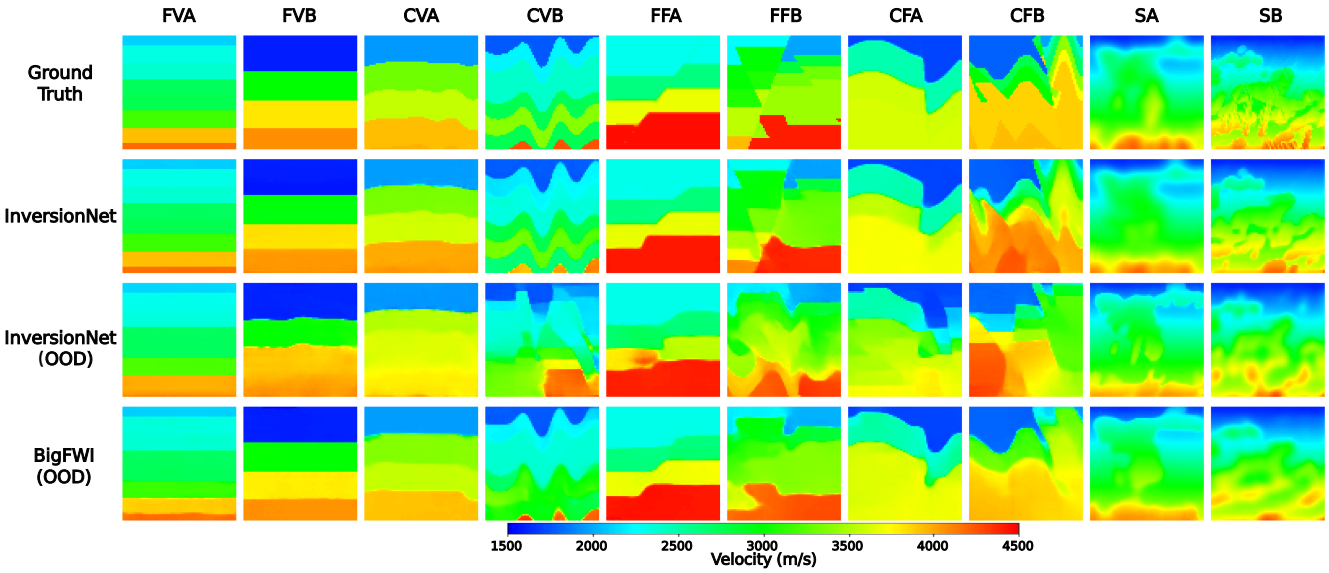


Figure 7. Comparison of ground truth (first row) and generalization results of different methods. From the first to last row: the InversionNet trained on the target datasets, the InversionNet trained on the datasets other than the target one, and the BigFWIs trained using leave-one-out settings. Our BigFWIs yield relatively reasonable velocity maps that are closer to the ones generated by the InversionNet trained on the target datasets.

In this experiment, we resize the original Marmousi and Overthrust velocity maps to match our sizes and generate the seismic data using the same configuration as in OpenFWI. Since the dataset SA is a smoothed version of SB in OpenFWI, we are also interested in the generalization performance of BigFWI on the smoothed versions of Marmousi and Overthrust. To this end, we follow the previous work[33] and apply Gaussian filters with a standard deviation 2 to the velocity maps to obtain the smoothed ones. For comparison, we compare the BigFWIs trained on OpenFWI-408K with the InversionNet models trained on SA and SB separately.

The generalization ability of our models to Marmousi and Overthrust are depicted in Figure 8. Generally, BigFWI yield more accurate inversion results compared to InversionNet. For the smoothed version of Marmousi, the results of BigFWI match the ground truth better in the shallow region. The BigFWI-M even generates some layered structures in the top-right corner. In the deep region, the results of the InversionNet models contain either too many false high-velocity predictions or a horizontal layer with relatively low velocity. In contrast, though the velocity in the results of BigFWI is lower than the ground truth, they capture the locations of high-velocity regions. For the original version of Marmousi, it is obvious that the performance of BigFWI is better than InversionNet. We observe the layered structures given by BigFWI, and we think this is learned from CVA and CVB.

