VSNet: Deep-learning inversion of velocity structures and source locations with passive seismic data

Hongyu Sun, Department of Earth, Environmental and Resource Sciences, The University of Texas at El Paso*

SUMMARY

Jointly determining source locations and velocity structures is essential in passive seismic inversion due to the coupling between hypocenters and velocity structures. Traditional methods, including travel-time tomography and full-waveform inversion (FWI), face challenges due to their sensitivity to initial models and high computational costs. Recently, deep learning has emerged as a promising alternative, offering improved efficiency and accuracy in seismic inversion. However, most existing data-driven FWI studies focus on active-source data, where the source information is known, making their direct application to passive seismic data challenging. Here I propose VSNet, a multi-task learning framework for the joint inversion of velocity structures and source locations with passive seismic data. VSNet consists of two separate sub-networks: VNet for velocity inversion and SNet for source localization, which are updated simultaneously through a soft-sharing mechanism. Using a passive seismic dataset simulated on the OpenFWI Kimberlina-CO_2 velocity models, I demonstrate that VSNet can directly estimate both velocity structures and source locations from full-waveform data. The results show that VSNet achieves high accuracy in both tasks while being computationally efficient once trained, highlighting the potential of deep learning in passive seismic inversion.

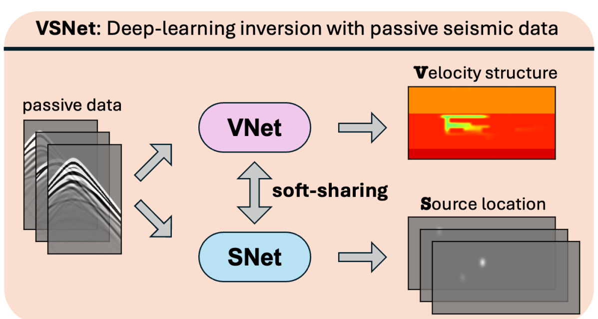


Figure 1: VSNet takes passive seismic records as its sole input, then separately generates a velocity structure via VNet and source locations via SNet. The two subnetworks exchange velocity and source information via a soft-sharing mechanism.

INTRODUCTION

The joint determination of source locations and velocity structures is crucial in passive seismic inversion due to the coupling between the earthquake hypocenters and velocity structures (Thurber, 1992). An inaccurate velocity model can lead to errors in source localization (Kamei and Lumley, 2014; Tong et al., 2016), while precise source locations form the foundation for passive seismic tomography. Passive data processing is generally classified into two main categories: travel-time-based and full-waveform-based methods. The former is widely used in earthquake seismology, where large-scale subsurface models are the primary focus. The latter is predominantly applied in microseismic monitoring, where high-resolution, small-scale features are of interest. Although full-waveform approaches are more computationally expensive, they are gaining increasing attention thanks to their ability to reveal more detailed geological structures. However, the joint inversion of passive seismic data in either way is inherently non-unique and challenging (Pavlis and Booker, 1980).

FWI with passive seismic data for velocity changes and source locations has been rapidly developed for microseismic monitoring (Behura, 2015). By incorporating time-reversal imaging of passive seismic sources into FWI (Sun et al., 2016; Lyu and Nakata, 2020) or traveltime inversion (Zheng et al., 2016), velocity models and event locations can be updated separately in an alternating manner. Further advancements include the incorporation of source time function inversion into the joint inversion workflow (Wang and Alkhalifah, 2018; Wang et al., 2020; Aghamiry et al., 2022). With a sufficiently dense receiver network capable of recording high-amplitude events, Diekmann et al. (2019) relaxed the requirement for high-quality priori velocity information by employing wavefront tomography to update the long-wavelength velocity in a similar workflow. Additionally, Song et al. (2019) linked velocity estimation to the focusing properties of source images obtained using the geometric-mean imaging condition (Nakata and Beroza, 2016), allowing for the simultaneous update of the long-wavelength velocity model and the location of a single point source. Despite the progress, the joint inversion of velocity and location still suffers from challenges such as significant computational resources, the lack of an accurate initial velocity model, and the feasibility to simultaneously locate multiple sources.

