Petroleum Science 23 (2026) 1890–1907

Contents lists available at ScienceDirect

Petroleum Science

journal homepage: www.keaipublishing.com/en/journals/petroleum-science

Original Paper

Deep reparameterization for full waveform inversion: Architecture benchmarking, robust inversion, and multiphysics extension

Feng Liu [a,b], Ya-Xing Li [c], Rui Su [b,*], Jian-Ping Huang [c,d], Lei Bai [b]

[a] School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, 200240, China
[b] Shanghai Artificial Intelligence Laboratory, Shanghai, 200232, China
[c] Key Laboratory of Earth Exploration and Information Technology of Ministry of Education, Chengdu University of Technology, Chengdu, 610059, Sichuan, China
[d] State Key Laboratory of Deep Oil and Gas, School of Geosciences, China University of Petroleum (East China), Qingdao, 266580, Shandong, China

ARTICLE INFO

Article history:
Received 29 June 2025
Received in revised form
26 October 2025
Accepted 15 December 2025
Available online 19 December 2025

Edited by Meng-Jiao Zhou

Keywords:
Deep reparameterization
Full waveform inversion
Multiparameter inversion
Sparse acquisition
Network architecture search

ABSTRACT

Full waveform inversion (FWI) is a high-resolution subsurface imaging technique, but its effectiveness is limited by challenges such as noise contamination, sparse acquisition, and artifacts from multiparameter coupling. To address these limitations, this study develops a deep reparameterized FWI (DR-FWI) framework, in which subsurface parameters are represented by a deep neural network. Instead of directly optimizing the parameters, DR-FWI optimizes the network weights to reconstruct them, thereby embedding network priors and facilitating optimization. To provide guidelines for the design and usage of DR-FWI, we benchmark two initial model embedding strategies: one involves pretraining the network to generate predefined initial models (pretraining-based), and the other directly adds the network outputs to the initial models, along with three representative architectures (UNet, CNN, MLP). Extensive ablation experiments show that combining CNN with pretraining-based initialization significantly enhances inversion accuracy, offering valuable insights into network design. To further understand the mechanism of DR-FWI, spectral bias analysis reveals that the network first captures low-wavenumber features and progressively reconstructs high-wavenumber details. This learning pattern supports adaptive multi-scale inversion and provides a physically interpretable view of the inversion process. Notably, the robustness of DR-FWI is validated under various noise levels and sparse acquisition scenarios, where its strong performance with limited shots and receivers demonstrates reduced reliance on dense observational data. Additionally, a "backbone-branch" structure is proposed to extend DR-FWI to multiparameter inversion, and its efficacy in mitigating cross-parameter interference is validated on a synthetic anomaly model and the Marmousi2 model. These results suggest a promising direction for joint inversion involving multiple parameters or multiphysics.

1. Introduction

Full waveform inversion (FWI) is a wave-equation-based high-resolution subsurface parameter estimation method that minimizes the residuals between simulated and observed seismic wavefields through iterative optimization (Lailly, 1983; Tarantola, 1984). Compared to conventional approaches such as ray-based tomography and travel-time imaging, FWI demonstrates superior capabilities by synergistically utilizing full waveform features including amplitude, phase, and travel-time information (Virieux and Operto, 2009; Tromp, 2019). This comprehensive data exploitation enables exceptional performance in complex geological structure inversion (e.g., salt domes (Kalita et al., 2019), fault systems (Brossier et al., 2009)) and resource exploration applications (particularly in hydrocarbon reservoirs (Routh et al., 2017) and geothermal fields (Schmelzbach et al., 2016)). The method's essence lies in its fundamental departure from the geometric approximations inherent in ray theory, instead employing numerical solutions of the wave equation to achieve

* Corresponding author.
E-mail address: surui@pjlab.org.cn (R. Su).
Peer review under the responsibility of China University of Petroleum (Beijing).

https://doi.org/10.1016/j.petsci.2025.12.027

high-fidelity parameter reconstruction. This capability provides critical petrophysical constraints for engineering decision-making process in exploration geophysics (Virieux and Operto, 2009).

Despite its theoretical potential, FWI faces substantial challenges in practical implementations. First, the highly nonlinear nature of the objective function makes the inversion process susceptible to entrapment in local minima (Gauthier et al., 1986; Luo and Schuster, 1991). Gradient-based algorithms often exhibit poor convergence robustness when the initial model substantially deviates from the true subsurface properties (Métivier et al., 2014; Xie et al., 2024). Second, the ill-posedness of the inverse problem manifests as solution non-uniqueness, where distinct parameter combinations can produce nearly identical wavefield responses, thereby amplifying solution uncertainty (Tarantola, 2005). Moreover, in multiparameter joint inversions (e.g., P-wave velocity (v_p), S-wave velocity (v_s), and density (ρ)), crosstalk effects between parameters can introduce spurious anomalies, critically undermining the physical reliability of inverted results (Operto et al., 2013; Yang et al., 2014; Chen et al., 2024b). These intertwined issues collectively dictate that FWI's practical efficacy remains heavily contingent upon initial model accuracy, observational system completeness, and the design of appropriate regularization strategy.

Recent advances in deep learning have provided new pathways to overcome traditional FWI limitations (Yu and Ma, 2021; Lin et al., 2023). Current methodologies for subsurface inversion can be broadly categorized into two paradigms: data-driven approaches that employ end-to-end networks to directly map observed data to subsurface medium parameters (Wu and Lin, 2020; Zhang and Lin, 2020; Saadat et al., 2024), yet often suffer from a strong dependency on annotated datasets and poor physical interpretability; physics-driven methods embed wave-equation constraints into hybrid inversion frameworks (Rasht-Behesht et al., 2022; Ren et al., 2020), offering greater resilience under sparse data conditions and improved consistency with physical laws. In addition, hybrid strategies that integrate physical constraints into deep learning have also been explored to enhance inversion performance (Taufik et al., 2024; Chen et al., 2024a, 2025). As an example, Fig. 1 illustrates a subset of research on data-driven and physics-driven FWI, including approaches based on generative models and end-to-end mapping for the former, and physics-informed neural networks (PINNs) as well as deep reparameterization techniques for the latter.

As a pivotal branch of the physics-driven paradigm, deep reparameterization reformulates inversion parameters as low-dimensional latent variables via neural network-based implicit representations. This reformulation not only mitigates the nonlinearity of the objective function but also introduces learnable regularization into the inversion process (He and Wang, 2021; Zhu et al., 2022; Sun et al., 2023a; Liu et al., 2025). Prior studies have successfully applied this strategy to construct prior constraints, reduce dependence on initial models, and mitigate multiparameter crosstalk. To enhance regularization, Wu and McMechan (2019) and He and Wang (2021) implemented multi-layer convolutional neural networks (CNNs) for velocity model parameterization, employing a two-phase training scheme: Phase I pretrains the network to map random vectors to initial velocity models, while Phase II integrates adjoint-state methods for physics-constrained optimization. Their synthetic experiments demonstrated improved peak signal-to-noise ratio compared to conventional FWI. Zhu et al. (2022) adopted shallower CNNs to generate velocity perturbations, which were superimposed on background models as inputs to the FWI process. This architecture exhibited enhanced noise robustness when tested on data contaminated with Gaussian noise. For addressing initial model dependency, Sun et al. (2023a) proposed an implicit FWI framework using fully-connected networks (FNNs) or multi-layer perceptrons (MLPs) to directly map spatial coordinates to subsurface velocity values, thereby removing the reliance on initial velocity models. Notably, Sun et al. (2021) pioneered a discriminative autoencoder for data-physics co-driven FWI, which was subsequently extended by Wang et al. (2023) and Taufik et al. (2024) through diffusion model-based generative architectures. While these hybrid methods alleviate the dependence on initial velocities, they still require massive paired velocity-seismic datasets for effective network training. In multiparameter inversion, Dhara and Sen (2022) developed an autoencoder-based collaborative framework, in which an encoder compresses multi-shot seismic records into latent representations, and parallel decoders generate distinct subsurface parameter perturbations for superposition on background models.

Despite recent progress, several critical limitations hinder the broader applicability of deep reparameterization-based FWI. First, the lack of theoretical guidance on network design and hyperparameter tuning often leads to heuristic architectures with limited transferability. Second, most existing studies focus predominantly on noise robustness, with insufficient investigation into other practical scenarios such as sparse acquisitions. Third, multiparameter inversion remains underexplored; existing approaches rely on case-specific designs for limited parameter sets, lacking a unified and scalable framework applicable to diverse physical properties. Moreover, the absence of open-source implementations restricts reproducibility and further development.

To address these challenges, this paper proposes a reference-guided deep reparameterization FWI framework (DR-FWI). The core innovation lies in the integration of prior geological knowledge via pretraining schemes and the design of multiparameter adaptive architectures. The main contributions of this work are summarized as follows.

1) Rethinking network design for prior-guided deep reparameterization: This work introduces the deep reparameterization-based full waveform inversion (DR-FWI) framework, leveraging deep neural networks to reparameterize subsurface models. Through systematic benchmarking, we demonstrate that shallow CNNs, when combined with a pretraining-based initial model embedding strategy, consistently outperform deeper and more complex architectures in terms of both accuracy and convergence. The prior-guided reparameterization benefits from implicit structural priors, and our spectral bias analysis reveals a progressive learning behavior from low to high frequencies, offering interpretable theoretical insights into the effectiveness of such architectures for inversion tasks.
2) Extending FWI applicability to data-constrained and noisy scenarios: DR-FWI achieves state-of-the-art inversion accuracy under extremely sparse acquisition and high noise conditions, significantly outperforming conventional FWI. This robustness greatly broadens the applicability of full waveform inversion, enabling its deployment in data-limited or logistically constrained environments where traditional methods fail.
3) A scalable and generalizable framework for multiparameter joint inversion: This study proposes a novel backbone-branch architecture for multiparameter joint inversion, explicitly separating shared structural features from parameter-specific components. This design minimizes parameter crosstalk and ensures consistent inversion across multiple physical domains. The framework demonstrates robust scalability and generalization in complex geological scenarios, offering a unified and practical approach to multi-physics inversion using deep learning.