For the Overthrust velocity maps, BigFWI consistently generates flat layers with geological faults in the deep region, which are more visually plausible than the results of InversionNet. We see that the behavior of BigFWI is greatly affected by FVB,

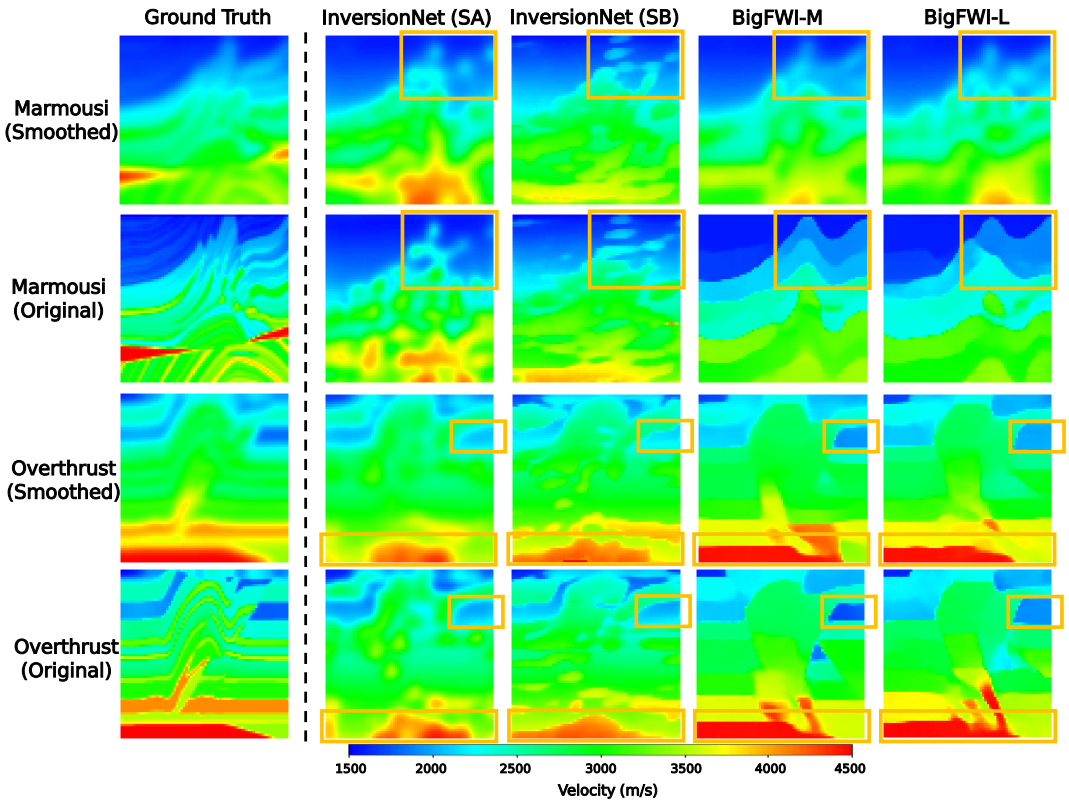


Figure 8. Comparison of ground truth (first column) and generalization results of different methods on Marmousi and Overthrust. For all velocity maps, Our BigFWIs yield more accurate results, especially in the shallow region of each velocity map as highlighted in squares.

which demonstrates the advantages of training models on large-scale multi-structural datasets. However, we also observe that BigFWI tends to follow one specific learned pattern per prediction; for example, in the predictions of the smoothed Overthrust, BigFWI still generates structures with sharp boundaries that exist in FVB. This indicates that there may be still much space for the improvement of BigFWI in terms of both the model architecture and the training data. The quantitative results are provided in Supplementary Table S5, which is consistent with our observations in the visualization results.