Deep learning has emerged as a powerful alternative for seismic inversion (Wu and Lin, 2019; Yang and Ma, 2019; Sun and Demanet, 2020, 2021; Jin et al., 2021; Sun et al., 2023b; Schuster et al., 2024). By training neural networks on large seismic datasets, deep learning methods can learn complex mappings between seismic wavefields and subsurface velocity structures, offering the potential to improve both the efficiency and accuracy of seismic inversion. With a well-designed deep-learning architecture and a diverse training dataset, AI-based inversion can directly infer subsurface velocity models from full-waveform data, without requiring the precise prior information that conventional methods depend on. Although generalization issues remain, as deep learning models are inherently limited by the diversity and quality of the training dataset, they hold promise for providing faster and more accurate inversion solutions.

However, most existing deep-learning inversion studies focus on active-source seismic data, where the source location is

10.1190/image2025-4316895.1

known. In contrast, passive seismic inversion presents significant challenges due to the uncertainty in seismic source parameters, making deep learning applications in passive seismic data inversion far more complex. Addressing this challenge requires deep-learning models to jointly estimate seismic source locations and velocity structures while incorporating advanced learning strategies to mitigate the inherent non-uniqueness of passive seismic inversion.

Unlike subsurface velocity model building, machine learning has been widely applied to passive seismic data for determining microseismic event locations, following two main approaches: picking-based earthquake monitoring enhanced by deep learning and end-to-end deep learning methods that directly map seismic waveforms to earthquake locations. The first approach incorporates deep learning into traditional earthquake monitoring workflows (Zhu and Beroza, 2019; Zhang et al., 2022; Sun et al., 2023a), where deep learning models assist in phase picking and arrival-time estimation, improving event detection and localization accuracy. The second approach directly predicts earthquake locations from seismic waveforms without requiring explicit arrival-time picking. For example, within the end-to-end deep learning category, Zhang et al. (2020) applied a convolutional neural network (CNN) to locate induced earthquakes in Oklahoma using a fixed seismic network. Wang and Alkhalifah (2021) trained one CNN to classify the number of seismic events in the data and another CNN to predict their source locations, peak frequencies, and amplitudes. Within the time-reversal imaging framework, Sun et al. (2022) trained neural operators to accelerate wavefield backpropagation for fast earthquake hypocenter determination. However, these methods assume that the velocity model is either explicitly known or implicitly embedded in the training dataset (Yang et al., 2024). When both earthquake locations and velocity structures are uncertain, it becomes crucial to estimate them simultaneously. Wamriew et al. (2022) determine microseismic event locations along with the velocity at those locations, using CNN with a single multi-channel output. However, this approach is limited to 1D layered velocity models and was primarily tested in a downhole microseismic monitoring setting. A general and robust learning framework for complex velocity models remains an open research challenge.

In this study, I propose a multi-task learning framework for the joint inversion of velocity structures and source locations using passive seismic data. I refer to this framework as VSNet, which consists of two separate networks for the two inversion tasks while being simultaneously updated through a soft-sharing mechanism (Figure 1). The source location task, SNet, generates Gaussian-distributed probability maps representing the likelihood of passive source locations. In contrast, the velocity inversion task, VNet, produces the subsurface velocity field. A numerical experiment using a passive seismic dataset simulated on the OpenFWI Kimberlina-CO_2 velocity models (Deng et al., 2022) demonstrates that the well-trained VSNet model can simultaneously predict seismic source locations and velocity structures from multi-source full-waveform data. Once trained, VSNet does not require initial velocity models and is computationally efficient.

METHOD

The joint inversion of velocity and source location naturally aligns with multi-task learning, as it involves predicting multiple related outputs from a single input dataset. In this work, each task requires distinct feature extraction patterns but benefits from high-level interactions due to velocity-source coupling. To leverage this, I propose a deep-learning model within the classic encoder-decoder framework, incorporating two independent sub-networks that share high-level extracted features via decoders with a soft-sharing mechanism. This design minimizes task interference, as the tasks are not highly correlated, while enabling distinct feature learning for velocity structure estimation and source localization. Additionally, it ensures cross-task information exchange, preserving the hypocenter-velocity coupling essential to the inversion process.