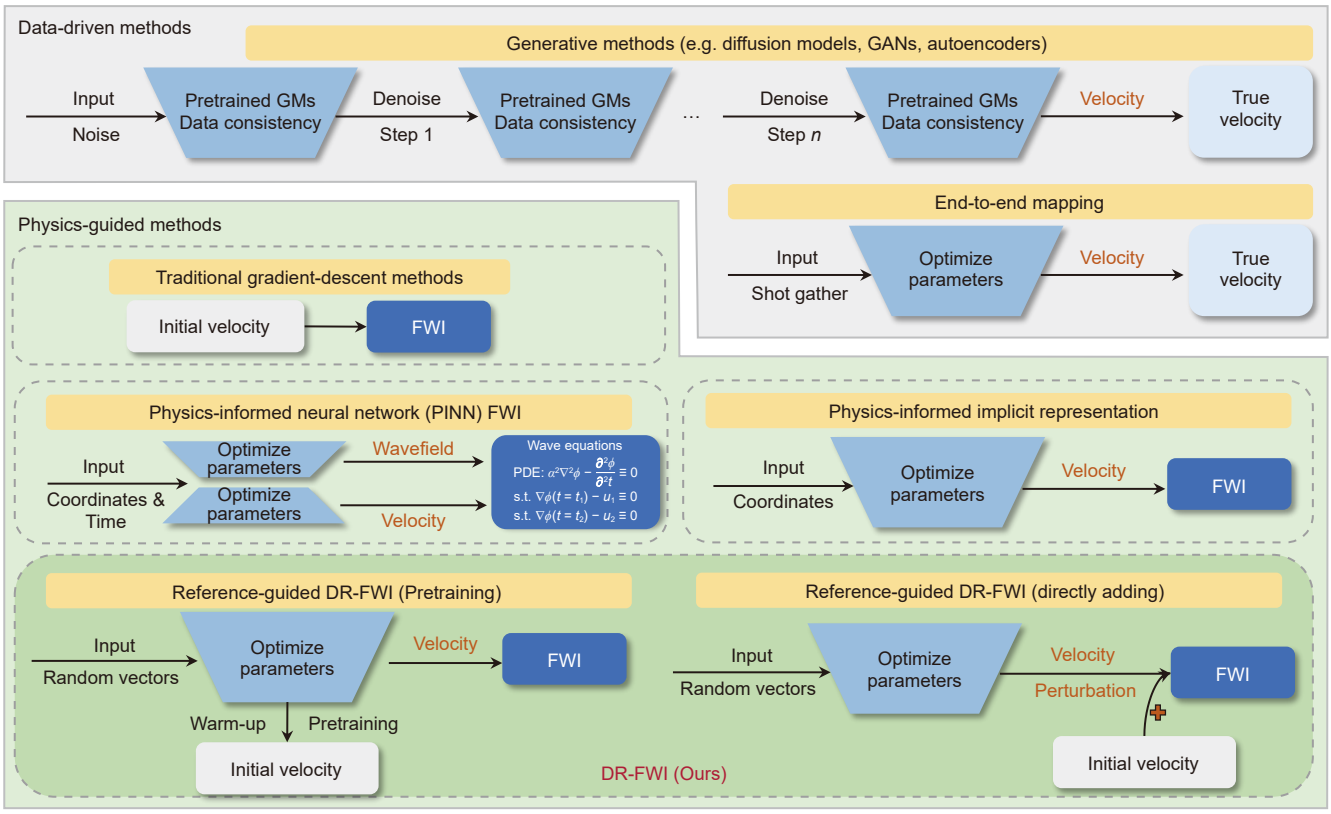

Fig. 1. Comparison of data-driven methods (e.g., generative model-based, end-to-end supervised learning) and physics-driven methods (e.g., traditional FWI, PINNs-based FWI, and implicit representation-based IFWI), along with the two forms of reference-guided DR-FWI proposed in this study: one involving pretraining the network to generate predefined initial models, and the other directly adding network outputs to the initial model.

This study advances physics-constrained deep learning inversion along three critical dimensions: principled network design informed by implicit priors, robust generalization under data-limited conditions, and extensibility to multiparameter and multi-physics settings. DR-FWI integrates seamlessly with conventional workflows and provides a theoretically grounded, practically effective solution for challenging geophysical inversion tasks.

2. Methodology

2.1. Physical principles of FWI

Traditional FWI is an entirely physics-driven inversion approach that employs gradient-based optimization methods to iteratively minimize the mismatch between observed and simulated seismic data. The underlying physical relationship between the subsurface properties and the recorded waveforms can be expressed in a general form as:

$$\mathbf{d} = F(\mathbf{m}), \tag{1}$$

where F represents the forward operator, $\mathbf{d}$ denotes the synthetic seismic data, and $\mathbf{m}$ describes the subsurface properties. Specifically, in an acoustic medium, $\mathbf{m}$ typically consists of the P-wave velocity (v_p) and density (ρ); in isotropic elastic medium, $\mathbf{m}$ is extended to include the P-wave velocity, S-wave velocity (v_s), and density; and in anisotropic elastic medium, $\mathbf{m}$ further incorporates anisotropic parameters such as ϵ, δ, and γ. A detailed implementation of different wave equation formulations can be found in Liu et al. (2025). All subsequent experiments in this study are conducted within the flexible ADFWI framework (Liu, 2024).

The classical FWI with a general regularization term can be formulated as a constrained optimization problem:

$$\mathbf{m}^* = \arg\min_{\mathbf{m}} \{\mathscr{D}(\mathbf{d}_{\text{obs}}, \mathbf{d}_{\text{cal}}(\mathbf{m})) + \lambda \mathscr{R}(\mathbf{m})\}, \tag{2}$$

where $\mathbf{m}^*$ represents the final inverted subsurface model, $\mathscr{D}$ measures the misfit between observed data $\mathbf{d}_{\text{obs}}$ and synthetic data $\mathbf{d}_{\text{cal}}(\mathbf{m})$, $\mathscr{R}(\mathbf{m})$ is a regularization term, and λ is the corresponding weight. The misfit function can be defined using various norms, such as the L_2-norm, global cross-correlation, or alternative metrics such as the envelope misfit (Lailly, 1983; Bozdağ et al., 2011; Choi and Alkhalifah, 2012).

The commonly used optimization strategy in FWI is gradient descent-based optimization, which iteratively refines an initial subsurface model ($\mathbf{m}_0$) until the synthetic data sufficiently match the observed data. Specifically, at the $(k+1)$th iteration, the model parameters are updated as:

$$\mathbf{m}_{k+1} = \mathbf{m}_k - \alpha_k \frac{\partial \mathscr{L}}{\partial \mathbf{m}_k}, \tag{3}$$

where $\mathscr{L}$ denotes the objective function to be minimized, and $\partial\mathscr{L}/\partial\mathbf{m}_k$ denotes the descent direction, which can be computed using the adjoint-state method (Fichtner et al., 2006; Liu and Tromp, 2006) or automatic differentiation (Sambridge et al., 2007; Liu et al., 2025), and α_k is the step length.

2.2. Deep reparameterization-based FWI

Originally developed for image reconstruction, the deep image prior (DIP) was first introduced by Lempitsky et al. (2018). Their work demonstrated that the architecture of a generator network

alone can capture a substantial amount of low-level image statistics, even prior to any learning process. Inspired by this insight, DIP has been successfully adapted to FWI by reparameterizing subsurface models via a generator network to capture hidden geostatistical features (He and Wang, 2021; Zhu et al., 2022; Sun et al., 2023b; Wang et al., 2023; Liu et al., 2025). In this framework (Fig. 2), the mapping from a fixed random input $\mathbf{I}_r$ to the velocity model $\mathbf{m}$ is expressed as

$$\mathbf{m} = \Phi_\theta(\mathbf{I}_r), \tag{4}$$

where θ denotes the trainable parameters of the generator $\Phi_\theta(\cdot)$. The random input is sampled once (e.g., $\mathbf{I}_r \sim \mathcal{N}(0, \mathbf{I})$) and kept fixed, while θ is optimized. The form of $\mathbf{I}_r$ depends on the architecture: a 1D vector for MLPs and CNNs (e.g., $\mathbf{I}_r \in \mathbb{R}^c$), and a 2D random field for UNet (i.e., $\mathbf{I}_r \in \mathbb{R}^{n_z \times n_x}$). Regardless of its input dimension, Φ_θ uses upsampling, convolution, and reshaping to ensure that the output matches the velocity model dimension.

With this reparameterization, the FWI objective in Eq. (2) is reformulated as

$$\theta^* = \arg\min_{\mathbf{m}} \{\mathcal{D}(\mathbf{d}_{obs}, \mathbf{d}_{cal}(\Phi_\theta(\mathbf{I}_r))) + \lambda \mathcal{R}(\Phi_\theta(\mathbf{I}_r))\}, \tag{5}$$

and the corresponding estimated model is $\mathbf{m}^* = \Phi_{\theta^*}(\mathbf{I}_r)$. This reformulation shifts the optimization from the model space $\mathbf{m}$ to the parameter space θ of the generator network, thereby altering how gradients are computed and propagated. In particular, the gradients of the objective function are backpropagated through the network via the chain rule, which links the conventional model-space gradient ($\partial\mathcal{L}/\partial\mathbf{m}$) to the parameter-space gradient ($\partial\mathbf{m}/\partial\theta$). Formally, this relationship is expressed as

$$\frac{\partial \mathcal{L}}{\partial \theta} = \frac{\partial \mathcal{L}}{\partial \mathbf{m}} \cdot \frac{\partial \mathbf{m}}{\partial \theta} = \frac{\partial \mathcal{L}}{\partial \Phi_\theta(\mathbf{I}_r)} \cdot \frac{\partial \Phi_\theta(\mathbf{I}_r)}{\partial \theta}, \tag{6}$$

where $\partial\Phi_\theta(\mathbf{I}_r)/\partial\theta$ denotes the Jacobian matrix of the network with respect to θ, which can be computed using automatic differentiation. The term $\partial\mathcal{L}/\partial\Phi_\theta(\mathbf{I}_r)$ corresponds to the gradient of the FWI objective with respect to $\mathbf{m}$, obtained via adjoint-state methods or automatic differentiation. Consequently, the iterative update is performed in the parameter space as

$$\theta_{k+1} = \theta_k - \alpha_k \frac{\partial \mathcal{L}}{\partial \theta_k}, \tag{7}$$

where α_k denotes the learning rate (equivalent to the step length in conventional FWI) at iteration k.

This formulation provides a rigorous foundation for deep reparameterization-based FWI. By replacing the conventional sensitivity matrix with the neural Jacobian, it embeds implicit priors and nonlinearity into gradient computation. Practical implementation remains challenging due to the strong nonlinearity of wave physics and the complex interaction between network parameterization and physical modeling. Consequently, systematic exploration of reparameterized network architectures and the mechanisms of this implicit regularization remains largely unexplored. Moreover, in many practical scenarios, a preliminary velocity model is available (e.g., a smoothed result from traveltime tomography), making it crucial to develop strategies that effectively leverage such prior information to accelerate inversion and mitigate the risk of convergence to local minima (Zhao et al., 2020).

2.3. Reference-guided deep reparameterization

Leveraging prior information, the reference-guided DIP approach replaces the random input of the vanilla DIP with a given reference (Zhao et al., 2020). Although the original reference-guided DIP proposed by Zhao et al. (2020) demonstrated enhanced performance compared to vanilla DIP by incorporating an initial solution closer to the final result, the network is still randomly initialized and must be trained from scratch. Consequently, this approach typically requires a large number of iterations (often thousands), and since that each iteration involves a full

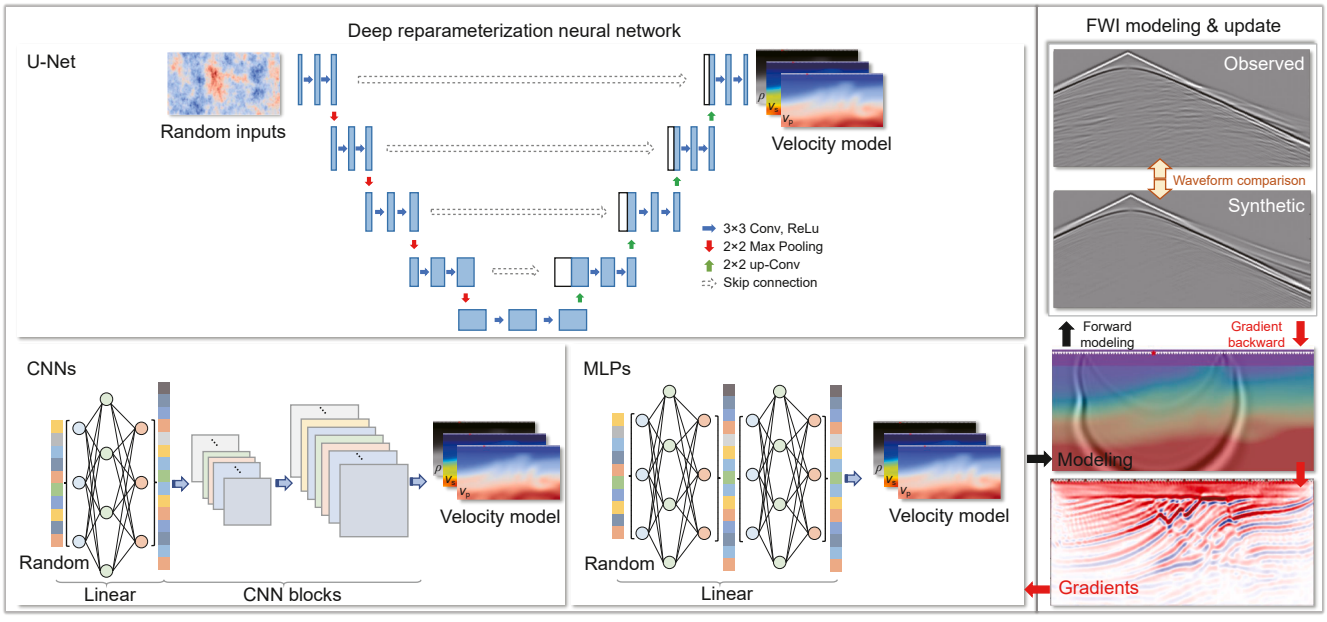

Fig. 2. Workflow of deep reparameterization-based full waveform inversion. The left panel illustrates three types of reparameterization neural networks (UNet, CNNs, and MLPs), which take random vectors or matrices as input and generate the corresponding subsurface models (e.g., v_p, v_s, and ρ). The right panel shows the FWI process, including forward modeling based on the generated velocity model, misfit computation between observed and synthetic waveforms, and gradient calculation using either the adjoint-state method or automatic differentiation.