Discussion

We presented an empirical study to determine the extent to which big data can benefit the deep learning models in FWI from three perspectives: model performance, the relationship between model size and data size, and model generalization. To accomplish this, we utilized the large-scale, publicly available datasets OpenFWI and designed the experiments to compare the performance of baseline InversionNet trained on relatively small-scale individual datasets with that of BigFWIs, which are trained on combined, large-scale datasets. Through both quantitative and qualitative analysis, our study has demonstrated that big data can significantly enhance the performance of deep learning models in FWI on both in-distribution and out-of-distribution data. Moreover, we have shown that model architectures need to be scaled with data size to achieve further improvement. We trust that our findings can provide valuable guidance for the future development of deep-learning-based FWI methods.

This study is a preliminary investigation into the influence of big data on deep learning FWI methods, and there still exist some limitations and promising future directions. First, our study is entirely based on OpenFWI, which brings us not only convenience but also several inherent limitations. Although the Style Family in OpenFWI has made an effort to simulate the real-world velocity maps, there is still a gap between the synthetic data and field data. Our experiments are thus limited to simulations. It is an ongoing challenge for the whole FWI community to bridge this gap by either providing more public field data or improving the fidelity of the simulation. Second, in the present study, we only made minimal modifications to the network architecture of BigFWI. As a potential direction of future work, we may develop different network architectures to further improve performance. For instance, we observed during the qualitative analysis that the cross-domain training

could lead to interference between datasets and inaccurate layer boundaries. Such issues could potentially be addressed by implementing an adaptive network architecture. Third, throughout the experiments, we observe that the evaluation of inversion results is very complicated, and sometimes the differences in visualization results cannot be reflected in the current quantitative metrics. Hence, new evaluation metrics should also be developed in future to better reflect inversion quality.

This study offers valuable insights into the inverse problem, which can contribute to the advancement of this concept in other domains, including medical imaging, climate modeling, and astronomy. The knowledge gained from this investigation can be leveraged to support the application of AI in scientific research and enhance its capabilities in these fields.

Methods

Full Waveform Inversion

Figure 1 provides an illustration of 2D data-driven FWI and forward modeling. The governing equation of the acoustic wave forward modeling in an isotropic medium with a constant density can be described as follows:

$$\nabla^2 p(\boldsymbol{r},t) - \frac{1}{v(\boldsymbol{r})^2}\frac{\partial^2 p(\boldsymbol{r},t)}{\partial t^2} = s(\boldsymbol{r},t)\ , \tag{1}$$

where ∇^2 is the Laplace operator, $p(\boldsymbol{r},t)$ denotes the pressure wavefield at spatial location $\boldsymbol{r}$ and time t, $v(\boldsymbol{r})$ represents the velocity map of wave propagation, and $s(\boldsymbol{r},t)$ is the source term. As shown in Fig. 1, the goal of forward modeling is to simulate seismic data $\tilde{p}$ from a given velocity map v. For simplicity, we formulate this process as:

$$\tilde{p} = f(v), \tag{2}$$

where $f(\cdot)$ represents the highly nonlinear forward operator. As mentioned above, data-driven FWI methods directly learn the inverse mapping as:

$$\hat{v} = g_\theta(p) = f^{-1}(p), \tag{3}$$

where $\hat{v}$ is the estimated velocity map and $g_\theta(\cdot)$ is the approximated inverse operator of $f(\cdot)$, which is usually implemented as neural networks parameterized by θ. BigFWI is developed to leverage large-scale datasets to obtain a more precise and universal approximation of the inverse operator.