VSNet: Multi-task learning with soft parameter sharing

Here I introduce the method using a 2D velocity model; however, its extension to 3D is straightforward. I also assume that there are three microseismic events randomly triggered in the subsurface. Given passive seismic records $X \in \mathbb{R}^{C\times T\times N}$, where C represents the number of sources, N denotes the number of receivers, and T is the number of temporal sampling points, our goal is to estimate the subsurface velocity structure $V \in \mathbb{R}^{1\times H\times W}$ and the probability distribution of source locations $S \in \mathbb{R}^{C\times H\times W}$ where H and W correspond to the number of grid points in depth and horizontal distance, respectively.

Overall, the VSNet architecture consists of one VNet for velocity inversion and one SNet for source localization. The specific architectures of VNet and SNet are not fixed; various classical deep learning models, such as UNet, ResNet, and Transformers, can be employed within this multi-task learning framework. In this work, I implement UNet and leverage its classic encoder-decoder structure for both VNet and SNet. Each task has a separate encoder for feature extraction:

$$Z_v = f_{\theta_v}(X), \tag{1}$$
$$Z_s = f_{\theta_s}(X), \tag{2}$$

where θ_v and θ_s are the encoder parameters. To enable feature fusion with the soft parameter sharing mechanism, I define a learnable fusion function $\mathcal{F}$ as follows:

$$\tilde{Z}_v = \mathcal{F}_v(Z_v, Z_s), \tag{3}$$
$$\tilde{Z}_s = \mathcal{F}_s(Z_s, Z_v). \tag{4}$$

Here, $\mathcal{F}$ represents a trainable fusion mechanism, which adaptively combines the extracted features from different tasks while preserving task-specific representations. Each task has a separate decoder for task-specific predictions:

$$\hat{V} = g_{\phi_v}(\tilde{Z}_v), \tag{5}$$
$$\hat{S} = g_{\phi_s}(\tilde{Z}_s), \tag{6}$$

where g_{ϕ_v} and g_{ϕ_s} are decoder functions. I use supervised learning to minimize the difference between neural network outputs and labels. Considering the velocity output is a velocity field and location is represented as a probability map, I use

10.1190/image2025-4316895.1

Mean Squared Error (MSE) for velocity inversion loss and Binary Cross-Entropy (BCE) for source localization loss. Then the total loss is defined as:

$$\mathcal{L} = \lambda_v \mathcal{L}_{\mathrm{MSE}}(\hat{V}, V) + \lambda_s \mathcal{L}_{\mathrm{BCE}}(\hat{S}, S), \tag{7}$$

where λ_v and λ_s are task weights. With this learning strategy, we can effectively disentangle the coupling between velocity and source location while preserving their information exchange in determining the final solutions to the joint inversion problem.

In addition, as a UNet architecture, the encoder employs a convolutional downsampling strategy, while each decoder integrates skip connections. The velocity decoder utilizes a Tanh activation, whereas the source decoders apply Sigmoid activations to produce probability maps.

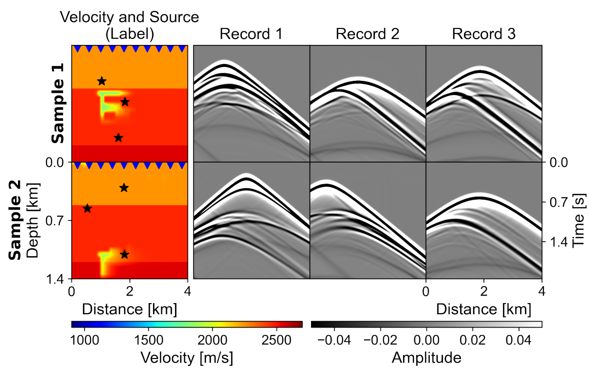


Figure 2: Examples of two samples from the dataset. Three microseismic events (marked by black stars) are randomly triggered in each model and recorded by 100 surface receivers. For clarity, only ten receivers are shown as blue triangles. These recorded waveforms serve as inputs to VSNet.