FWI calculation, the overall computational cost remains substantial.

To address these issues and further improve inversion results, we propose and compare several strategies for integrating the reference (initial) velocity model into the deep reparameterization framework.

1) Direct input of the reference model (Zhao et al., 2020): The reference subsurface model $\mathbf{m}_0$ is directly used as the input to the generator neural network, and the reparameterization subsurface model is expressed as:

$$\mathbf{m} = \Phi_\theta(\mathbf{m}_0). \tag{8}$$

2) Perturbation learning (Zhu et al., 2022): The generator neural network takes random noise as input and learns to predict the perturbation relative to the reference subsurface model. The velocity model used in the FWI workflow is obtained by adding the network output to the reference velocity model

$$\mathbf{m} = \Phi_\theta(\mathbf{I}_\mathrm{r}) + \mathbf{m}_0, \tag{9}$$

and the estimated model can be represented as $\mathbf{m}^* = \Phi_{\theta^*}(\mathbf{I}_\mathrm{r}) + \mathbf{m}_0$. For clarity, the workflow of perturbation-based DR-FWI is summarized in Algorithm 1.

3) Pre-training (Warm-Up) with reference model (He and Wang, 2021): The generator neural network is first pre-trained to learn a mapping from random noise to the reference velocity model, expressed as

$$\theta^* = \arg\min_{\mathbf{m}} \left(\|\Phi_\theta(\mathbf{I}_\mathrm{r}) - \mathbf{m}_0\|_2^2 \right). \tag{10}$$

After this warm-up phase, the pretrained network is used in the FWI process to generate the input velocity model, i.e., $\mathbf{m}^* = \Phi_{\theta^*}(\mathbf{I}_\mathrm{r})$, and is subsequently updated during inversion. The workflow of pretraining-based DR-FWI is summarized in Algorithm 2.

Most existing DIP-FWI studies focus on the regularization effects under noisy conditions (Zhu et al., 2022; Sun et al., 2023a; Liu et al., 2025), while its broader potential for addressing practical FWI challenges remains largely unexplored. Notably, the problem of missing data in seismic acquisition shares structural similarity with image inpainting (Lempitsky et al., 2018). In this work, we further explore the effectiveness of reference-guided DR-FWI in handling missing data scenarios, including sparse shot gathers and missing traces.

In the following sections, we conduct a detailed comparative analysis of these proposed strategies, focusing on their inversion performance and convergence behavior. Furthermore, we assess their robustness under data-missing conditions, demonstrating the advantages of proposed approach in improving inversion results in challenging acquisition scenarios.

Algorithm 1: Deep reparameterization-based FWI with perturbation strategy

Input: Observed seismic data $\mathbf{d}_\mathrm{obs}$, initial velocity model $\mathbf{m}_0$, learning rate schedule $\{\alpha_k\}$, maximum iterations K.

Output: Optimized network parameters θ^* and estimated velocity model $\mathbf{m}^*$.

Part I: Network initialization and input construction

// Select generator network architecture

Choose $\Phi(\cdot)$ (e.g., FCNN, CNN, UNet);

// Initialize network parameters

Set θ_0 (e.g., Xavier initialization);

// Construct fixed input tensor

Sample $\mathbf{I}_\mathrm{r} \sim \mathcal{N}(\mu, \sigma^2)$;

// Set hyperparameter

Regularization weight λ, optimizer type, and stopping criteria;

Part II: FWI iterative optimization

For $k = 0, 1, 2, \ldots, K$ **do**:

// Generate velocity model perturbation

$\mathbf{m}_k \leftarrow \Phi_\theta(\mathbf{I}_\mathrm{r}) + \mathbf{m}_0$;

// Forward modeling

$\mathbf{d}_\mathrm{syn} \leftarrow F(\mathbf{m}_k)$;

// Compute misfit

$\mathcal{L} \leftarrow \mathcal{D}(\mathbf{d}_\mathrm{obs}, \mathbf{d}_\mathrm{syn}) + \lambda \mathcal{R}(\mathbf{m}_k)$;

// Gradient backpropagation

$\frac{\partial \mathcal{L}}{\partial \theta_k} \leftarrow \frac{\partial \mathcal{L}}{\partial \mathbf{m}_k} \cdot \frac{\partial \mathbf{m}_k}{\partial \theta_k}$;

// Update network parameters

$\theta_{k+1} \leftarrow \theta_k - \alpha_k \frac{\partial \mathcal{L}}{\partial \theta_k}$;

Return $\theta^*, \mathbf{m}^* = \Phi_{\theta^*}(\mathbf{I}_\mathrm{r}) + \mathbf{m}_0$

Algorithm 2: Deep reparameterization-based FWI with pre-training strategy

Input: Observed seismic data $\mathbf{d}_{\text{obs}}$, initial velocity model $\mathbf{m}_0$, FWI learning rate schedule $\{\alpha_k\}$ and maximum iterations K, pre-training learning rate $\{\alpha_{\text{p}}\}$ and epochs P.

Output: Optimized network parameters θ^* and estimated velocity model $\mathbf{m}^*$.

Part I: Network initialization and pre-training

// Select generator network architecture

Choose $\Phi(\cdot)$ (e.g., FCNN, CNN, UNet);

// Initialize network parameters

Set θ_0 (e.g., Xavier initialization);

// Construct fixed input tensor

Sample $\mathbf{I}_{\text{r}} \sim \mathcal{N}(\mu, \sigma^2)$;

// Set hyperparameter

Regularization weight λ, optimizer type, and stopping criteria;

// Pre-training: map random input to initial model

For $i = 0, 1, 2, \ldots, P$ **do**:

// Generate velocity model from network

$\mathbf{m}_i \leftarrow \Phi_\theta(\mathbf{I}_{\text{r}})$;

// Compute pre-training loss

$\mathcal{L}_{\mathcal{P}} \leftarrow ||\mathbf{m}_i - \mathbf{m}_0||_2^2$;

// Update network parameters via gradient descent

$\theta_{i+1} \leftarrow \theta_i - \alpha_{\text{p}} \frac{\partial \mathcal{L}_{\mathcal{P}}}{\partial \theta_i}$;

Part II: FWI iterative optimization

For $k = 0, 1, 2, \ldots, K$ **do**:

// Generate velocity model perturbation

$\mathbf{m}_k \leftarrow \Phi_\theta(\mathbf{I}_{\text{r}})$;

// Forward modeling

$\mathbf{d}_{\text{syn}} \leftarrow F(\mathbf{m}_k)$;

// Compute FWI misfit

$\mathcal{L} \leftarrow \mathcal{D}(\mathbf{d}_{\text{obs}}, \mathbf{d}_{\text{syn}}) + \lambda \mathcal{R}(\mathbf{m}_k)$;

// Gradient backpropagation

$\frac{\partial \mathcal{L}}{\partial \theta_k} \leftarrow \frac{\partial \mathcal{L}}{\partial \mathbf{m}_k} \cdot \frac{\partial \mathbf{m}_k}{\partial \theta_k}$;

// Update network parameters via gradient descent

$\theta_{k+1} \leftarrow \theta_k - \alpha_k \frac{\partial \mathcal{L}}{\partial \theta_k}$;

Return $\theta^*, \mathbf{m}^* = \Phi_{\theta^*}(\mathbf{I}_{\text{r}})$

2.4. Deep reparameterization for multiparameter FWI

The success of deep reparameterization in single-parameter inversion has motivated its extension to multiparameter FWI. One of the primary challenges in multiparameter FWI is parameter crosstalk, which arises when data residuals caused by inaccuracies in one physical parameter are misattributed to another (Operto et al., 2013). This phenomenon impedes convergence and compromises inversion accuracy. Traditional approaches to mitigating crosstalk typically involve weighting updates across different physical parameters or post-processing gradients using Hessian-based corrections, both of which are either computationally intensive or difficult to implement (Yang et al., 2014; Zhang et al., 2022).

In contrast, deep reparameterization provides an alternative framework in which multiple physical parameters (e.g., v_{p} and ρ in acoustic wave equations, v_{p}, v_{s}, and ρ in elastic wave equations) are jointly represented by a generative neural network:

$$\left[v_{\text{p}}, v_{\text{s}}, \rho, \ldots\right]^{\text{T}} = \Phi_\theta(\mathbf{I}_{\text{r}}), \tag{11}$$

where each output channel of the network corresponds to a specific physical parameter. Rather than explicitly computing and manually weighting the gradients of individual parameters (e.g., $w_{v_{\text{p}}}.\partial\mathscr{L}/\partial \text{v}_{\text{p}}$; $w_{v_{\text{s}}}.\partial\mathscr{L}/\partial v_{\text{s}}$; $w_{\rho}.\partial\mathscr{L}/\partial \rho$), multiparameter DR-FWI performs updates in the network parameter domain ($\partial\mathscr{L}/\partial\theta$, as in Eq. (6)). During backpropagation, the network implicitly learns adaptive weightings among parameters, naturally mitigating crosstalk and effectively replacing the manual gradient scaling traditionally required to balance parameter sensitivities.

