ACCSCIENCE PUBLISHING

Journal of Seismic Exploration

ARTICLE

A physics-constrained autoencoder for full-waveform inversion using axial self-attention

Yunbo Niu[1,2], Yingming Qu[1,2]*, and Zhenchun Li[1,2]

[1]State Key Laboratory of Deep Oil and Gas, China University of Petroleum (East China), Qingdao, Shandong, China

[2]School of Geosciences, China University of Petroleum (East China), Qingdao, Shandong, China

(This article belongs to the *Special Issue: Full Waveform Inversion Methods and Applications for Seismic Data in Complex Media*)

Abstract

Full-waveform inversion (FWI) is highly sensitive to the initial model and low-frequency content, and it often suffers from cycle skipping and degraded resolution in complex media. We propose a physics-constrained autoencoder-based FWI with axial self-attention (AxPCAE-FWI). In a unified encoder–decoder architecture, a differentiable acoustic wave-equation solver is explicitly embedded, and the data-domain waveform misfit is used as the primary objective, so that training is consistently governed by wave physics and does not rely on paired seismic–velocity labels. The encoder extracts inversion-relevant, low-dimensional features, while the decoder reconstructs physically admissible velocity models. To capture long-range spatiotemporal dependencies in the time–offset plane, axial multi-head self-attention is introduced in the encoder, where global attention is computed separately along the time and receiver axes; two one-dimensional global attentions approximate a single two-dimensional (2D) global attention, reducing the computational complexity relative to full 2D attention while preserving global context. This design improves the representation of complex wavefield phenomena, including multiples, converted waves, and far-offset reflections, thereby alleviating cycle skipping when low-frequency information is limited. Numerical experiments on the Marmousi2 and Society of Exploration Geophysicists salt-dome models demonstrate stable convergence and high structural similarity with improved geological plausibility. Compared to conventional physics-informed adaptive extended FWI under the same iteration budget, AxPCAE-FWI yields clearer salt-body boundaries and better imaging of structurally complex regions, with improved robustness to noise.

Keywords: Full-waveform inversion; Deep learning; Self-attention; Physics-constrained autoencoder; Autoencoder

***Corresponding author:**
Yingming Qu
(20180041@upc.edu.cn)

Citation: Niu Y, Qu Y, Li Z. A physics-constrained autoencoder for full-waveform inversion using axial self-attention. *J Seismic Explor*. 2026;35(1):269-281. doi: 10.36922/JSE025480119

Received: November 24, 2025

Revised: January 9, 2026

Accepted: January 9, 2026

Published online: February 23, 2026

1. Introduction

High-resolution imaging of the subsurface is a key step in improving the reliability of structural interpretation, reducing the overall cost of geophysical exploration, and enhancing the efficiency of resource development. As exploration targets have gradually shifted from simple structures to deeper, more complex formations and unconventional reservoirs, conventional imaging methods based on wave-equation migration and

ray theory have become increasingly inadequate in terms of resolution and quantitative inversion capability. Consequently, full-waveform inversion (FWI) has attracted widespread attention. By solving the forward wave equation and minimizing the waveform misfit between simulated and observed seismic records, FWI enables high-resolution reconstruction of subsurface medium parameters and is regarded as an important bridge between seismic imaging and subsurface property inversion.[1-4] In its classical framework, FWI typically adopts gradient-based or quasi-Newton optimization schemes to iteratively update the velocity model by minimizing a data-misfit objective function, and in theory can recover subsurface structures at wavelength or even subwavelength scales. In practical applications, however, acoustic FWI remains the most widely used form. It solves only the acoustic wave equation and inverts for the P-wave velocity model, simplifying the Earth as an isotropic, lossless scalar medium. This approximation is relatively effective in marine environments, but in complex onshore settings, the subsurface usually exhibits pronounced elastic behavior, anisotropy, and strong heterogeneity. Under such conditions, the acoustic approximation cannot explicitly account for S-waves and other elastic parameters, and it treats multiparameter coupling effects, such as those between velocity and density, in an oversimplified manner, thereby limiting the physical consistency and interpretational reliability of the inversion results. Mora[5] likened FWI to a combination of migration and tomography: the FWI process is equivalent to performing migration imaging and reflection tomography simultaneously, where tomography updates the background velocity field, and migration moves reflectors to their correct positions. Thus, the quality of an FWI result can be judged by how well the recovered model geometry matches the true reflector geometry. Nevertheless, FWI is intrinsically a strongly nonlinear inverse problem and faces multiple challenges. It is highly sensitive to the initial model: if the starting velocity field significantly deviates from the true subsurface, or if the observed data lack sufficient low-frequency content, large phase and cycle discrepancies arise between simulated and real wavefields. This leads to a highly multimodal objective function, making the inversion prone to being trapped in local minima, a phenomenon known as cycle skipping.[4] To provide a more suitable initial model for FWI, Xu *et al.*[6] proposed reflection waveform inversion, which fully exploits reflection data to simultaneously update the background velocity and delineate reflectors, and the resulting model can then serve as the initial model for FWI. Moreover, the computational cost of FWI is extremely high: large-scale forward and adjoint wavefield simulations must be repeatedly performed at high spatial and temporal resolution, imposing stringent demands on memory and computational resources.

In recent years, deep learning techniques have been successfully applied across numerous research fields.[7] In the field of geophysics, classical multilayer perceptron and convolutional neural network (CNN) architectures have been used for reflectivity inversion,[8,9] velocity inversion,[10-15] seismic tomography,[16-20] and impedance inversion.[21,22] In the context of velocity model building and waveform-based inversion, learning-based studies related to FWI can be described according to how neural networks are used with the physics-driven FWI workflow. A line of work targets data-driven seismic inversion, where a network learns a direct mapping from seismic waveforms to velocity models and produces rapid velocity estimates in a machine-learning paradigm.[23,24] These outputs are not necessarily generated using an iterative FWI update loop; however, the predicted models can be used to support FWI in a practical manner, for example, by providing profile- or model-level initialization for subsequent FWI iterations.[25] Another line of work uses deep learning to construct informative starting models for conventional physics-based FWI, such as CNN-based starting-model building for near-surface two-dimensional (2D) FWI.[26] In addition, deep learning can act within an iterative FWI procedure by introducing learned regularizers or plug-in refinement modules to improve stability and convergence, including diffusion-model-based learned regularization for multiparameter elastic FWI[27] and the "deep velocity generator" that refines starting or intermediate velocity models (together with migrated images) during FWI to mitigate cycle skipping and accelerate convergence.[28] Finally, a growing body of research couples neural networks more tightly with wave-equation physics, enabling training driven by data-domain waveform misfit and physical constraints without requiring paired seismic–model labels, such as physics-informed generative adversarial network-based FWI,[24] latent representation learning in physics-informed neural networks for FWI,[29] parametric CNN-domain FWI,[30] and deep neural network-based reparameterized FWI frameworks for regularization and uncertainty quantification.[31-33] Collectively, these developments indicate that integrating neural networks with wave-equation physics provides a viable route to stable, accurate velocity model building.