Dataset and Training

To train VSNet, I simulate a synthetic passive seismic dataset using the Kimberlina-CO_2 velocity model family from the OpenFWI dataset (Deng et al., 2022). These 2D acoustic velocity models describe the spatial and temporal migration of the supercritical CO_2 plume within a geologic carbon sequestration reservoir over a time frame of 200 years. I reserve models with continuous time evolution within a 20-year interval and randomly select 4,500 models for training and 619 models for testing. Since the OpenFWI dataset contains only active-source seismic surveys, I use only the velocity models from the dataset and generate passive seismic data specifically for this study.

Figure 2 presents examples of two samples from the dataset. Each velocity model uses a uniform grid spacing of 10 m. Each data sample is simulated on a velocity model with various passive sources triggered at random locations in the subsurface. The true source locations are retained to generate probability maps based on Gaussian distributions, indicating the likelihood of source positions.

Passive seismic data are simulated by solving 2D acoustic wave equations using the Deepwave software (Richardson, 2023). A Ricker wavelet is used as the source time function, and data is recorded with a sampling rate of 0.002 s. In total, 1024 time samples are recorded. There are 100 receivers evenly spaced along the surface of each velocity model, with a spacing of 40 m. I simulate three seismic events on each velocity model. The dimension of the input data is $3 \times 1024 \times 100$. The output source image has a dimension of $3 \times 141 \times 401$, while the velocity output has a dimension of $1 \times 141 \times 401$. The subsurface velocity distributions are normalized to $[-1, 1]$ before being fed into the neural networks.

The VSNet parameters are optimized using Adam with an initial learning rate of 2×10^{-5}. A step decay learning rate schedule is applied, reducing the learning rate by half every 50 epochs. The model is trained for 100 epochs with a batch size of 4 on a single GPU. Performance on velocity inversion is evaluated using root mean square error (RMSE), mean absolute error (MAE), and structural similarity index (SSIM), while source localization accuracy is evaluated based on both vertical $(z_{True} - z_{VSNet})$ and horizontal location $(x_{True} - x_{VSNet})$ errors.

RESULTS

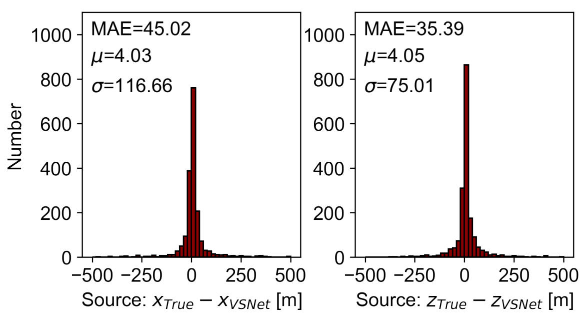
The trained VSNet model is applied to the test dataset for performance evaluation. The final source location is inferred from the probability map, where the maximum probability indicates the source position, assuming each output channel represents a single source. The histogram of all test samples shows that the MAE of location errors is 45.02 m in horizontal distance and 35.39 m in depth (Figure 3a). It is important to note that the resolution of the source image (grid spacing) affects the final localization accuracy, with finer grids leading to more precise source positions. Despite a discretization of 10 m in the source image, VSNet demonstrates high accuracy and robustness in source localization.

Figure 3b compares the accuracy of the predicted velocity models against the ground truth. The average RMSE between the predicted and true velocity models is 53.50 m/s, while the average SSIM index is 0.96, demonstrating the high accuracy of VSNet in velocity model building using passive seismic data. The histogram shows that the majority of samples have an SSIM greater than 0.9, confirming the effectiveness of the proposed method, particularly when both the training and test datasets follow the same data distribution.