Experimental results indicate that the coupling problem in multiparameter inversion can be significantly mitigated using a simple multi-layer CNN with a "backbone-branch" architecture (Fig. 3), provided an appropriate parameterization strategy is adopted. Two key strategies include: 1) Hierarchical representation: A backbone network captures shared structural features and facilitates automatic coupling among different physical parameters, while separate branch networks learn parameter-specific variations. 2) Parameter normalization: Normalizing the output of each branch network stabilizes the inversion process, while subsequent inverse normalization ensures that the predictions

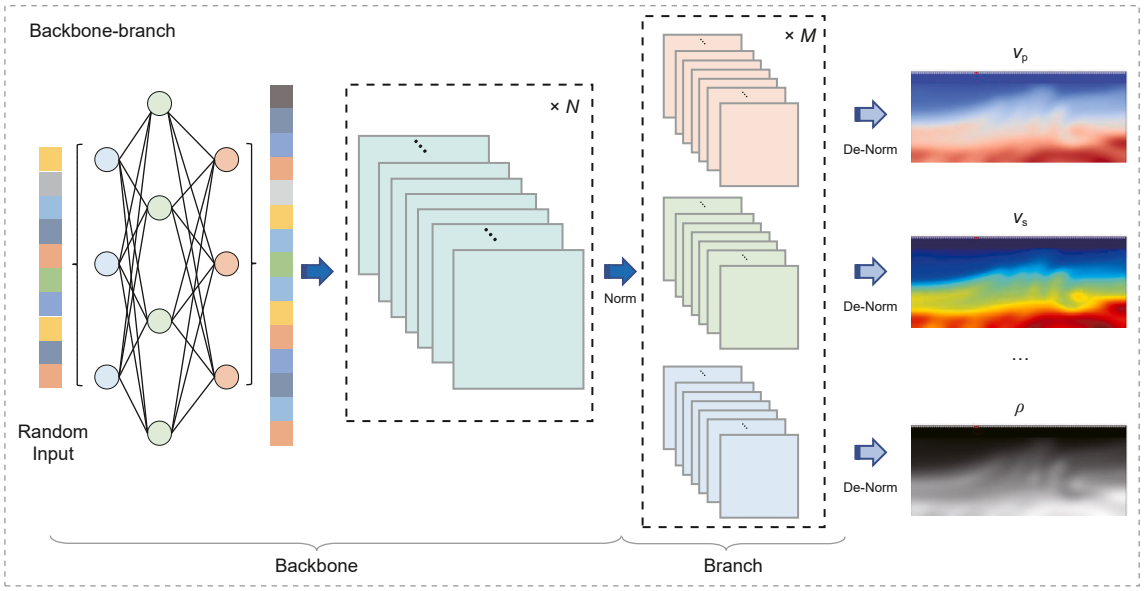

Fig. 3. Overview of the backbone-branch architecture for multiparameter reparameterization. Random inputs are first projected into a high-dimensional latent space via a linear layer to form an expressive representation. The shared backbone, composed of multiple convolutional blocks, extracts common structural features across different physical parameters. Each branch, also composed of convolutional blocks, captures parameter-specific variations. Normalization and denormalization operations account for differences in physical scales.

remain consistent with the physical scales of each parameter. By integrating these strategies, the proposed deep reparameterization framework improves the stability of multiparameter inversion while reduces the reliance on traditional crosstalk correction techniques.

3. Experiments and results

3.1. Baseline models and quantitative metrics

To evaluate the effectiveness of reference-guided deep reparameterization, we select three well-known 2D geophysical models: the Marmousi2 model (Martin et al., 2006), the Overthrust model (Aminsadeh, 1996), and the Foothill model (Gray and Marfurt, 1995). The corresponding initial velocity models are obtained by applying a Gaussian smoothing operation. More details about these models and the corresponding observation system configurations can be found in Table 1. All experiments are conducted on a single NVIDIA GPU and implemented using PyTorch. Forward modeling for acoustic and elastic wave propagation is performed using the ADFWI toolbox (Liu, 2024), and the global-correlation misfit function is adopted to measure the discrepancy between the observed and synthetic waveforms (Choi and Alkhalifah, 2012).

For conventional FWI, the Adam optimizer (Kingma and Ba, 2017) is used with an initial learning rate of 10, decayed by multiplying a factor of 0.75 every 100 iterations, with a maximum of 300 inversion iterations (Liu et al., 2025). For DR-FWI, we employ neural network-based reparameterization, which can be instantiated as CNNs, UNets, or MLPs. In the perturbation-based DR-FWI (Algorithm 1), the Adam optimizer is used with an initial learning rate of $1e^{-4}$, StepLR decaying the learning rate by a factor of 0.75 every 100 iterations, and a maximum of 300 inversion iterations. In the pretraining-based DR-FWI (Algorithm 2), the network is first pretrained to learn the mapping from random vectors to the initial velocity model using an Adam optimizer with an initial learning rate of $5e^{-4}$, StepLR decaying the learning rate by 0.5 every 1000 iterations for a total of 10,000 pretraining iterations. During the subsequent inversion, the training hyperparameters and optimizer settings are identical to those of the perturbation-based DR-FWI.

To assess the inversion performance of different methods, we employ the mean absolute percentage error (MAPE) (Hyndman and Koehler, 2006), the structural similarity index measure (SSIM) (Wang et al., 2004), and the signal-to-noise ratio (SNR) to quantify the discrepancy between the inverted and true models:

$$\mathrm{MAPE}(v,\widehat{v}) = \frac{1}{M\times N}\sum_{i=1}^{M}\sum_{j=1}^{N}\left|\frac{v_{i,j}-\widehat{v}_{i,j}}{v_{i,j}}\right|\times 100\%, \tag{12}$$

Table 1
Key parameters of the tests for DR-FWI in various types of earth models.

	Acoustic Marmousi2	Acoustic Overthrust	Acoustic FootHill	Elastic anomaly	Elastic Marmousi2
Number of grids	76 × 200	90 × 200	97 × 267	100 × 200	68 × 200
Grid spacing, m	40	50	40	30	45
Time steps	2500	1600	2000	2500	2500
Time interval, s	0.003	0.003	0.003	0.0025	0.003
Number of receivers	200	200	267	200	200
Number of sources	40	40	54	40	40
Source dominant frequency, Hz	5	5	5	5	5
Smooth window size	240 m × 240 m	300 m × 300 m	240 m × 240 m	–	180 m × 180 m
Inverted parameters	v_p	v_p	v_p	v_p, v_s, ρ	v_p, v_s, ρ

$$\mathrm{SSIM}(v,\widehat{v}) = \frac{(2\mu_v\mu_{\widehat{v}} + c_1)(2\sigma_{v\widehat{v}} + c_2)}{(\mu_v^2 + \mu_{\widehat{v}}{}^2 + c_1)(\sigma_v^2 + \sigma_{\widehat{v}}{}^2 + c_2)}, \tag{13}$$

$$\mathrm{SNR}(v,\widehat{v}) = 10\log_{10}\frac{\|v\|_2^2}{\|\widehat{v} - v\|_2^2}, \tag{14}$$

where M and N denote the size of velocity model, $v_{i,j}$ and $\widehat{v}_{i,j}$ are the true and inverted velocity values, respectively, μ_v and σ_v are the local mean and standard deviation for the true model, $\mu_{\widehat{v}}$ and $\sigma_{\widehat{v}}$ are those for the inverted model, and $\sigma_{v,\widehat{v}}$ is the cross-covariance between the true and inverted models. Higher SSIM and SNR values, along with lower MAPE, indicate better inversion performance.

3.2. Ablation studies on reparameterization network architectures and strategies

Previous studies have demonstrated the potential of deep reparameterization networks in FWI, with different architectures and strategies exhibiting varying levels of performance. Commonly adopted architectures include MLPs, CNNs, and UNet, while two prevalent reparameterization strategies involve learning the velocity perturbation ($\Delta\mathbf{m}$, Eq. (9)) or directly training the neural network to learn the velocity model ($\mathbf{m}$, Eq. (10)).

A fundamental question thus arises: which combination of reparameterization network architecture and training strategy yields the most effective FWI performance? (He and Wang, 2021). To address this, we evaluate three distinct neural network architectures.

1) UNet: The input is a randomly initialized 2-D array, from which the network learns the statistical representation of the velocity model through a sequence of down-sampling and up-sampling operations. Skip connections are incorporated to retain fine-grained details by linking corresponding layers in the down-sampling and up-sampling paths.
2) MLPs: The input is a 1-D random vector, which is transformed into the velocity model through multiple fully connected layers.
3) CNNs: The input consists of 1-D random arrays, which are first processed through a fully connected layer to generate an initial 1-D representation of the velocity model. The representation is then reshaped into a 2-D format and refined using multiple convolutional layers. Finally, a 1×1 convolutional layer merges multi-channel representations into the final velocity model output.

Detailed structures are shown in Supplementary Tables S1–S3.

Combining the above architectures with the two deep reparameterization strategies leads to six different combinations: CNN-v_p, CNN-Δv_p, MLP-v_p, MLP-Δv_p, UNet-v_p, and UNet-Δv_p. We first evaluate all six combinations on the Marmousi2 model. To reduce the impact of network hyperparameter choices (e.g., the number of layers and channels), we perform a Network Architecture Search (NAS) for each architecture and select the best-performing configuration for comparison. The ranges of layers and channels were chosen based on prior studies (Zhu et al., 2021; Sun et al., 2023a; Liu et al., 2025) and practical considerations. We initially explored broader network structures, and then narrowed the search space to achieve a balance between model expressiveness, training stability, and computational efficiency. Based on this strategy, we adopted the following optimized ranges: CNN architectures with 1–4 layers, each having 64, 128, 256, and 512 channels, totaling 16 different variants; MLPs with 1 or 2 layers, each with 100, 1000, 2000, and 3000 neurons per layer, resulting in 8 variants; and UNet with 1–4 encoder-decoder blocks, each with 64, 128, 256, and 512 channels, yielding 16 variants. The best inversion results for each reparameterization strategy on the Marmousi2 model are illustrated in Fig. 4. Panel (a) depicts the true Marmousi2 model (v_p), while panel (b) presents the inversion result using traditional FWI. Panels (c) and (d) correspond to the UNet-based inversion results using the v_p and Δv_p initial model embedding strategies, respectively. Panels (e) and (f) show the results for the MLP-based variants, respectively. Panels (g) and (h) depict the CNN-based results.

To further assess the inversion performance of different strategies, we analyze the evolution of data misfit and model residuals, as shown in Fig. 5. The colored lines represent the average performance across different network variants, while the shaded regions indicate the range of fluctuations in inversion results due to hyperparameter variations. Panel (a) presents the evolution of the global-correlation misfit (GC-loss), demonstrating that all inversion strategies successfully converge in the data domain. Panels (b), (c), and (d) show model residuals in terms of MAPE, SNR, and SSIM, respectively. Beyond the Marmousi2 model, similar experiments were conducted on the Overthrust and Foothill models. The best performance achieved by each reparameterization strategy on these velocity models is summarized in Table 2, while detailed information on network architectures, parameter counts, memory consumption, and convergence iterations is provided in Supplementary Table S4–S6. Two key findings emerge from these results: 1) The warm-up strategy (CNN–v_p, MLP–v_p, UNet–v_p) consistently outperforms the perturbation-learning strategy (CNN–Δv_p, MLP–Δv_p, UNet–Δv_p) across all network architectures. 2) A simple CNN-based reparameterization network achieves superior inversion performance compared to the baseline and even surpasses more complex architectures such as UNet. To further validate these findings, additional tests using a smoother initial velocity model (Gaussian smoothing with a scale of 320 m × 320 m) were conducted, and the corresponding visualizations and metrics are provided in Supplementary Figs. S4 and S5.

3.3. Robustness analysis under noise and acquisition sparsity

Neural networks are known to exhibit a spectral bias during the learning process in the frequency domain, prioritizing low-frequency signal components before gradually resolving high-frequency details (Rahaman et al., 2018; Chakrabarty and Maji, 2019). This property is aligned with the multi-scale principles of FWI, where low-frequency data constrain large-scale structures prior to high-frequency refinements. In DR-FWI, the neural network progressively learns statistical characteristics of the velocity model through iterative inversion, enhancing its structural representation capability. The spectral bias inherently imposes a prior constraint, reducing the risk of local minima entrapment by initially recovering smooth velocity models (low-frequency dominance) and later refining high-frequency details. This regularization mechanism not only stabilizes the inversion workflow but also improves robustness against noise and data sparsity. To validate these claims, rigorous evaluations are conducted using the Marmousi2 model under controlled noise and sparsity conditions.