In this study, we propose a method that integrates the acoustic wave equation into an autoencoder architecture equipped with axial self-attention mechanisms. Physically, coherent reflections, multiples, and converted-wave arrivals follow kinematically consistent trajectories on the time–offset plane, governed by propagation paths and limited-aperture (Fresnel-zone) coherence. The axial self-attention exploits this structure by aggregating

information separately along time and offset: time-axis attention emphasized phase-consistent temporal patterns, while offset-axis attention adaptively selects an effective receiver aperture and downweights kinematically inconsistent energy. As a result, the network could better capture long-range spatiotemporal correlations and preserve event continuity and phase consistency in the time–offset domain. The acoustic wave equation was used to generate synthetic seismic data for a given velocity model, while the observed seismic records were used as inputs to the network. Under these physical constraints, the network was trained to infer the corresponding velocity model. The reconstructed velocity model was then input into an acoustic finite-difference partial differential equation forward simulator to generate synthetic seismic data. The network loss is defined as the discrepancy between the synthetic data predicted from the current velocity model and the reference (observed) seismic data. By backpropagating this misfit, we computed the gradients of the loss with respect to the network weights, enabling the network to learn how to transform latent features extracted from the observed data into an accurate velocity model. This physics-informed framework did not require an explicit initial model or low-frequency data to achieve convergence, thereby mitigating the cycle-skipping problem. We demonstrated the effectiveness of the proposed method on the Marmousi2 model and the Society of Exploration Geophysicists (SEG) salt model.

2. Methods

In conventional FWI, time-domain elastic velocity-stress equations are used for forward modeling to generate synthetic seismic data. The ultimate goal is to minimize the misfit between observed and synthetic data to invert for the velocity model. The data misfit is typically defined by the following objective function (**Equation 1**):

$$\min \mathcal{J}(m) = \sum_{k=1}^{N_s} \sum_{j=1}^{N_r} \sum_{n=1}^{N_t} \left\| d_{obs} - d_{pred} \right\|_2^2 \quad (1)$$

where d_{obs} denotes the observed seismic data and d_{pred} represents the synthetic data generated from the current model.

To minimize the objective function in **Equation 1**, the inversion is typically initialized with a starting model and then iteratively updated according to the following rule (**Equation 2**):

$$m^{k+1} = m^k - \alpha \nabla C(m^k) \quad (2)$$

Where α is the step size, and $\nabla C(m^k)$ denotes the gradient of the objective function with respect to the model parameters. The gradient is typically computed using the adjoint-state method,[34] which enables efficient evaluation of sensitivity kernels for updating the model.

We utilized the physics-constrained autoencoder-based FWI with axial self-attention (AxPCAE-FWI) network architecture as illustrated in Figure 1. We followed the optimization philosophy of conventional FWI and took the observed seismic records d_{obs} as input. Through a data-driven, physics-constrained encoder–decoder framework, we obtained the subsurface parameter model m. The encoder was used to extract a low-dimensional latent representation that was most relevant to inversion from d_{obs}, while the decoder mapped the extracted latent features,

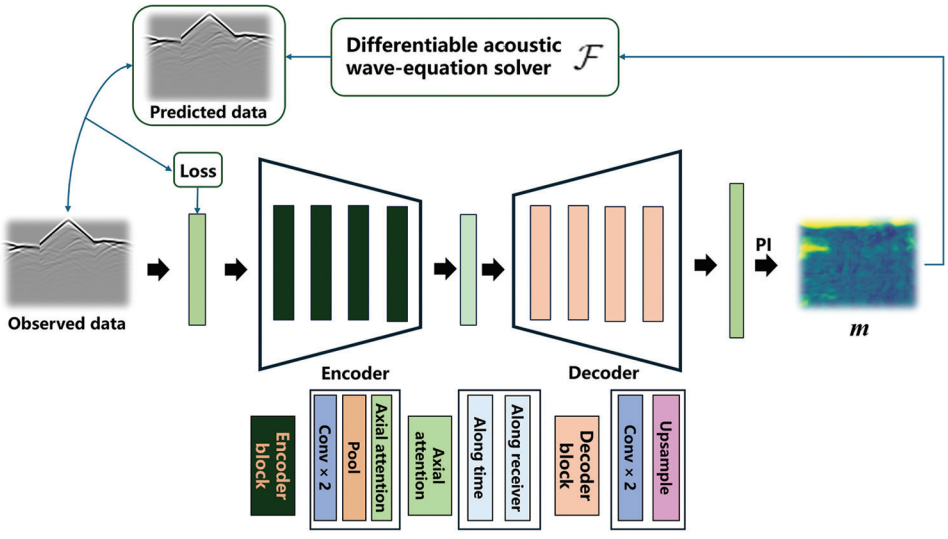

Figure 1. Architecture of the physics-constrained autoencoder-based full-waveform inversion with axial self-attention
Abbreviation: PI: Physics-informed.

under explicit physical constraints, to the desired velocity model. The predicted velocity model was then passed to a differentiable acoustic wave-equation solver $\mathcal{F}$ to generate synthetic seismic data d_{pred} corresponding to the current model. A data-domain consistency loss was constructed between d_{pred} and d_{obs}, thereby enforcing physical laws during network training.

Once the observed seismic data d_{obs} were input to the network, they passed through four pooling layers in the encoder, where inversion-related features were progressively extracted and compressed into a latent representation, denoted as $\hat{z} \in (0,1)$. To enhance the modeling capability for long-range spatiotemporal correlations, we inserted axial multi-head self-attention blocks after selected pooling layers: scaled dot-product attention was computed independently along the time and receiver (offset) axes. The resulting attention features were fused with the backbone features via residual connections. The key idea was to approximate a full 2D global attention operation using two one-dimensional global attentions; compared to 2D global attention, axial multi-head self-attention substantially reduced computational cost. Meanwhile, for seismic records defined on the time–offset plane, which exhibited a strong one-dimensional sequential structure in time and were subjected to clear physical constraints along offset, axial self-attention was better suited than purely local convolutions for capturing long-range spatiotemporal dependencies associated with multiples, converted waves, and far-offset reflections, while preserving the continuity and phase consistency of reflection events in both time and space. Each attention block was followed by a point-wise feedforward network and residual connections to further enhance the representational power. The resulting feature maps were then fed into the decoder, which contained four upsampling layers that gradually restored the original spatial resolution of the latent representation. Finally, an affine transformation mapped the normalized network output to a physically admissible parameter interval [m_{min}, m_{max}], yielding the target model m (**Equation 3**):

$$m = m_{\min} + \sigma(\hat{z}) \odot \left(m_{\max} - s m_{\min}\right) \tag{3}$$

Where $\sigma(\cdot)$ is the Sigmoid function and $\odot$ denotes element-wise (Hadamard) multiplication. This mapping was implemented using the Sigmoid layer followed by linear rescaling at the end of the decoder, which guaranteed that the output parameters fell within the prescribed physical bounds. The Sigmoid function is defined as **Equation 4**:

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{4}$$

The obtained model m was then fed into the differentiable acoustic wave-equation solver $\mathcal{F}$ to generate the simulated seismic record d_{pred}, as shown in **Equation 5**:

$$d_{\text{pred}} = \mathcal{F}(m) \tag{5}$$

Accordingly, the objective function used in all experiments is the data-domain waveform-fitting loss (**Equation 6**):

$$\mathcal{L}_{\text{data}} = \frac{1}{N}\sum_{i=1}^{N}\left(\left\| d_{pred}(i) - d_{obs}(i)\right\|^2\right) \tag{6}$$

Notably, we did not introduce an explicit hidden/latent-space regularization term in this work; therefore, its weight was set to zero. The latent bottleneck and the physically admissible output mapping in **Equation 3** acted as implicit regularization.