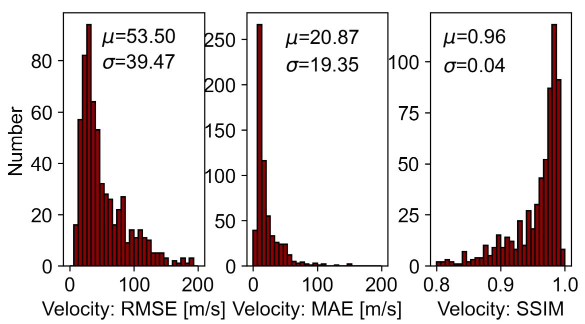
Figure 4 presents examples of VSNet's predictions on two test samples. The results demonstrate that VSNet successfully reconstructs the velocity distribution and accurately localizes passive sources using multiple passive records. The CO_2 reservoirs and the source probability maps are well recovered, showing high consistency with the ground truth.

Then I analyze the impact of velocity errors on source localization, assuming the velocity model is given as prior information. To achieve this, I adapt the input layer of SNet to accommodate both passive seismic records and the velocity model, allowing for the prediction of passive source locations. This design enables the manual introduction of errors into the input velocity model, facilitating the analysis of how velocity inaccuracies affect source localization.

Figure 5 shows that source location errors increase with growing errors in the input velocity model in both the horizontal and depth directions. Moreover, the localization error when using an accurate velocity model remains higher than that of VSNet, likely due to the lower generalization capability of a single network compared to a multi-task learning framework. Nonetheless, this experiment underscores the critical role of accurate velocity information in passive seismic source localization. Therefore, a joint inversion approach that leverages the coupling between velocity and source location is essential for enhancing accuracy in both tasks.



(a)

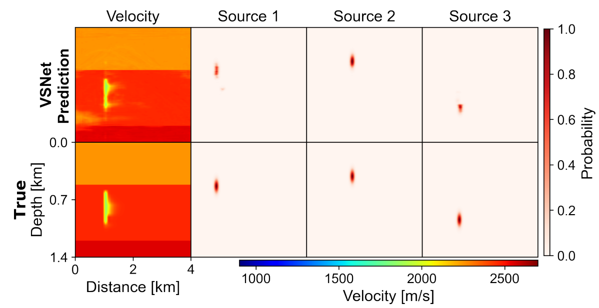


(b)

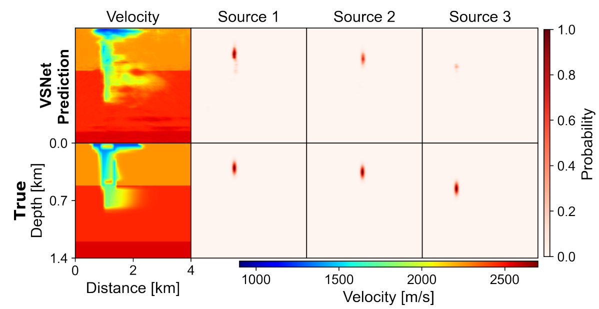
Figure 3: Performance on the test dataset. (a) Source location errors in horizontal $x_{\text{True}} - x_{\text{VSNet}}$ and vertical $z_{\text{True}} - z_{\text{VSNet}}$ directions. (b) Accuracy of velocity predictions measured by RMSE, MAE, and SSIM. The mean (μ), standard deviation (σ), and MAE of each histogram are labeled.

CONCLUSIONS

I propose VSNet, a multi-task deep learning framework for the joint inversion of subsurface velocity structures and seismic source locations using passive seismic data. Instead of relying on iterative optimization between velocity and source parameters, VSNet adopts a multi-task learning approach with a soft-parameter sharing mechanism to estimate both simultaneously from full-waveform data. Numerical experiments show that VSNet achieves high accuracy in both velocity model building and source localization. These results underscore the potential of deep learning to advance passive seismic inversion and contribute to the broader development of AI-driven FWI methods.



(a)



(b)

Figure 4: Examples of VSNet predictions on the test dataset.