First, the noise robustness of the deep reparameterization method is quantified by injecting Gaussian white noise into the observation data. The noise levels are defined by a fixed mean $\mu = \mu_0$ and varying standard deviations from $\sigma = 1\sigma_0$ to $\sigma = 6\sigma_0$, where μ_0 and σ_0 correspond to the mean and standard deviation of the clean observation data, respectively. Fig. 6(a) compares a clean shot gather with its noise-contaminated counterpart. Building on prior results demonstrating the superiority of the CNN–v_p architecture (CNN-based velocity parameterization with pretraining-based initialization), we adopt this strategy for the noise

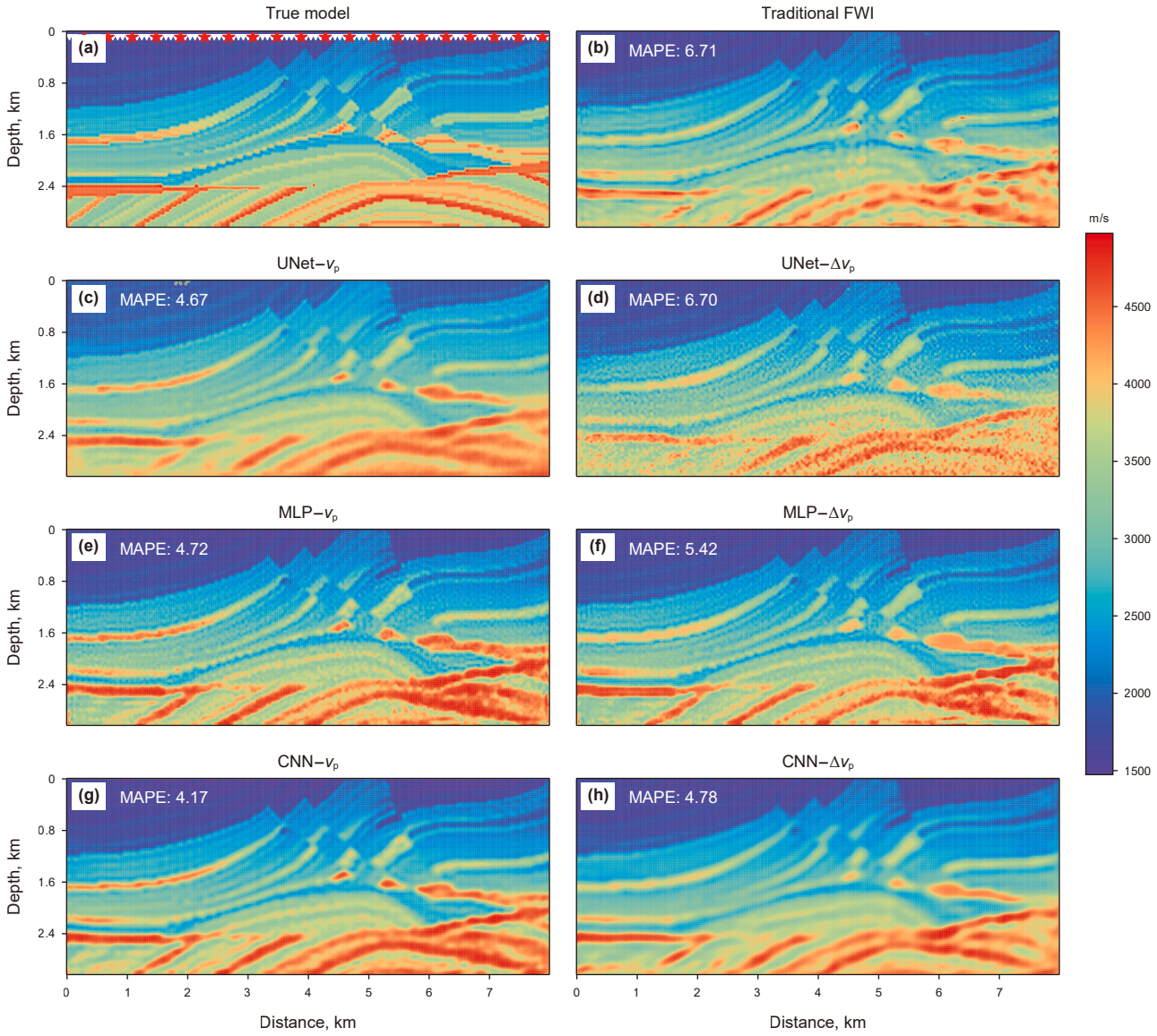

Fig. 4. Comparison of inversion results using different deep reparameterization architectures and strategies for the Marmousi2 model. **(a)** True velocity model. **(b)** Baseline inversion result using conventional FWI. **(c)**–**(d)** Results obtained using UNet-based reparameterization for v_p and Δv_p. **(e)**–**(f)** Results obtained using MLP-based reparameterization for v_p and Δv_p. **(g)**–**(h)** Results obtained using CNN-based reparameterization for v_p and Δv_p.

robustness study. Fig. 6(b) presents the MAPE of inversion results across increasing noise levels: conventional FWI (black line) versus CNN–v_p reparameterization (blue line), with shaded regions indicating architectural variability across CNNs. Fig. 6(c) and (d) visualize inversion results for the $6\sigma_0$ noise case, highlighting the enhanced stability and accuracy of the reparameterized approach under extreme-noise conditions. In addition to these synthetic noise tests, we further benchmark the method under actual noise conditions by superimposing real field noise onto synthetic data, with detailed results and discussions provided in Supplementary Text S1. For completeness, we also compared DR-FWI with traditional regularized FWI that employs both Tikhonov-type and Total Variation (TV)-type regularizations on the model parameters. As detailed in Supplementary Text S2, the performance of these traditional regularization methods is highly sensitive to the choice of the regularization factor, and their optimal results remain inferior to those of DR-FWI. In contrast, DR-FWI achieves higher accuracy without explicitly introducing additional regularization terms, indicating that the deep parameterization itself serves as an implicit prior that enhances robustness.

While the robustness of DR-FWI to noise has been extensively validated (Zhu et al., 2022; Sun et al., 2023a), its effectiveness under data sparsity remains less explored. To investigate this, we design two sparsity scenarios: (1) sparse sources (20, 10, 5, and 2 sources paired with 200 receivers), and (2) sparse receivers (10 sources paired with 200, 50, or 20 receivers). Fig. 7(a) summarizes seven acquisition configurations: four fixed-receiver setups (200 receivers, 2 to 20 sources) and three fixed-source setups (10 sources, 20 to 200 receivers). Fig. 7(b) shows MAPE curves for both the sparse-source (left) and the sparse-receiver (right) cases, comparing conventional and DR-FWI. Fig. 7(c) and (d) display the inverted velocity models for an extreme receiver-sparse case (10 sources, 20 receivers), highlighting the DR-FWI superior structural recovery (Fig. 7(d)). Overall, under moderate source sparsity conditions (>2 sources), DR-FWI exhibits significantly higher stability and accuracy than conventional FWI. Even under extreme sparsity

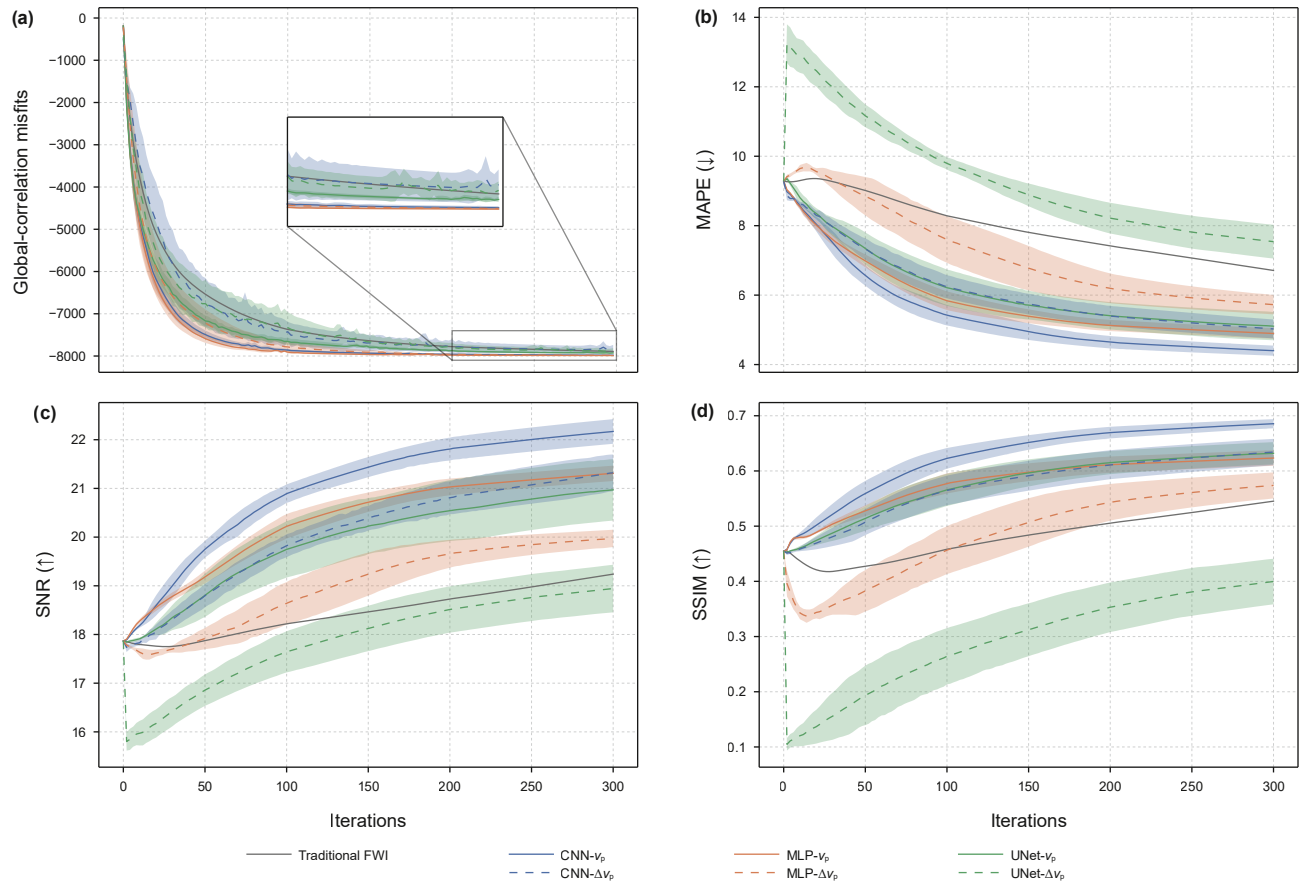

Fig. 5. Comparison of inversion performance metrics for different deep reparameterization architectures and strategies. **(a)** Evolution of the global-correlation misfit. **(b)**–**(d)** Model residuals in terms of MAPE, SNR, and SSIM, respectively. Solid (CNN–v_p, MLP–v_p, UNet–v_p) and dashed (CNN–Δv_p, MLP–Δv_p, UNet–Δv_p) lines represent the mean performance for each reparameterization strategy, while the shaded areas indicate variations induced by different hyperparameter choices (e.g., the number of layers and channels).

Table 2
Quantitative comparison of reparameterization strategies on the Marmousi2 model based on GC-loss, MAPE, SSIM, and SNR.