Network Architecture

Our BigFWI shares a similar network architecture with the baseline InversionNet. Overall, the model is comprised of an encoder $\mathcal{E}$ and a decoder $\mathcal{D}$. The encoder $\mathcal{E}$ first extracts the spatial-temporal features from the seismic input $p \in \mathbb{R}^{S\times T\times R}$ and compresses them into a latent vector $z = \mathcal{E}(x) \in \mathbb{R}^{L\times 1\times 1}$. Here, S equals the number of sources used in seismic surveys or simulation, T represents the number of samples recorded by each receiver, R denotes the number of receivers, and L is the length of the latent vector. The decoder $\mathcal{D}$ then transforms the latent vector z into spatial domain and generates the estimation of the velocity map $\hat{v} = \mathcal{D}(z) \in \mathbb{R}^{1\times W\times H}$, where W and H denote the horizontal (i.e. length) and vertical (i.e. depth) dimensions of the velocity map. Both the encoder $\mathcal{E}$ and the decoder $\mathcal{D}$ are fully based on 2D convolutional and deconvolution layers, and the details are presented as follows. The visualized network architecture of BigFWI is provided in Supplementary Figure S1.

In the encoder $\mathcal{E}$, since $T = 1000$ is much larger than $R = 70$ in the seismic data p of OpenFWI, we first reduce temporal dimension and extract temporal features by stacking seven convolutional layers with $n \times 1$ kernels, where $n = 7$ in the first layer, and $n = 3$ in the following six layers. The stride along the temporal dimension is set to 2 for every other layer to reduce the temporal dimension until it is close to the spatial dimension. We then stack six layers with 3×3 kernels to extract spatial-temporal features at the same time. Stride 2 is now applied to both dimensions every other layer. In BigFWI-M, instead of stacking six layers, we stack nine layers where an additional 3×3 layer with stride 1 is added after every two layers so as to increase model capacity without changing the dimensions of the original feature maps. In BigFWI-L, we stack eight layers where two additional 3×3 layer with 1024 features maps are appended at the end. In BigFWI-B and BigFWI-M, we use a layer with an 8×9 kernel to flatten the feature maps of the last to a 512-length latent vector z. In BigFWI-L, a layer with 4×5 kernel is used, and the length of latent vector is 1024.

The decoder $\mathcal{D}$ includes five deconvolution layers for upsampling, and each of them is followed by one convolutional layer with 3×3 kernels in BigFWI-B and two convolutional layers in BigFWI-M. The first deconvolution layer with kernel size 5 transforms the latent vector z into a $512 \times 5 \times 5$ tensor. The rest of the deconvolution layers with kernel size 4 and stride 2 upsample the feature maps by a factor of 2, resulting in an $80 \times 80 \times 32$ tensor. We then apply center cropping followed by a 3×3 convolutional layer to output a single channel 70×70 velocity map. In BigFWI-L, the kernel size of the first deconvolution layer is 2, and there are six groups of deconvolution and convolutional layers.

All the convolutional and deconvolution layers are followed by batch normalization and LeakyReLU as the activation function, except for the last output layer, which uses Tanh to generate the velocity map between $[-1, 1]$.

Loss Function

The original InversionNet model was trained with pixel-wise ℓ_1 loss or ℓ_2 loss between the ground truth of velocity maps v and the predictions $\hat{v}$. In this paper, we trained the baseline InversionNet and BigFWI using a combination of two loss functions to leverage the advantages from both sides according to the previous study[25]. The loss function can be written as:

$$\mathcal{L}(v,\hat{v}) = \frac{1}{W \cdot H}\sum_{i=1}^{W}\sum_{j=1}^{H}|v_{ij}-\hat{v}_{ij}| + \frac{1}{W \cdot H}\sum_{i=1}^{W}\sum_{j=1}^{H}\sqrt{(v_{ij}-\hat{v}_{ij})^2}, \tag{4}$$

where W and H denote the number of grids in horizontal length and depth directions, and v_{ij} and $\hat{v}_{ij}$ represent the ground truth velocity and the prediction at the grid (i, j).