The gradients of this loss with respect to the network weights were computed through backpropagation, and the Adam optimizer was then used to update ω, as shown in **Equation 7**:

$$\omega_{t+1} = \omega_t - \alpha \cdot \frac{\partial \mathcal{L}_{data}}{\partial \omega} \tag{7}$$

Where α denotes the learning rate.

As the iterations progressed, the network learned to optimize a latent representation that could generate synthetic seismic records closely matching the observed data and to decode this representation into the velocity model, thereby yielding accurate FWI results.

We assessed the performance of the proposed AxPCAE-FWI framework using the structural similarity index (SSIM) and the peak signal-to-noise ratio (PSNR). SSIM quantifies perceptual similarity by jointly comparing luminance, contrast, and structural information between two images, and thus reflects how well edges and fine spatial details are preserved. The SSIM is defined as **Equation 8**:

$$\text{SSIM}(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{8}$$

where μ_x and μ_y denote the mean values of the two images, σ_x^2 and σ_y^2 represent their variances, and σ_{xy} is the covariance between them.

An SSIM value closer to 1 indicates a higher degree of structural similarity. The PSNR is defined as **Equation 9**:

$$\text{PSNR} = 10 \cdot \log_{10}\left(\frac{M^2}{\text{MSE}}\right) \tag{9}$$

Where M denotes the maximum possible (peak) value of the original image. A larger PSNR value indicates higher numerical fidelity and better reconstruction quality.

3. Results

3.1. Marmousi2 model

We evaluated the inversion performance of AxPCAE on the Marmousi2 model. The selected Marmousi2 model had a size of 1,790 × 1,410 grid points, with both horizontal and vertical grid spacing set to 10 m. The source was a Ricker wavelet with a dominant frequency of 20 Hz.

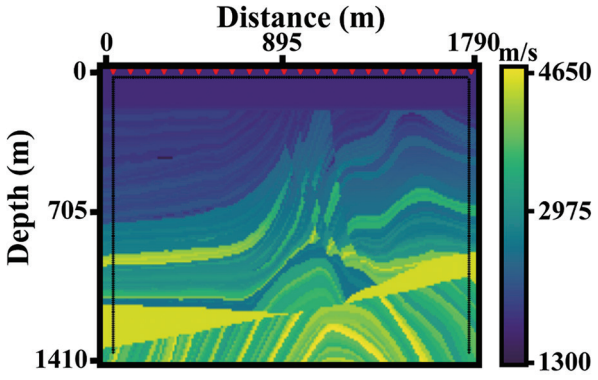

Figure 2. Marmousi2 model

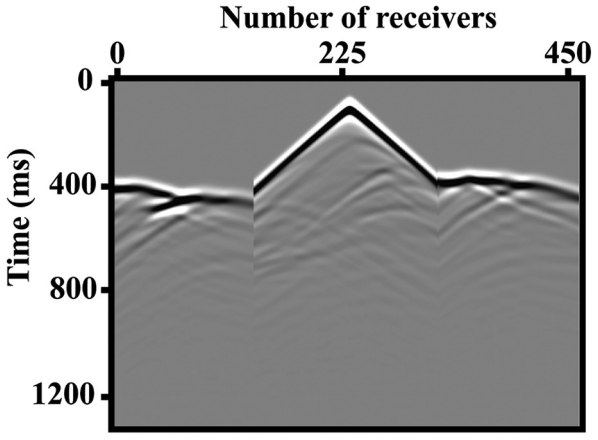

Figure 3. Seismic records of the Marmousi2 model

The acquisition geometry is shown in Figure 2, where red inverted triangles indicate source locations and black line denotes receiver positions. This acquisition configuration provides seismic records with richer information content (Figure 3).

The observed seismic records were then fed into AxPCAE, and the corresponding inversion results are shown in Figure 4. Figure 4A presents the result after the first iteration, Figure 4B after 100 iterations, and Figure 4C after 1,000 iterations. These results showed that the proposed method can perform FWI without requiring an explicit initial model. After 100 iterations (Figure 4B), the lateral velocity variations in the shallow and middle parts of the Marmousi2 model were clearly delineated: the continuity and geometry of several major reflectors were well reconstructed, and some high-wavenumber components began to be superimposed on the low-wavenumber background, making reflector boundaries sharper and structural undulations more distinct. At this stage, however, the deeper structures remained relatively smooth and several local features remained insufficiently resolved, indicating that the inversion was in a transition phase from predominantly low-frequency to mid–high-frequency updates, and that the multi-scale inversion had not yet fully converged.

When the number of iterations increased to 1,000 (Figure 4C), the model exhibited richer high-frequency details and finer lateral heterogeneities. Small-scale velocity fluctuations within the shallow sedimentary layers were clearly resolved, the undulations and bending of interfaces in structurally complex zones were well recovered, and the main deep reflectors together with underlying structures showed high resolution and good phase continuity. Meanwhile, the recovered velocity values remained within a physically plausible range, without obvious unphysical high-/low-velocity anomalies or strong striping artifacts. These observations demonstrate that, by incorporating axial self-attention and physical constraints, AxPCAE-FWI can progressively inject high-frequency information while preserving an accurate low-frequency background, thereby achieving consistent multi-scale

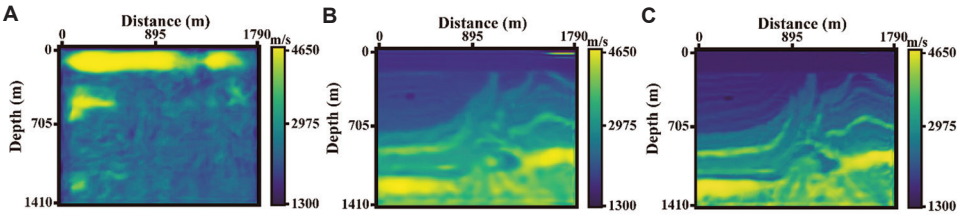

Figure 4. Physics-constrained autoencoder-based full-waveform inversion with axial self-attention inversion results for the Marmousi2 model. (A) Inversion result after the first iteration. (B) Inversion result after 100 iterations. (C) Inversion result after 1,000 iterations.

inversion that balances global convergence with high-resolution imaging capability.