	Traditional FWI	UNet–Δv_p	UNet–v_p	MLP–Δv_p	MLP–v_p	CNN–Δv_p	CNN–v_p
GC-loss (Marmousi2)	−7892.49	−7937.52	−7944.44	−7989.94	−7991.62	−7987.98	−7990.07
MAPE (↓)	6.71	6.70	4.67	5.42	4.72	4.78	4.17
SSIM (↑)	0.55	0.44	0.66	0.60	0.64	0.66	0.70
SNR (↑)	19.24	19.38	21.73	20.24	21.54	21.95	22.57
GC-loss (Overthrust)	−7957.52	−7985.95	−7993.14	−7984.25	−7984.37	−7993.53	−7991.66
MAPE (↓)	3.91	4.43	2.99	3.05	2.75	2.52	2.38
SSIM (↑)	0.66	0.45	0.63	0.74	0.75	0.76	0.81
SNR (↑)	25.83	25.08	28.21	27.83	28.46	29.41	29.52
GC-loss (Foothill)	−14388.7	−14378.7	−14379.1	−14366.1	−14407.4	−14410.7	−14407.5
MAPE (↓)	1.63	3.34	2.73	1.82	1.41	1.41	1.37
SSIM (↑)	0.84	0.50	0.57	0.84	0.85	0.82	0.87
SNR (↑)	25.03	26.17	27.68	30.34	32.20	32.25	32.32

regimes (2 sources or 20 receivers), it maintains a measurable performance advantage, with particularly pronounced benefits observed in receiver-sparse scenarios (Fig. 7(d)).

3.4. Multiparameter inversion and crosstalk mitigation

The superior performance of DR-FWI in single-parameter inversion stems from the inherent spectral bias and implicit prior regularization of deep neural networks during training. Beyond these properties, the flexibility of this framework enables seamless extension to multiparameter inversion by incorporating a backbone-branch structure. In multiparameter DR-FWI, a shared backbone network captures coupled representations of subsurface parameters (e.g., v_p, v_s, ρ), while dedicated branch networks disentangle parameter-specific features through channel-wise normalization tailored to their respective physical ranges. This design is validated through two synthetic experiments: a geometrically idealized anomaly model and the Marmousi2 benchmark.

To evaluate the crosstalk mitigation capabilities of the proposed multiparameter inversion framework, we first design a synthetic elastic model inspired by the CTS model (Dokter et al., 2017) and STH model (Dhara and Sen, 2022). The designed model consists of an elastic layer overlying a homogeneous half-

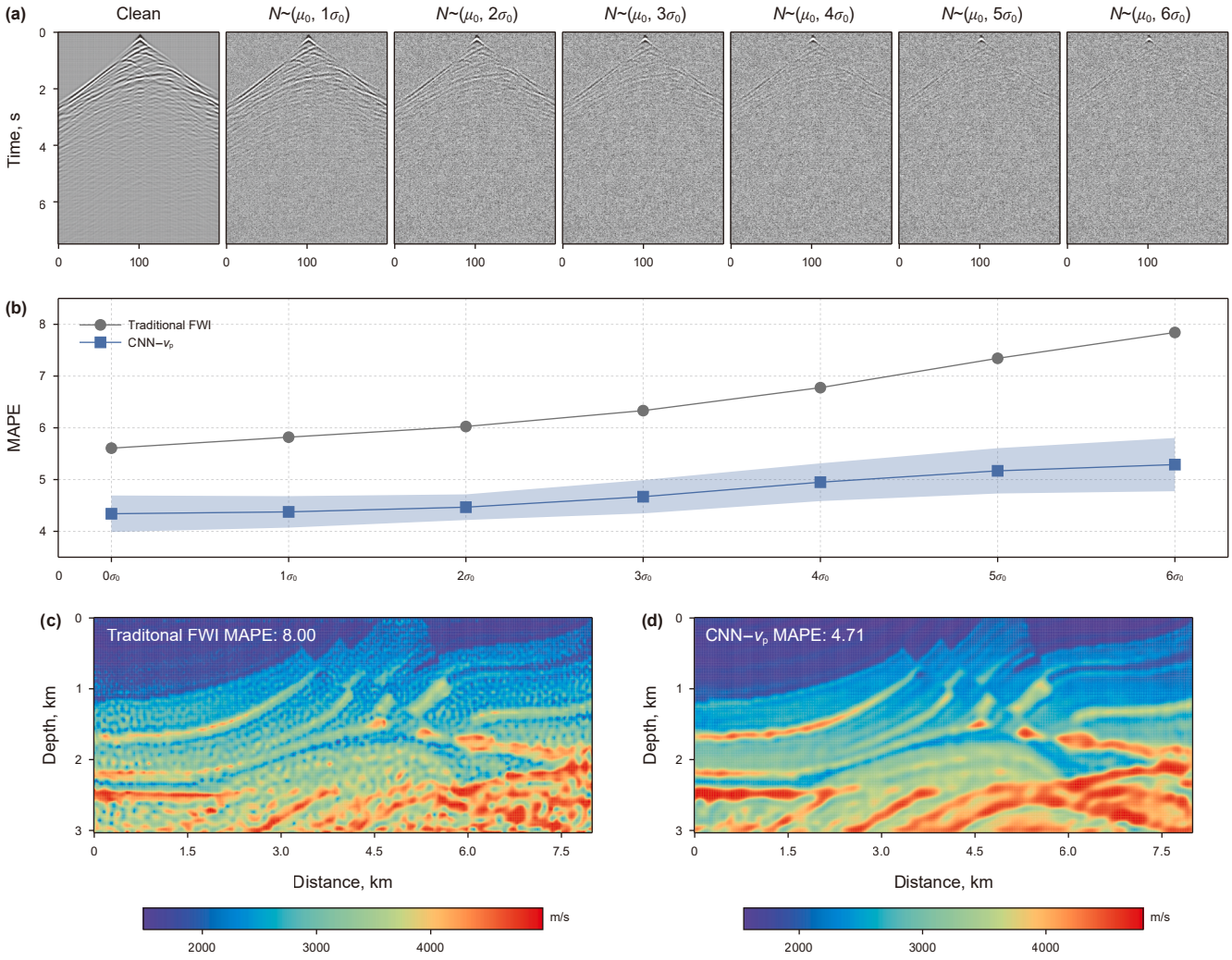

Fig. 6. Robustness evaluation of DR-FWI under noisy data scenarios on the Marmousi2 model. **(a)** Example shot gather showing clean and noisy data with Gaussian noise levels from $1\sigma_0$ to $6\sigma_0$. **(b)** MAPE of inversion results with and without the CNN–v_p deep reparameterization strategy under different noise conditions. Shaded areas represent MAPE variations across different CNN architectures. **(c)** and **(d)** Inversion results with $6\sigma_0$ Gaussian noise, highlighting the improved stability of the deep reparameterization method under high noise levels.

space and embeds three distinct types of geometric anomalies: five circular v_p anomalies, five square v_s anomalies, and five triangular ρ anomalies. Despite their idealized nature, these non-overlapping and geometrically distinct anomalies (Fig. 8(a)–(c)) provide a controlled testbed for isolating and quantifying crosstalk effects. The model spans a 100×200 grid with 30-m spacing, using 40 sources and 200 receivers. We perform simultaneous inversion for v_p, v_s, and ρ, a parameter triplet commonly adopted in elastic FWI. All inversions are initialized from homogeneous background models devoid of any embedded anomalies, while using the same learning rate for all parameters in traditional FWI. The results of conventional FWI (Fig. 8(d)–(f)) reveal severe crosstalk artifacts: v_p and ρ models show spurious structures mirroring the ρ anomalies, while the ρ model, despite preserving the intended geometry, suffers from unstable background regions (MAPE > 13). In contrast, the CNN–v_p reparameterized FWI (Fig. 8(g)–(i)) achieves near-perfect recovery of all three parameters (MAPE < 0.5), eliminating crosstalk and preserving background stability. These results demonstrate that the proposed deep reparameterization framework effectively decouples multi-parameter interactions while enforcing physically consistent relationships across multiple elastic properties.

Building upon the validated effectiveness of deep reparameterization in suppressing crosstalk in the simplified anomaly model, we further apply this approach to the more geologically realistic models, including the Marmousi2 model (Fig. 9) and the Valhall model (Supplementary Fig. S6). The true v_p, v_s, and ρ models of Marmousi2 are shown in Fig. 9(a)–(c). Fig. 9(d)–(f) display the results obtained by conventional FWI without reparameterization. Due to strong parameter coupling, the deeper regions of the v_p and v_s models exhibit significant blurring, while the ρ model, being less sensitive to the seismic data, appears structurally incoherent and highly disordered. In contrast, the inversion results obtained by the proposed deep reparameterization method (Fig. 9(g)–(i)) show substantial improvements. All three parameters are recovered with enhanced structural fidelity and reduced artifacts, closely matching the ground truth (Fig. 9(a)–(c)). These results highlight the robustness and generalizability of the proposed reparameterization framework in complex geological settings, further confirming its applicability to multiparameter inversion tasks. However, it should be noted that its performance depends on the selected network architecture and hyperparameter settings, and may deteriorate if the network s representational bias does not align with the true subsurface structures.