Training Details

We use identical hyperparameters to train all the models in our experiments. Specifically, we employ AdamW optimizers with momentum parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$, and a weight decay of 1×10^{-4} to update the parameters of each model. The base learning rate is set to 8×10^{-4}, and the models are trained for 170 epochs. In the first five warm-up epochs, we linearly increase the learning rate from 1×10^{-4}, and we decay the learning rate by a factor of 10 at epoch 150 and epoch 160, respectively. The batch size is set to 256. All the models are implemented in PyTorch and trained on 4 NVIDIA Tesla V100 GPUs. We employ the natural logarithmic transformation to make the intensity of seismic data more balanced and normalize the data to range $[-1, 1]$ before they are fed into the network. The velocity maps are also normalized to the same scale before we compute the loss.

Data and Codes Availability

OpenFWI data set can be downloaded from the website (https://openfwi-lanl.github.io/). InversionNet codes are released and can be downloaded from the Website (https://github.com/lanl/OpenFWI/).

References

1. Brown, T. *et al.* Language models are few-shot learners. *Adv. neural information processing systems* **33**, 1877–1901 (2020).
2. Sun, C., Shrivastava, A., Singh, S. & Gupta, A. Revisiting unreasonable effectiveness of data in deep learning era. In *Proceedings of the IEEE international conference on computer vision*, 843–852 (2017).
3. Kolesnikov, A. *et al.* Big transfer (bit): General visual representation learning. In *European conference on computer vision*, 491–507 (Springer, 2020).
4. Schuhmann, C. *et al.* Laion-5b: An open large-scale dataset for training next generation image-text models. In *Thirty-sixth Conference on Neural Information Processing Systems Datasets and Benchmarks Track*.
5. Jumper, J. *et al.* Highly accurate protein structure prediction with alphafold. *Nature* **596**, 583–589 (2021).
6. Fichtner, A. *Full seismic waveform modelling and inversion* (Springer Science & Business Media, 2010).
7. Zhang, Z., Huang, L. & Lin, Y. A wave-energy-based precondition approach to full-waveform inversion in the time domain. In *SEG Technical Program Expanded Abstracts 2012*, 1–5 (Society of Exploration Geophysicists, 2012).
8. Ma, Y., Hale, D., Gong, B. & Meng, Z. Image-guided sparse-model full waveform inversion. *Geophysics* **77**, R189–R198 (2012).
9. Zhang, Z. & Huang, L. Double-difference elastic-waveform inversion with prior information for time-lapse monitoring. *Geophysics* **78**, R259–R273 (2013).
10. Feng, S. & Schuster, G. T. Transmission+ reflection anisotropic wave-equation traveltime and waveform inversion. *Geophys. Prospect.* **67**, 423–442 (2019).
11. Feng, S., Fu, L., Feng, Z. & Schuster, G. T. Multiscale phase inversion for vertical transverse isotropic media. *Geophys. Prospect.* **69**, 1634–1649 (2021).
12. Lin, Y. & Huang, L. Acoustic-and elastic-waveform inversion using a modified Total-Variation regularization scheme. *Geophys. J. Int.* **200**, 489–502 (2014).
13. Lin, Y. & Huang, L. Quantifying subsurface geophysical properties changes using double-difference seismic-waveform inversion with a modified Total-Variation regularization scheme. *Geophys. J. Int.* **203**, 2125–2149 (2015).

14. Hu, W., Abubakar, A. & Habashy, T. M. Simultaneous multifrequency inversion of full-waveform seismic data. *Geophysics* **74**, R1–R14 (2009).

15. Guitton, A. Blocky regularization schemes for full-waveform inversion. *Geophys. Prospect.* **60**, 870–884 (2012).

16. Chen, Y. *et al.* Multiscale reflection phase inversion with migration deconvolution. *Geophysics* **85**, R55–R73 (2020).

17. Wang, W. & Ma, J. Velocity model building in a crosswell acquisition geometry with image-trained artificial neural networks. *Geophysics* **85**, U31–U46 (2020).

18. Liu, B. *et al.* Deep-learning seismic full-waveform inversion for realistic structural modelsdl seismic fwi. *Geophysics* **86**, R31–R44 (2021).