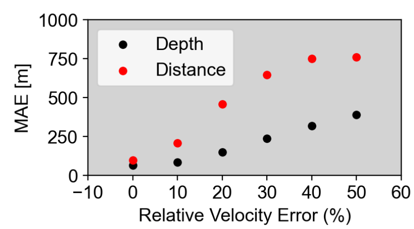


Figure 5: The impact of velocity errors on source localization in SNet, assuming the velocity model is known and provided as a secondary input alongside seismic records for determining source locations from passive seismic data.

ACKNOWLEDGMENTS

Hongyu Sun acknowledges UT System of STARs program and UTEP University Research Institute Grant for funding.

10.1190/image2025-4316895.1

REFERENCES

Aghamiry, H. S., A. Gholami, S. Operto, and A. Malcolm, 2022, ADMM-based full-waveform inversion for microseismic imaging: Geophysical Journal International, **228**, 259–274, doi: https://doi.org/10.1093/gji/ggab332.

Behura, J., 2015, Expedited FWI of microseismic data for mapping velocity changes and hypocenters, SEG Technical Program Expanded Abstracts 2015: Society of Exploration Geophysicists, 2461–2466, doi: https://doi.org/10.1190/segam2015-5908346.1.

Deng, C., S. Feng, H. Wang, X. Zhang, P. Jin, Y. Feng, Q. Zeng, Y. Chen, and Y. Lin, 2022, Openfwi: Large-scale multi-structural benchmark datasets for full waveform inversion: Advances in Neural Information Processing Systems, 6007–6020.

Diekmann, L., B. Schwarz, A. Bauer, and D. Gajewski, 2019, Source localization and joint velocity model building using wavefront attributes: Geophysical Journal International, **219**, 995–1007, doi: https://doi.org/10.1093/gji/ggz342.

Jin, P., X. Zhang, Y. Chen, S. X. Huang, Z. Liu, and Y. Lin, 2021, Unsupervised learning of full-waveform inversion: Connecting CNN and partial differential equation in a loop: arXiv preprint arXiv:2110.07584.

Kamei, R., and D. Lumley, 2014, Passive seismic imaging and velocity inversion using full wavefield methods: SEG International Exposition and Annual Meeting, SEG, SEG–2014.

Lyu, B., and N. Nakata, 2020, Iterative passive-source location estimation and velocity inversion using geometric-mean reverse-time migration and full-waveform inversion: Geophysical Journal International, **223**, 1935–1947, doi: https://doi.org/10.1093/gji/ggaa428.

Nakata, N., and G. C. Beroza, 2016, Reverse time migration for microseismic sources using the geometric mean as an imaging condition: Geophysics, **81**, KS51–KS60, doi: https://doi.org/10.1190/geo2015-0278.1.

Pavlis, G. L., and J. R. Booker, 1980, The mixed discrete-continuous inverse problem: Application to the simultaneous determination of earthquake hypocenters and velocity structure: Journal of Geophysical Research: Solid Earth, **85**, 4801–4810, doi: https://doi.org/10.1029/JB085iB09p04801.

Richardson, A., 2023, Deepwave (v0.0.20) [software].

Schuster, G. T., Y. Chen, and S. Feng, 2024, Review of physics-informed machine-learning inversion of geophysical data: Geophysics, **89**, T337–T356, doi: https://doi.org/10.1190/geo2023-0615.1.

Song, C., T. Alkhalifah, and Z. Wu, 2019, Microseismic event estimation and velocity analysis based on a source-focusing function: Geophysics, **84**, KS85–KS94.

Sun, H., and L. Demanet, 2020, Extrapolated full-waveform inversion with deep learning: Geophysics, **85**, R275–R288, doi: https://doi.org/10.1190/geo2019-0195.1.

Sun, H., and L. Demanet, 2021, Deep learning for low-frequency extrapolation of multicomponent data in elastic FWI: IEEE Transactions on Geoscience and Remote Sensing, **60**, 1–11.

Sun, H., Z. E. Ross, W. Zhu, and K. Azizzadenesheli, 2023a, Phase neural operator for multi-station picking of seismic arrivals: Geophysical Research Letters, **50**, e2023GL106434.