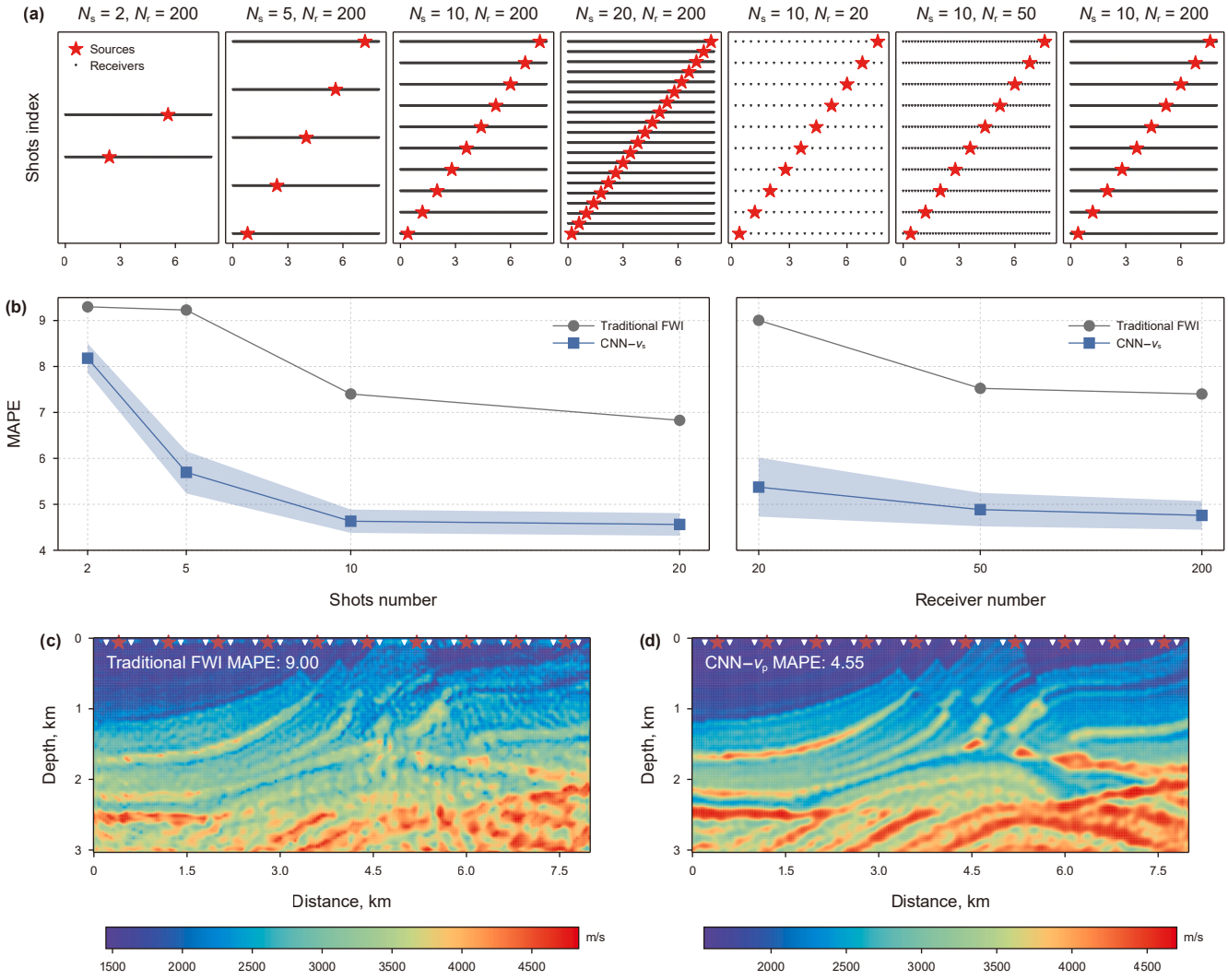

Fig. 7. Robustness evaluation of DR-FWI under sparse-data scenarios on the Marmousi2 model. **(a)** Seven observation configurations: the first four fix 200 receivers with source counts ranging from 2 to 20; the last three fix 10 sources with receiver counts increasing from 20 to 200. **(b)** MAPE curves comparing inversion accuracy for sparse-source (left: 2 to 20 sources, 200 receivers) and sparse-receiver (right: 10 sources, 20 to 200 receivers) scenarios, with (blue line) and without (black line) deep reparameterization. Shaded regions indicate MAPE variability across CNN architectures. **(c)** Conventional FWI results and **(d)** DR-FWI results for an extreme receiver-sparse configuration (10 sources, 20 receivers).

The effects of network design and parameter selection on multi-parameter inversion performance are further examined in the Discussion section.

4. Discussion

4.1. Insights on deep reparameterization and DIP mechanism

Although empirical results across multiple FWI studies have highlighted the regularization benefits of deep reparameterization, the underlying theoretical mechanisms remain insufficiently investigated (Zhu et al., 2022; Sun et al., 2023a). This work bridges this knowledge gap through systematic spectral dynamics analysis, demonstrating that deep reparameterization implicitly imposes progressive frequency regularization via the spectral learning properties of deep neural networks. To quantitatively analyze the spectral behavior of deep reparameterization, we investigate the frequency-dependent convergence characteristics during inversion. Prior studies on deep image prior have shown that deep neural networks do not learn all frequency components simultaneously; instead, they reconstruct low-frequency structures first and gradually recover high-frequency details as optimization proceeds (Lempitsky et al., 2018). This phenomenon, known as the "spectral bias" or "frequency principle", implies that the network itself acts as a form of implicit frequency regularization. To capture this behavior quantitatively in the context of FWI, we employ the Frequency-Band Correspondence (FBC) metric (Shi et al., 2022), which evaluates how closely the Fourier spectrum of the inverted model matches that of the true model across different wavenumber ranges. Specifically, the two-dimensional Fourier spectra of the true and inverted velocity models at iteration i are computed as

$$= \mathrm{FFT}(\mathbf{m}_{\mathrm{true}}), \widehat{\mathbf{m}}_i = \mathrm{FFT}(\mathbf{m}_i), \tag{15}$$

where $\mathrm{FFT} : \mathbb{R}^{H\times W} \to \mathbb{C}^{H\times W}$ denotes the two-dimensional discrete Fourier transform applied to the velocity model. The element-wise amplitude ratio is then defined as

$$\mathscr{H}(k_x, k_z) = \frac{|\widehat{\mathbf{m}}_i(k_x, k_z)|}{|\widehat{\mathbf{m}}_{\mathrm{true}}(k_x, k_z)|}, \tag{16}$$

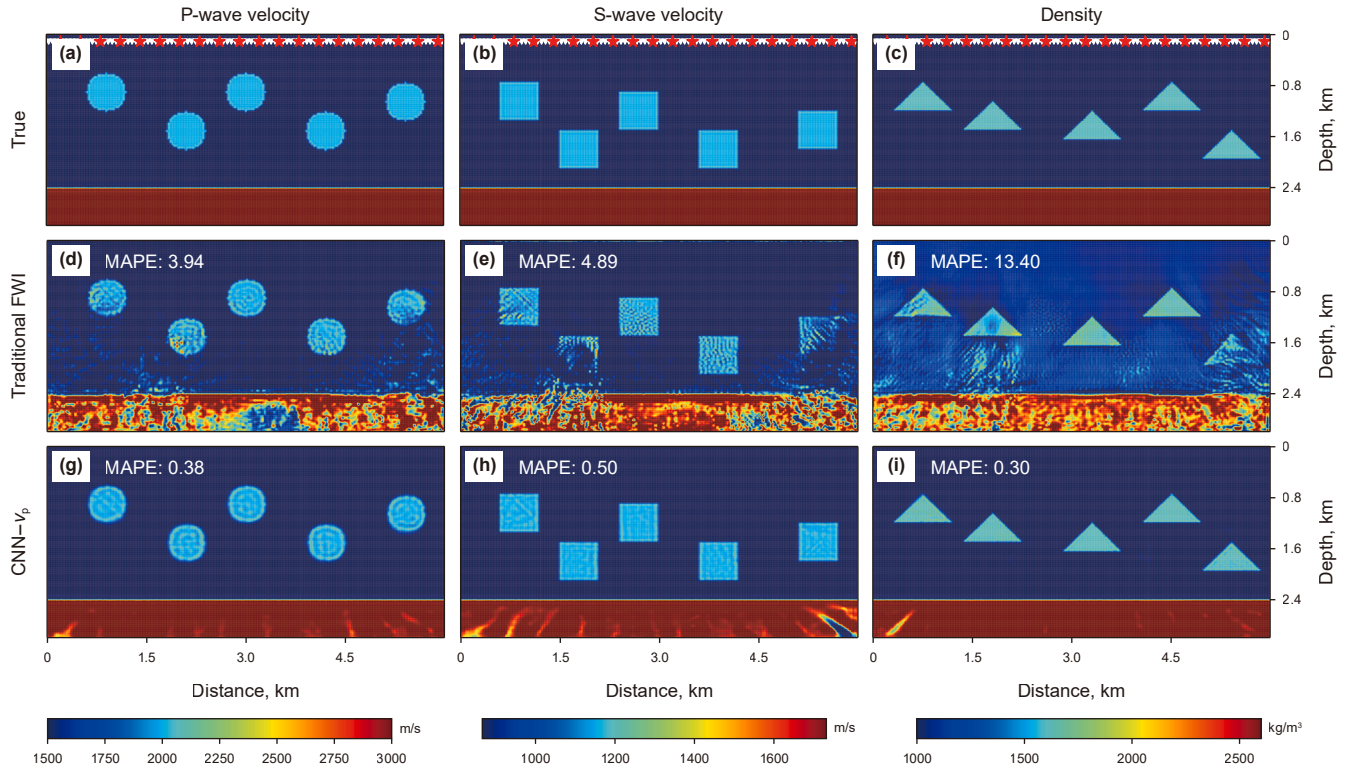

Fig. 8. Multiparameter inversion results on an anomaly model. **(a)**–**(c)** Ground-truth models of v_p, v_s, and ρ, respectively. **(d)**–**(f)** Inversion results obtained using traditional FWI. **(g)**–**(i)** Inversion results obtained using DR-FWI. Red stars and white triangles denote source and receiver locations, respectively.

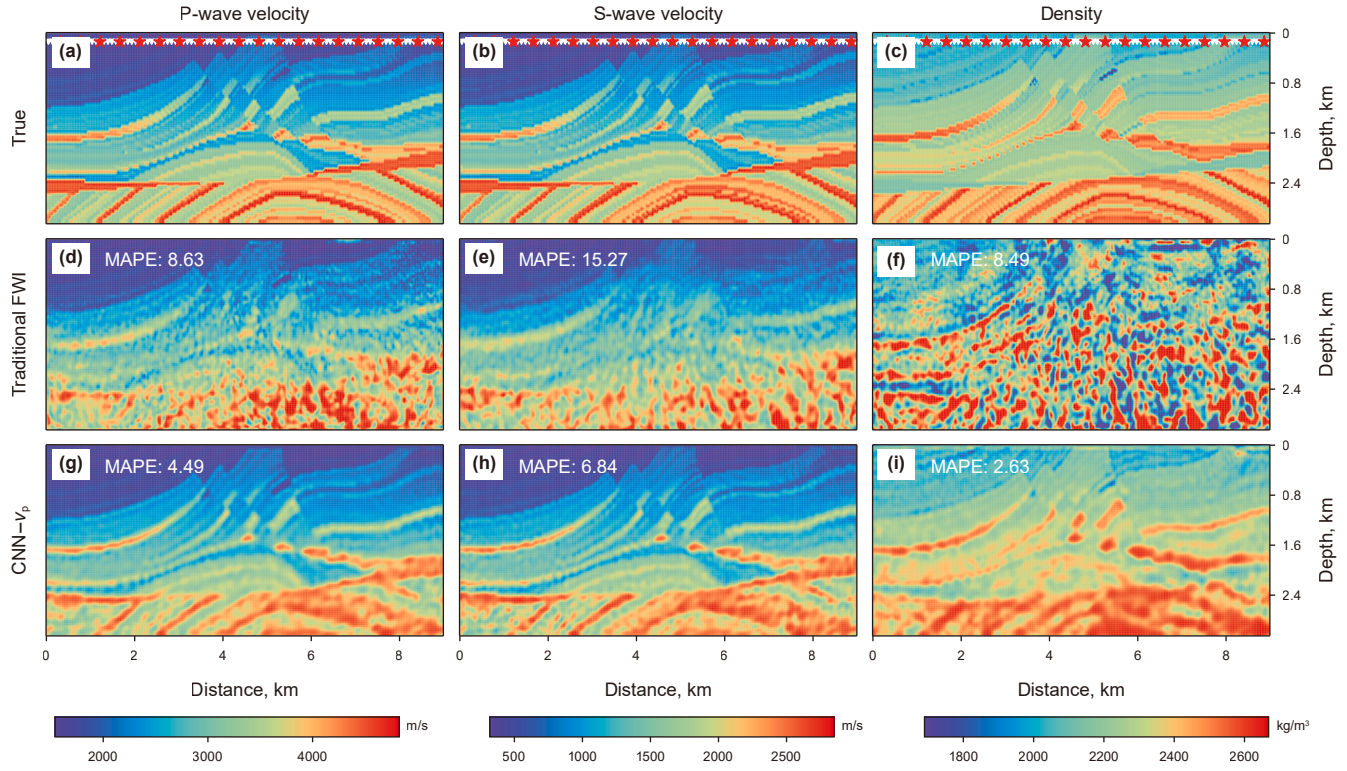

Fig. 9. Multiparameter inversion on Marmousi2 model. **(a)**–**(c)** Ground-truth models of v_p, v_s, and ρ, respectively. **(d)**–**(f)** Inversion results obtained using traditional FWI. **(g)**–**(i)** Inversion results obtained using DR-FWI. Red stars and white triangles denote source and receiver locations, respectively.