19. Araya-Polo, M., Jennings, J., Adler, A. & Dahlke, T. Deep-learning tomography. *The Lead. Edge* **37**, 58–66 (2018).

20. Yang, F. & Ma, J. Deep-learning inversion: A next-generation seismic velocity model building method. *Geophysics* **84**, R583–R599 (2019).

21. Ren, Y., Nie, L., Yang, S., Jiang, P. & Chen, Y. Building complex seismic velocity models for deep learning inversion. *IEEE Access* **9**, 63767–63778 (2021).

22. Geng, Z. *et al.* Deep learning for velocity model building with common-image gather volumes. *Geophys. J. Int.* **228**, 1054–1070 (2022).

23. Lin, Y., Theiler, J. & Wohlberg, B. Physics-guided data-driven seismic inversion: Recent progress and future opportunities in full-waveform inversion. *IEEE Signal Process. Mag.* **40**, 115–133 (2023).

24. Wu, Y. & Lin, Y. InversionNet: An efficient and accurate data-driven full waveform inversion. *IEEE Transactions on Comput. Imaging* **6**, 419–433 (2019).

25. Zhang, Z. & Lin, Y. Data-driven seismic waveform inversion: A study on the robustness and generalization. *IEEE Transactions on Geosci. Remote. sensing* **58**, 6900–6913 (2020).

26. Araya-Polo, M. *et al.* Automated fault detection without seismic processing. *The Lead. Edge* **36**, 208–214 (2017).

27. Deng, C. *et al.* Openfwi: Large-scale multi-structural benchmark datasets for full waveform inversion. *Adv. Neural Inf. Process. Syst.* **35**, 6007–6020 (2022).

28. Wang, Z., Bovik, A. C., Sheikh, H. R. & Simoncelli, E. P. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* **13**, 600–612 (2004).

29. Brougois, A. *et al.* Marmousi, model and data. In *EAEG workshop-practical aspects of seismic data inversion*, cp–108 (European Association of Geoscientists & Engineers, 1990).

30. Martin, G. S., Wiley, R. & Marfurt, K. J. Marmousi2: An elastic upgrade for marmousi. *The leading edge* **25**, 156–166 (2006).

31. Aminzadeh, F., Burkhard, N., Long, J., Kunz, T. & Duclos, P. Three dimensional seg/eaeg models—an update. *The Lead. Edge* **15**, 131–134 (1996).

32. Aminzadeh, F., Weimer, P. & Davis, T. 3-d salt and overthrust seismic models. *Stud. Geol.* **42**, 247–256 (1996).

33. Feng, S., Lin, Y. & Wohlberg, B. Multiscale data-driven seismic full-waveform inversion with field data study. *IEEE Transactions on Geosci. Remote. Sens.* 1–14, DOI: 10.1109/TGRS.2021.3114101 (2021).

34. Jin, P. *et al.* Unsupervised learning of full-waveform inversion: Connecting CNN and partial differential equation in a loop. In *Proc. Tenth International Conference on Learning Representations (ICLR)* (2022).

35. Feng, Y. *et al.* An intriguing property of geophysics inversion. In *Proc. Thirty-ninth International Conference on Machine Learning (ICML)* (2022).

Acknowledgements

This work was funded by the U.S. Department of Energy (DOE) Office of Fossil Energy's Carbon Storage Research Program via the Science-Informed Machine Learning to Accelerate Real Time Decision Making for Carbon Storage (SMART-CS) Initiative.

Author contributions statement

Y.C. and P.J. proposed the idea and initialized the project. P.J. conducted the experiments and plotted the figures. S.F. and H.W. contributed geophysical expertise. P.J., Y.F., S.F., and H.W. analyzed the results and wrote the manuscript. Y.C., B.C., and Y.L. provided comments on the manuscript. Y.C., Z.L., and Y.L. supervised the project.

Competing interests

The authors declare no competing interests.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementary.pdf