Figure 5 presents single-trace velocity comparisons at x = 100 m and x = 600 m in the Marmousi2 model. For both profiles, the predicted curves closely followed the overall velocity gradient and the primary interface locations over the entire depth range, with only slight smoothing and mildly reduced amplitudes observed in structurally complex mid-to-deep sections. This indicates that AxPCAE exhibits strong inversion capability on the Marmousi2 model and can produce accurate inversion results.

Figure 6 compares the inversion results on the Marmousi2 model, where Figure 6A shows the physics-informed autoencoder (PIAE) result after 1,000 iterations and Figure 6B shows the AxPCAE result at the same iteration count. Due to the lack of high-wavenumber information, the PIAE result exhibited discontinuous undulations of shallow layered

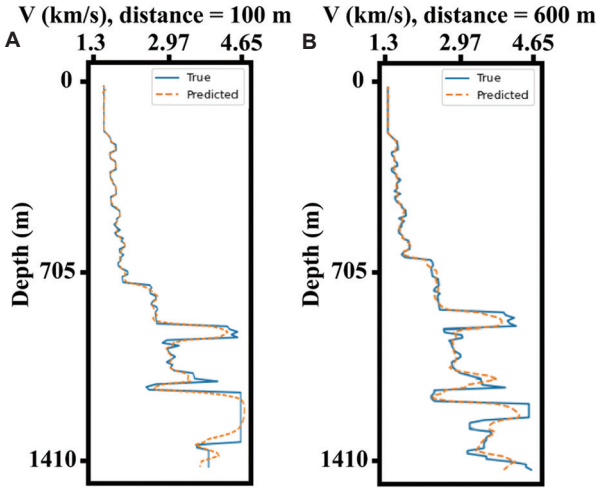

Figure 5. Trace-by-trace comparison between the inversion result of physics-constrained autoencoder-based full-waveform inversion with axial self-attention and the true model. (A) Single-trace comparison at x = 100 m. (B) Single-trace comparison at x = 600 m.

structures, blurred interface boundaries, and locally uneven layer thicknesses. Below the main reflector, extensive low-velocity patches and locally unrealistic velocity anomalies were observed within the high-velocity layer, breaking intra-layer continuity. In contrast, by introducing axial attention, AxPCAE could effectively constrain the focusing of wavefield energy along the true interfaces, thereby enhancing the imaging resolution and geological plausibility of the Marmousi2 model.

3.2. SEG's salt-dome model

We evaluated the performance of AxPCAE in the presence of strong overburden shielding using the SEG salt model (Figure 7). The model comprised 1,760 × 1,510 grid points, with both horizontal and vertical grid spacings set to 10 m. The source was a Ricker wavelet with a dominant frequency of 20 Hz, and the corresponding observed seismic records used as input are shown in Figure 8.

The inversion results obtained using AxPCAE-FWI for the SEG salt model are shown in Figure 9. Figure 9A presents the result after the first iteration, Figure 9B after 500 iterations, and Figure 9C after 1,000 iterations. As the iterations progressed to 500 steps, the overall geometry of the salt body was already well established (Figure 9B): the position and undulating features of the top-salt interface were reasonably recovered, the steeply dipping flanks on both sides of the salt-dome gradually emerged, and the velocity contrast within the overburden and parts of the subsalt region was enhanced, indicating that mid–high-wavenumber information began to be superimposed on the low-wavenumber background. However, a certain degree of subsalt under-illumination remained, and the resolution of some local structures remained limited, suggesting that, under strong velocity contrasts and complex multi-path wave propagation, the subsalt velocity field was still being progressively adjusted.

Examining the final inversion result (Figure 9C), AxPCAE-FWI achieved an accurate and high-resolution

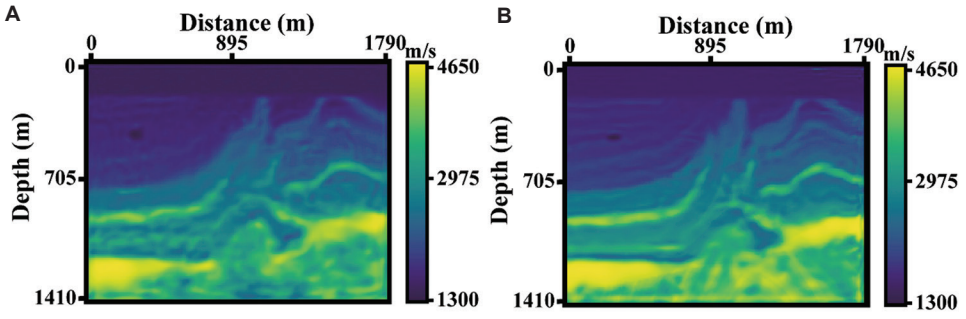

Figure 6. Comparison of inversion results for the Marmousi2 model obtained using (A) physics-informed autoencoder-full-waveform inversion and (B) physics-constrained autoencoder-based full-waveform inversion with axial self-attention after 1,000 iterations.

image of both the salt dome and subsalt structures in the presence of a strong shielding salt body. The top and base of the salt body, as well as the lateral flanks, were sharply delineated, and their geometry closely matched that of the target model. Layering within the overburden was well characterized, while the main subsalt reflectors and the underlying structures exhibited good phase continuity and a realistic description of lateral heterogeneities. The overall velocity distribution was smooth yet not overly smoothed, with no obvious nonphysical high- or low-velocity streaks or numerical oscillations. These results demonstrate that, in the challenging setting of a typical salt-body overburden, AxPCAE-FWI can build a stable low-wavenumber background and gradually compensate for wavefield distortion and illumination loss caused by the salt, thereby enabling effective inversion of subsalt targets and showcasing strong robustness and multi-scale imaging capability.

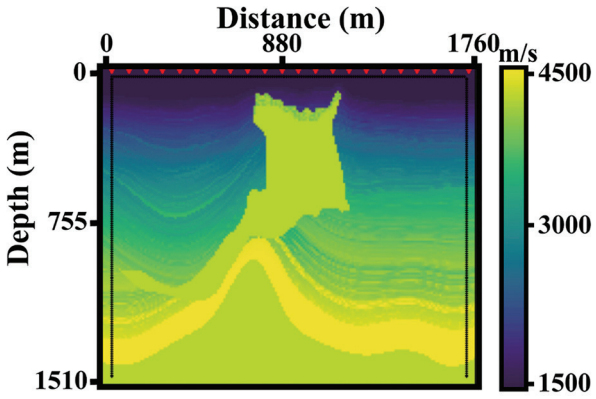

Figure 7. Society of Exploration Geophysicists salt-dome model

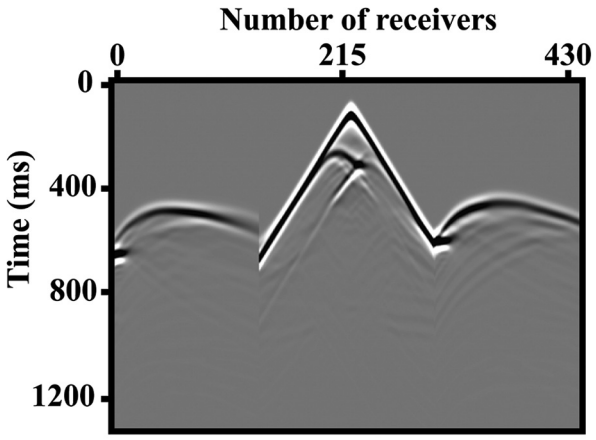

Figure 8. Seismic records of the Society of Exploration Geophysicists salt-dome model