Sun, H., Y. Sun, R. Nammour, C. Rivera, P. Williamson, and L. Demanet, 2023b, Learning with real data without real labels: a strategy for extrapolated full-waveform inversion with field data: Geophysical Journal International, **235**, 1761–1777, doi: https://doi.org/10.1093/gji/ggad330.

Sun, H., Y. Yang, K. Azizzadenesheli, R. W. Clayton, and Z. E. Ross, 2022, Accelerating time-reversal imaging with neural operators for real-time earthquake locations: arXiv preprint arXiv:2210.06636.

Sun, J., Z. Xue, T. Zhu, S. Fomel, and N. Nakata, 2016, Full-waveform inversion of passive seismic data for sources and velocities: SEG International Exposition and Annual Meeting, SEG, SEG–2016, doi: https://doi.org/10.1190/segam2016-13959115.1.

Thurber, C. H., 1992, Hypocenter-velocity structure coupling in local earthquake tomography: Physics of the Earth and Planetary Interiors, **75**, 55–62, doi: https://doi.org/10.1016/0031-9201(92)90117-E.

Tong, P., D. Yang, Q. Liu, X. Yang, and J. Harris, 2016, Acoustic wave-equation-based earthquake location: Geophysical Supplements to the Monthly Notices of the Royal Astronomical Society, **205**, 464–478, doi: https://doi.org/10.1093/gji/ggw026.

Wamriew, D., M. Charara, and D. Pissarenko, 2022, Joint event location and velocity model update in real-time for downhole microseismic monitoring: A deep learning approach: Computers & Geosciences, **158**, 104965, doi: https://doi.org/10.1016/j.cageo.2021.104965.

Wang, H., and T. Alkhalifah, 2018, Microseismic imaging using a source function independent full waveform inversion method: Geophysical Journal International, **214**, 46–57, doi: https://doi.org/10.1093/gji/ggy121.

Wang, H., and T. Alkhalifah, 2021, Direct microseismic event location and characterization from passive seismic data using convolutional neural networks: Geophysics, **86**, KS109–KS121, doi: https://doi.org/10.1190/geo2020-0636.1.

Wang, H., Q. Guo, T. Alkhalifah, and Z. Wu, 2020, Regularized elastic passive equivalent source inversion with full-waveform inversion: Application to a field monitoring microseismic data set: Geophysics, **85**, KS207–KS219, doi: https://doi.org/10.1190/geo2019-0738.1.

Wu, Y., and Y. Lin, 2019, InversionNet: An efficient and accurate data-driven full waveform inversion: IEEE Transactions on Computational Imaging, **6**, 419–433, doi: https://doi.org/10.1109/TCI.2019.2956866.

Yang, F., and J. Ma, 2019, Deep-learning inversion: A next-generation seismic velocity model building method: Geophysics, **84**, R583–R599, doi: https://doi.org/10.1190/geo2018-0249.1.

Yang, Y., O. M. Saad, and T. Alkhalifah, 2024, Deep learning-based 3D microseismic event direct location using simultaneous surface and borehole data: Journal of Geophysical Research: Machine Learning and Computation, **1**, e2024JH000365.

Zhang, M., M. Liu, T. Feng, R. Wang, and W. Zhu, 2022, LOC-FLOW: An end-to-end machine learning-based high-precision earthquake location workflow: Seismological Society of America, **93**, 2426–2438.

Zhang, X., J. Zhang, C. Yuan, S. Liu, Z. Chen, and W. Li, 2020, Locating induced earthquakes with a network of seismic stations in Oklahoma via a deep learning method: Scientific Reports, **10**, 1941, doi: https://doi.org/10.1038/s41598-020-58908-5.

Zheng, Y., Y. Wang, and X. Chang, 2016, Wave equation based microseismic source location and velocity inversion: Physics of the Earth and Planetary Interiors, **261**, 46–53, doi: https://doi.org/10.1016/j.pepi.2016.07.003.

Zhu, W., and G. C. Beroza, 2019, PhaseNet: A deep-neural-network-based seismic arrival-time picking method: Geophysical Journal International, **216**, 1831–1841, doi: https://doi.org/10.1093/gji/ggy529.