Figure 10 shows single-trace velocity comparisons at x = 100 m (Figure 10A) and x = 600 m (Figure 10B). For the profile at 100 m, the inverted curve closely matched the true model over the entire depth range, with only slight smoothing observed in a few thin layers, indicating that both velocity gradients and fine-scale details could be accurately recovered in well-illuminated regions. In contrast, the profile at 600 m still aligned reasonably well with the true model in the shallow section, but exhibited pronounced discrepancies in the deeper subsalt zone: the sharp high-velocity contrasts were significantly smoothed, and their amplitudes were underestimated, reflecting that illumination loss induced by the overlying salt body continued to limit the inversion resolution in the subsalt region.

Similarly, we compared the proposed method with PIAE, and the results are shown in Figure 11. The PIAE-FWI result was clearly inferior (Figure 11A). The top-salt interface appeared relatively blurred, and the high-velocity anomaly diffused to varying degrees into the overlying sediments. The velocity transition zone between the salt body and the surrounding rocks was overly smoothed, resulting in insufficient constraints on the position and geometry of the interface. The imaging of the salt flanks exhibited trailing and smearing artifacts, and the high-wavenumber energy failed to adequately focus along the true boundaries. In the subsalt region, the velocity distribution contained numerous artifacts, showed poor lateral continuity, and had indistinct boundaries of the main structural units. Overall, the imaging quality was significantly lower than that of AxPCAE-FWI (Figure

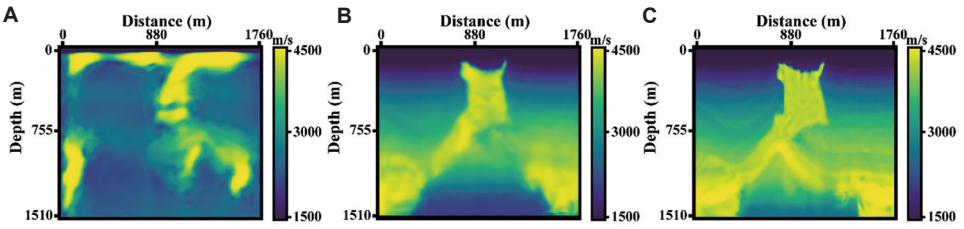

Figure 9. Physics-constrained autoencoder-based full-waveform inversion with axial self-attention results for the Society of Exploration Geophysicists salt-dome model. (A) Inversion result after the first iteration. (B) Inversion result after 500 iterations. (C) Inversion result after 1,000 iterations.

11B). These issues indicate that PIAE, which relied solely on convolutional architectures, faced difficulty capturing the long-range, anisotropic correlations of the wavefield under conditions of strong velocity contrasts and complex multi-path propagation.

Table 1 compares the SSIM and PSNR values of PIAE-FWI and AxPCAE-FWI for the Marmousi2 and SEG salt-dome models. The proposed method achieved significantly better recovery of the overall velocity gradient, interface locations, and fine structural textures than PIAE, resulting in higher structural similarity and numerical accuracy. Meanwhile, in the complex salt-dome environment characterized by strong velocity variations and severe illumination loss, AxPCAE effectively enhanced the imaging quality and structural consistency of the salt body and subsalt structures without increasing overall error, demonstrating stronger robustness and multi-scale imaging capability.

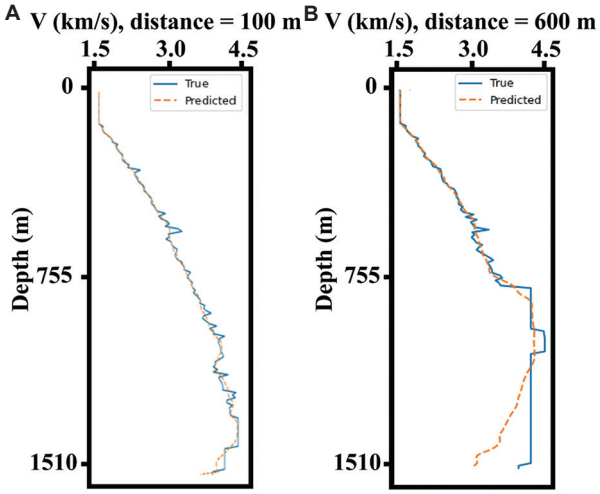

Figure 10. Trace-by-trace comparison between the physics-constrained autoencoder-based full-waveform inversion with axial self-attention result and the true model. (A) Single-trace comparison at x = 100 m. (B) Single-trace comparison at x = 600 m.

Table 2 compares the computational cost of PIAE and AxPCAE on the Marmousi2 and SEG salt-dome models. All experiments were conducted on an NVIDIA RTX 3090 GPU, and the reported runtimes correspond to the total wall-clock time for 1,000 iterations for each method. As shown, AxPCAE exhibited a runtime comparable to PIAE, with a slightly longer iteration time, while achieving superior inversion quality and improved recovery of complex structures. It should be emphasized that the "computational savings" referred to in this work were mainly relative to standard Transformers with full 2D global attention. By adopting axial self-attention, we substantially reduced the computational complexity of attention, thereby enhancing the representation of long-range spatiotemporal correlations at a manageable overall cost.

4. Discussion

We clarify here the key differences between AxPCAE-FWI and several closely related approaches. First, numerous deep learning–assisted FWI methods treat machine learning as a "plug-in" to a conventional update workflow,[32] for example, by generating an initial model or by providing an external regularization term to support a standard model-update loop. In contrast, AxPCAE-FWI embeds a differentiable acoustic wave-equation solver within a unified encoder–decoder framework and uses the data-domain waveform misfit as the primary constraint, thereby governing the optimization using wave-propagation physics. Second, supervised inversion networks typically learn an end-to-end mapping from seismic data to velocity models using paired "seismic–model" labels;[11] AxPCAE-FWI, however, does not require labeled velocity models. It is trained in an unsupervised manner by generating synthetic seismic records using the embedded differentiable solver and updating network parameters via backpropagation of the waveform misfit.[30] Third, physics-informed neural network-based or "neural-solver" variants often replace

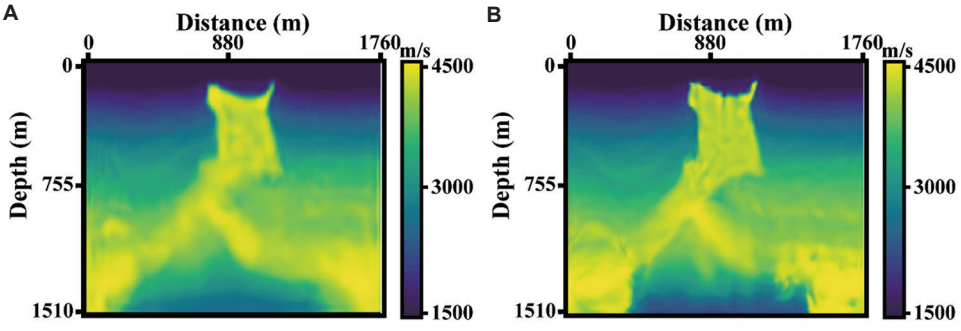

Figure 11. Comparison of inversion results for the SEG salt-dome model obtained using (A) physics-informed autoencoder-full-waveform inversion and (B) physics-constrained autoencoder-based full-waveform inversion with axial self-attention after 1,000 iterations

the numerical partial differential equation solver with a physics-informed network and may perform updates in a low-dimensional latent space while introducing an explicit latent-space regularization term.[29] In contrast, AxPCAE-FWI retains a conventional differentiable simulator and does not introduce explicit latent-space regularization; its stability mainly arises from implicit regularization provided by the latent bottleneck and the physically admissible output mapping, i.e., by limiting model capacity and constraining predictions to a feasible parameter range to mitigate ill-posedness. Finally, although AxPCAE-FWI is conceptually related to inversion methods that re-parameterize the model using a neural network generator,[31] our framework is explicitly data-conditioned: the encoder takes observed seismic records as input, and axial self-attention performs separable modeling along the temporal and offset axes to capture long-range correlations on the time–offset plane. This design enables effective representation of complex wave phenomena, such as multiples, converted waves, and far-offset reflections, while keeping the computational overhead manageable.

Building on these observations, we next investigated the robustness of AxPCAE-FWI under noise interference. Specifically, we contaminated the observed seismic records with substantial random noise and evaluated the corresponding inversion performance. We used the Marmousi2 and SEG salt-dome models, with model dimensions and acquisition geometry identical to those described above. The noisy observed data for the Marmousi2 model are shown in Figure 12, and the corresponding inversion results are presented in Figure 13.

Table 1. Similarity comparison of inversion results obtained using PIAE-FWI and AxPCAE from different models

Models	SSIM		PSNR (dB)	
	PIAE-FWI	AxPCAE-FWI	PIAE-FWI	AxPCAE-FWI
Marmousi2	0.545352	0.809973	22.87	28.07
SEG salt-dome	0.676215	0.728408	19.07	19.05

Abbreviations: AxPCAE-FWI: Physics-constrained autoencoder-based full-waveform inversion with axial self-attention; PIAE-FWI: Physics-informed autoencoder-full-waveform inversion; PSNR: Peak signal-to-noise ratio; SEG: Society of Exploration Geophysicists; SSIM: Structural similarity index.

Table 2. Comparison of computational cost between AxPCAE-FWI and PIAE-FWI

Models	Average time per iteration (s)		Total inversion time (h)	
	PIAE-FWI	AxPCAE-FWI	PIAE-FWI	AxPCAE-FWI
Marmousi2	39.98	69.54	11.11	19.32
SEG salt-dome	49.43	72.59	13.73	20.16

Abbreviations: AxPCAE-FWI: Physics-constrained autoencoder-based full-waveform inversion with axial self-attention; PIAE-FWI: Physics-informed autoencoder-full-waveform inversion; SEG: Society of Exploration Geophysicists.

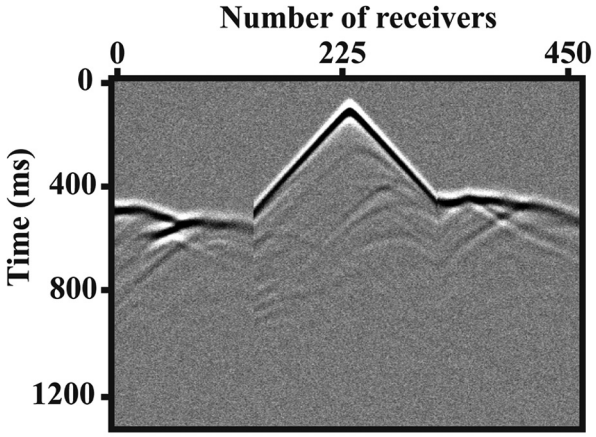

Figure 12. Seismic records for the Marmousi2 model contaminated with random noise

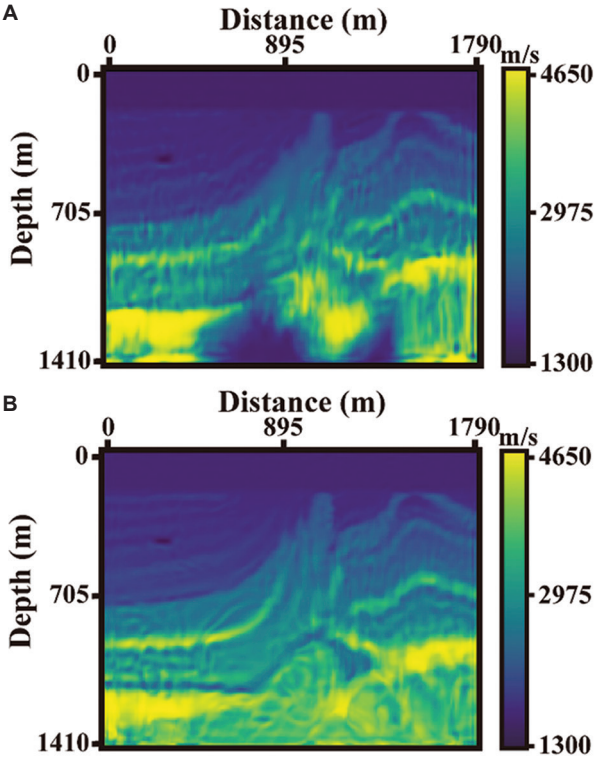

Figure 13. Comparison of Marmousi2 model inversion results under random noise after 1,000 iterations using (A) physics-informed autoencoder-full-waveform inversion and (B) physics-constrained autoencoder-based full-waveform inversion with axial self-attention

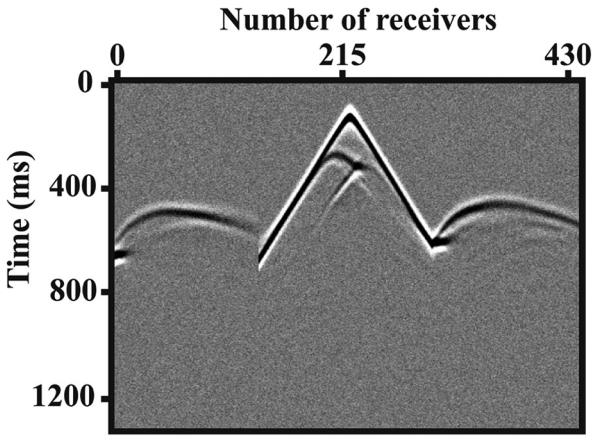

Figure 14. Seismic records for the Society of Exploration Geophysicists salt-dome model contaminated with random noise

In the Marmousi2 experiment with strong random noise, compared to the PIAE result (Figure 13A), the AxPCAE result (Figure 13B) achieved a more robust integration of low-wavenumber information, yielding a more stable macro-velocity trend and traveltime kinematics and mitigating noise-induced background undulation. Meanwhile, high-wavenumber details were selectively reinforced within the axial context, producing more concentrated interface gradients and narrower transition zones, thereby markedly improving the continuity and geometric fidelity of dipping strata and channelized structures. We attributed this to attention weights that adaptively aggregated coherent energy while suppressing incoherent speckle noise, thereby avoiding the bandwidth degradation in conventional results where "high-frequency edge spreading" coexisted with over-smoothing.

The noisy observed records for the SEG salt-dome model are shown in Figure 14, and the corresponding inversion results are presented in Figure 15.

Under strong random noise, the PIAE inversion in Figure 15A exhibited insufficient low-frequency constraints, leading to an unstable macro-velocity trend and traveltime kinematics. High-frequency information was obscured by noise due to amplitude compression; the continuity of subsalt wedge-shaped strata was poor, and speckle-like artifacts were observed. In contrast, the AxPCAE result in Figure 15B was more robust at both low- and high-wavenumber ends: gradients along the salt top and flanks were more concentrated with sharper contours, the base was better localized, subsalt layering was more coherent, and background texture noise was markedly reduced, yielding improved broadband consistency and noise robustness overall.

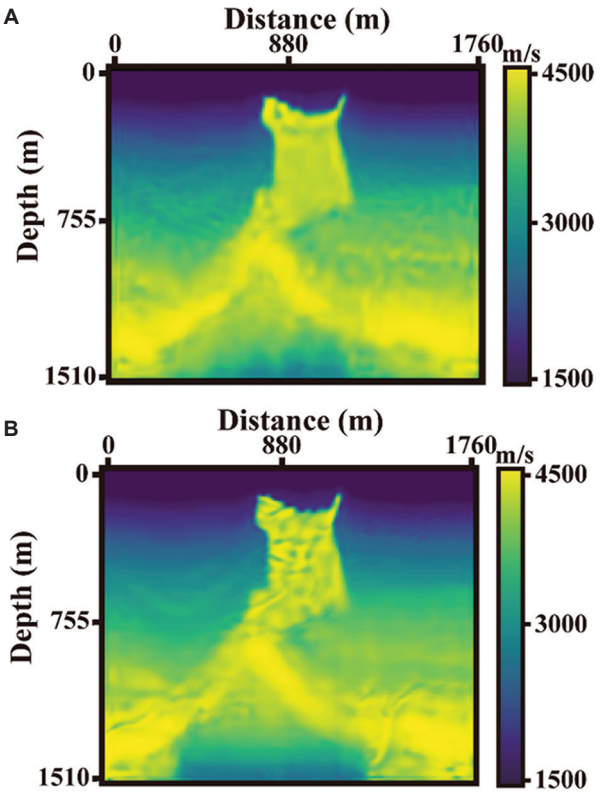

Figure 15. Comparison of SEG salt-dome model inversion results under random noise after 1,000 iterations using (A) physics-informed autoencoder-full-waveform inversion and (B) physics-constrained autoencoder-based full-waveform inversion with axial self-attention

Nevertheless, localized velocity discontinuities are observable within the salt body and its shadow zones, reflecting the anisotropy of axial attention that can amplify both noise and fine details at high frequencies. Overall, relative to PIAE, AxPCAE under high noise improves resolution and boundary delineation.

5. Conclusion

This study proposes AxPCAE-FWI, a physics-constrained autoencoder framework for FWI with axial self-attention. In a unified encoder–decoder framework, a differentiable acoustic wave-equation solver was explicitly embedded, and data-domain waveform fitting was used as the primary constraint, ensuring that the network training process was consistently governed using the wave equation rather than relying solely on empirical features or black-box fitting. In this way, the network automatically learned the nonlinear mapping from observed seismic records to subsurface velocity models. The encoder extracts low-dimensional latent features, while the decoder reconstructs the velocity model under a physically admissible constraint. A Sigmoid function followed by linear rescaling was applied to ensure

that the inverted parameters fell within a prescribed velocity range, thereby suppressing nonphysical anomalies at the implementation level.

In terms of network architecture, AxPCAE-FWI introduces axial multi-head self-attention modules in the encoding stage, where global attention is computed separately along the time and receiver axes. Two one-dimensional global attention operations were used to approximate a single two-dimensional global attention, enabling the method to capture long-range spatiotemporal correlations associated with multiples, converted waves, and far-offset reflections at a substantially reduced computational cost, and to enhance the representational capacity for complex wavefields. In contrast to conventional FWI, which performed gradient-based updates directly in the model space, the proposed approach conducted "implicit inversion" in the latent space and achieved stable convergence without requiring an explicit initial model or ultra-low-frequency data, thereby alleviating cycle skipping from a mechanistic perspective. Numerical experiments on the Marmousi2 and SEG salt-dome models show that, under complex structural settings and strong shielding from high-velocity bodies, AxPCAE-FWI achieves higher structural similarity and greater geological plausibility than conventional PIAE-FWI. It also demonstrates greater robustness to noise, confirming the effectiveness and practical promise of coupling axial self-attention with physical constraints for FWI.

Acknowledgments

None.

Funding

This study is supported by the National Natural Science Foundation of China (42574163), CNPC Science and Technology Innovation Foundation (Grant No. 2024DQ02-0135), and Taishan Scholars Program (Grant No. tsqn202312117).

Conflict of interest

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

Author contributions

Conceptualization: Yunbo Niu, Yingming Qu, Zhenchun Li
Formal analysis: Yunbo Niu
Investigation: Yunbo Niu
Methodology: Yunbo Niu, Yingming Qu
Writing–original draft: Yunbo Niu
Writing–review & editing: Yunbo Niu, Yingming Qu, Zhenchun Li

Availability of data

All data are available upon reasonable request from the corresponding authors.

References

1. Lailly P. The seismic inverse problem as a sequence of before stack migrations. In: Bednar JB, Redner R, Robinson E, Weglein A, editors. *Conference on Inverse Scattering: Theory and Application. Society for Industrial and Applied Mathematics*; 1983. p. 206-220.

2. Tarantola A, Valette B. Generalized nonlinear inverse problems solved using the least squares criterion. *Rev Geophys*. 1982;20(2):219-232.

   doi: 10.1029/RG020i002p00219

3. Tarantola A. Inversion of seismic reflection data in the acoustic approximation. *Geophysics*. 1984;49(8):1259-1266.

   doi: 10.1190/1.1441754

4. Virieux J, Operto S. An overview of full-waveform inversion in exploration geophysics. *Geophysics*. 2009;74(6):WCC1-WCC26.

   doi: 10.1190/1.3238367

5. Mora P. Inversion=migration+tomography. *Geophysics*. 1989;54(12):1575-1586.

   doi: 10.1190/1.1442625

6. Xu S, Wang D, Chen F, Lambaré G, Zhang Y. Inversion of reflected seismic waves. In: *SEG Technical Program Expanded Abstracts 2012*. Las Vegas, NV. Society of Exploration Geophysicists; 2012. p. 1-7.

   doi: 10.1190/segam2012-1473.1

7. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436-444.

   doi: 10.1038/nature14539

8. Kim Y, Nakata N. Geophysical inversion versus machine learning in inverse problems. *Lead Edge*. 2018;37(12):894-901.

   doi: 10.1190/tle37120894.1

9. Wu Y, McMechan GA, Wang Y. Adaptive feedback convolutional neural network-based high-resolution reflection-waveform inversion. *J Geophys Res Solid Earth*. 2022;127(6):e2022JB024138.

   doi: 10.1029/2022JB024138

10. Wang W, Yang F, Ma J. Velocity model building with a modified fully convolutional network. In: *SEG Technical Program Expanded Abstracts 2018; SEG International Exposition and 88th Annual Meeting*. Anaheim, CA: Society of Exploration Geophysicists; 2018. p. 2086-2090.

    doi: 10.1190/segam2018-2997566.1

11. Wu Y, Lin Y, Zhou Z. InversionNet: Accurate and efficient

seismic-waveform inversion with convolutional neural networks. In: *SEG Technical Program Expanded Abstracts 2018; SEG International Exposition and 88th Annual Meeting*. Anaheim, CA. Society of Exploration Geophysicists; 2018. p. 2096-2100.

doi: 10.1190/segam2018-2998603.1

12. Park MJ, Sacchi MD. Automatic velocity analysis using convolutional neural network and transfer learning. *Geophysics*. 2020;85(1):V33-V43.

doi: 10.1190/geo2018-0870.1

13. Li S, Liu B, Ren Y, Chen Y. Deep-learning inversion of seismic data. *IEEE Trans Geosci Remote Sens*. 2020;58(3):2135-2149.

doi: 10.1109/TGRS.2019.2953473

14. Wang W, Ma J. Velocity model building in a crosswell acquisition geometry with image-trained artificial neural networks. *Geophysics*. 2020;85(2):U31-U46.

doi: 10.1190/geo2018-0591.1

15. Lu J, Wu C, Zhang H, Qu Y, Xu Q, Zhu M. *Phi-Net: An Aggregation Network for Seismic Velocity Model Building. SSRN*; 2024. Available from: https://ssrn.com/abstract=5076729 [Last accessed on 2025 Nov 29].

doi: 10.2139/ssrn.5076729

16. Araya-Polo M, Jennings J, Adler A, Dahlke T. Deep-learning tomography. *Lead Edge*. 2018;37(1):58-66.

doi: 10.1190/tle37010058.1

17. Song C, Geng H, Wang Y, *et al.* Simultaneous P- and S-wave seismic traveltime tomography using physics-informed neural networks. *Geophys Prospect*. 2025;73(6):e70034.

doi: 10.1111/1365-2478.70034

18. Xue Z, Wang Y, Wu X, Ma J. Multi-geophysical information neural network for seismic tomography. *Geophysics*. 2025;90(3): R89-R100.

doi: 10.1190/geo2024-0039.1

19. Zhang X, Wang Y, Zhang H. Rapid probabilistic seismic tomography using graph mixture density networks. *J Geophys Res Solid Earth*. 2025;130(8):e2025JB031129.

doi: 10.1029/2025JB031129

20. Das V, Pollack A, Wollner U, Mukerji T. Convolutional neural network for seismic impedance inversion. *Geophysics*. 2019;84(6):R869-R880.

doi: 10.1190/geo2018-0838.1

21. Wu B, Meng D, Wang L, Liu N, Wang Y. Seismic impedance inversion using fully convolutional residual network and transfer learning. *IEEE Geosci Remote Sens Lett*. 2020;17(12):2140-2144.

doi: 10.1109/LGRS.2019.2963106

22. Taufik MH, Wang F, Alkhalifah T. Learned regularizations for multi-parameter elastic full waveform inversion using diffusion models. *J Geophys Res Machine Learn Comput*. 2024;1(1):e2024JH000125.

doi: 10.1029/2024JH000125

23. Kazei V, Ovcharenko O, Plotnitskii P, Peter D, Zhang X, Alkhalifah T. Mapping full seismic waveforms to vertical velocity profiles by deep learning. *Geophysics*. 2021;86(5):R711-R721.

doi: 10.1190/geo2019-0473.1

24. Yang F, Ma J. FWIGAN: Full-waveform inversion via a physics-informed generative adversarial network. *J Geophys Res Solid Earth*. 2023;128(4):e2022JB025493.

doi: 10.1029/2022JB025493

25. Zhang Z, Lin Y. Data-driven seismic waveform inversion: A study on the robustness and generalization. *IEEE Trans Geosci Remote Sens*. 2020;58(10):6900-6913.

doi: 10.1109/TGRS.2020.2977635

26. Yang Y, Engquist B, Sun J, *et al.* Application of optimal transport and the quadratic Wasserstein metric to full-waveform inversion. *Geophysics*. 2018;83(1):R43-R62.

doi: 10.1190/geo2016-0663.1

27. Vantassel JP, Kumar K, Cox BR. Using convolutional neural networks to develop starting models for near-surface two-dimensional full waveform inversion. *Geophys J Int*. 2022;231(1):72-90.

doi: 10.1093/gji/ggac179

28. Yang F, Ma J. Deep-learning inversion: A next-generation seismic velocity model building method. *Geophysics*. 2019;84(4):R583-R599.

doi: 10.1190/geo2018-0249.1

29. Taufik MH, Huang X, Alkhalifah T. Latent representation learning in physics-informed neural networks for full waveform inversion. *Earth Space Sci*. 2025;12(9):e2024EA004107.

doi: 10.1029/2024EA004107

30. Wu Y, McMechan GA. Parametric convolutional neural network-domain full-waveform inversion. *Geophysics*. 2019;84(6):R881-R896.

doi: 10.1190/geo2018-0224.1

31. Zhu W, Xu K, Darve E, *et al.* Integrating deep neural networks with full-waveform inversion: Reparameterization, regularization, and uncertainty quantification. *Geophysics*. 2022;87(1):R93-R109.

doi: 10.1190/geo2020-0933.1

32. Wang Y, Jiang B, Wei Z, *et al.* Deep velocity generator:

A plug-in network for FWI enhancement. *IEEE Trans Geosci Remote Sens*. 2023;61:1-17.

doi: 10.1109/TGRS.2023.3247880

33. He Q, Wang Y. Reparameterized full-waveform inversion using deep neural networks. *Geophysics*. 2021;86(1):V1-V13.

doi: 10.1190/geo2019-0382.1

34. Tromp J, Tape C, Liu Q. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels. *Geophys J Int*. 2005;160(1):195-216.

doi: 10.1111/j.1365-246X.2004.02453.x