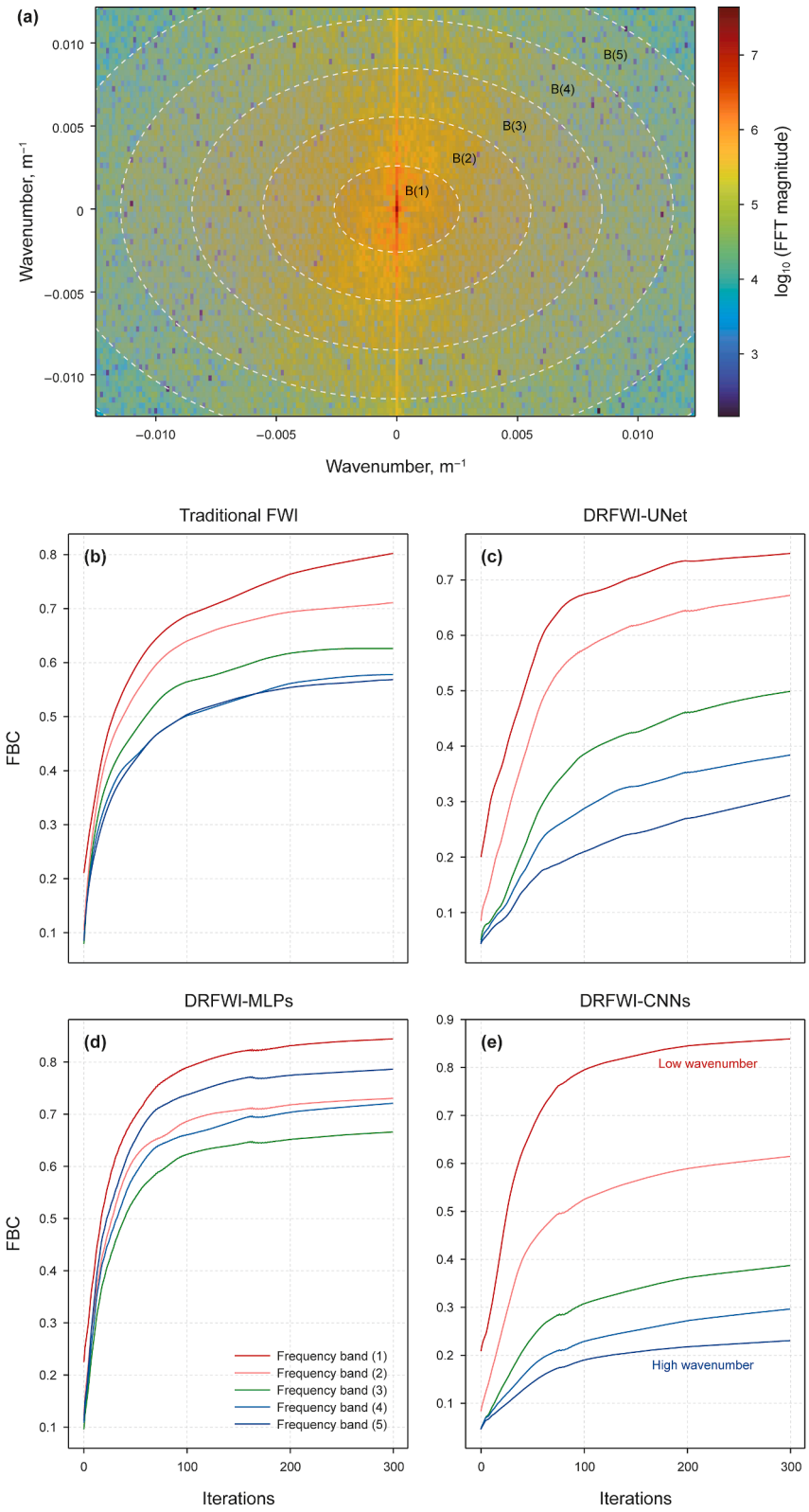

Fig. 10. Frequency-band correspondence (FBC) analysis for conventional FWI and DR-FWI with different reparameterization networks. **(a)** Schematic illustration of the FBC frequency-band division. The background shows the logarithmic Fourier amplitude spectrum of the true Marmousi2 model. The spectrum is divided into five concentric bands, $\mathscr{B}^{(1)} - \mathscr{B}^{(5)}$, ranging from low-to high-wavenumber components. **(b)**–**(e)** Evolution of the mean FBC values within each spectral band during inversion for **(b)** conventional FWI and DR-FWI using **(c)** UNet, **(d)** MLPs, and **(e)** CNNs.

where (k_x, k_z) denotes the spatial frequency index, and values of $\mathscr{K}(k_x, k_z)$ close to 1 indicate strong spectral similarity. Given the central symmetry of the Fourier spectrum, we divide it into $N = 5$ concentric, non-overlapping wavenumber bands $\mathscr{B}^{(1)}, \ldots, \mathscr{B}^{(5)}$ (Fig. 10(a)), ranging from low-to high-wavenumber regions. The FBC value for each band is obtained by averaging $\mathscr{K}(k_x, k_z)$ over the corresponding annular region:

$$\mathrm{FBC}^{(n)} = \frac{1}{\left|\mathscr{B}^{(n)}\right|} \sum_{(k_x,k_z)\in\mathscr{B}^{(n)}} \mathscr{K}(k_x, k_z), n = 1, \ldots, 5. \tag{17}$$

This averaged measure quantifies the degree of spectral correspondence within each wavenumber band.

Fig. 10(b)–(e) show the evolution of FBC during inversion for conventional FWI and DR-FWI with different reparameterization networks. In conventional FWI, FBC curves across all wavenumber bands exhibit nearly uniform slopes, indicating that low- and high-wavenumber components are updated at comparable rates (Fig. 10(b)). DR-FWI using UNet (Fig. 10(c)) and CNNs (Fig. 10(e)) exhibits a hierarchical spectral learning behavior, with low-wavenumber components converging rapidly and high-wavenumber components progressively reconstructed. Notably, compared with UNet, CNN achieves faster convergence and higher FBC values for low-wavenumber bands, indicating more effective early recovery of large-scale structures. This frequency-dependent learning sequence mitigates convergence to local minima by prioritizing large-scale structure recovery before finer details, while slightly lower FBC values at high wavenumbers reflect the limited high-frequency content of the target model (Lempitsky et al., 2018; Shi et al., 2022). By contrast, DR-FWI using MLPs updates low- and high-wavenumber components more uniformly, exhibiting little low-frequency bias (Fig. 10(d)). This behavior likely arises from the large number of neurons in fully connected layers, which enables MLPs to fit low- and high-wavenumber components simultaneously, thereby reducing the low-frequency preference relative to convolutional architectures. Overall, these results quantitatively confirm that deep reparameterization acts as an implicit low-to-high wavenumber regularizer and elucidate how convolutional architectures naturally induce hierarchical spectral learning dynamics in DR-FWI.

4.2. Structure ablation of multiparameter reparameterization

The incorporation of a branch network structures and normalization or denormalization operations extends the deep reparameterization approach to multiparameter FWI. This method enables parameter-specific feature extraction and scale alignment across different physical properties, thereby effectively addressing the crosstalk problem in multiparameter inversion. Its effectiveness has been validated on the synthetic anomaly, Marmousi2 and Valhall models. In this section, we investigate the impact of CNN hyperparameters—such as the number of layers in the backbone and branch networks and the number of feature extraction channels—on inversion performance. Two network architectures are

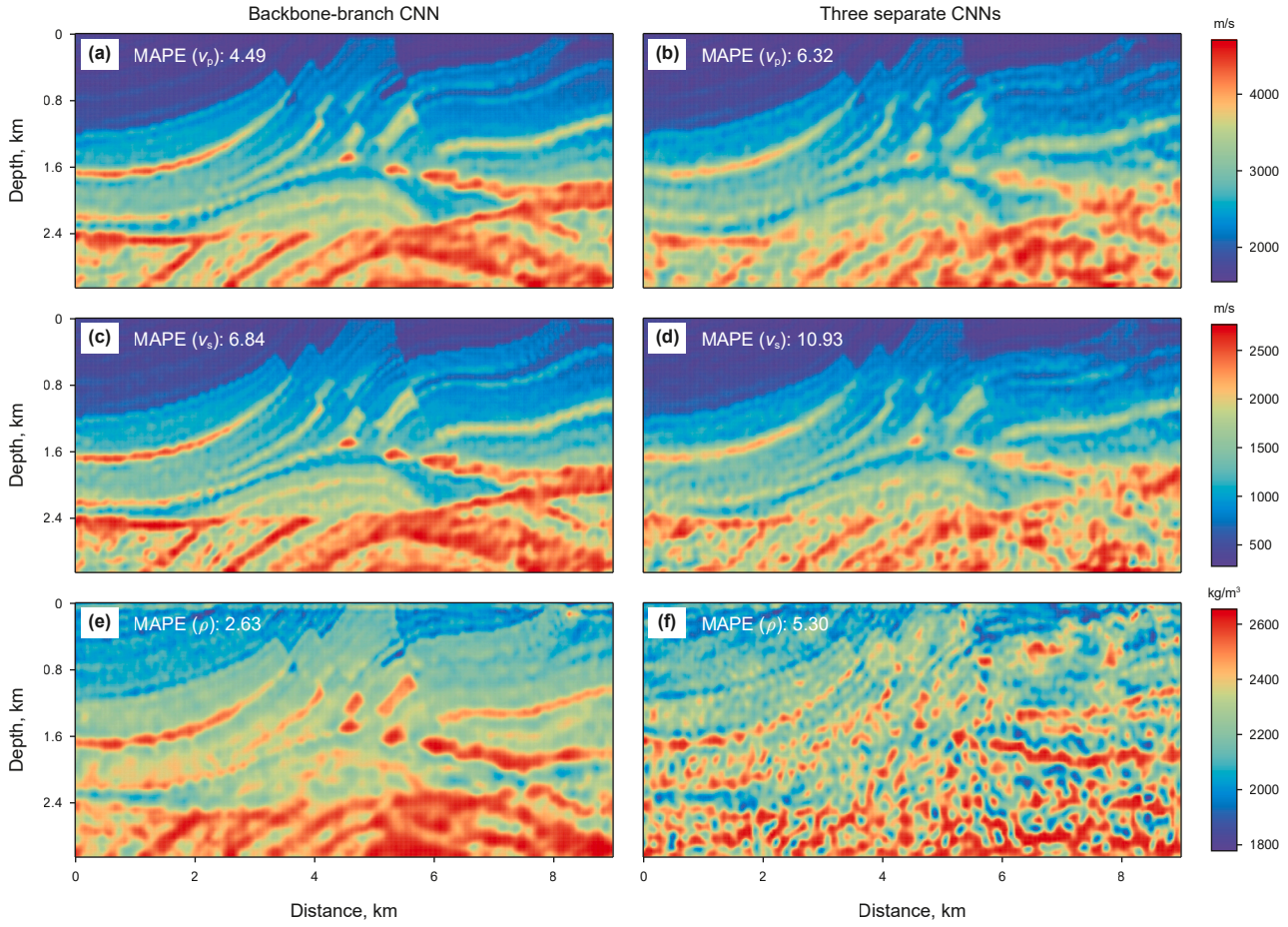

Fig. 11. Comparison of inversion results for v_p, v_s, and ρ obtained from two reparameterization strategies in multiparameter FWI: **(a)**, **(c)**, **(e)** results from a single backbone-branch network, and **(b)**, **(d)**, **(f)** results from three independent CNNs. Subfigures **(a)** and **(b)** show v_p; **(c)** and **(d)** show v_s; and **(e)** and **(f)** show ρ.

compared: (1) a unified backbone-branch CNN that jointly generates all physical parameters (e.g., v_p, v_s, ρ), and (2) multiple independent CNNs, each responsible for generating a single parameter. We perform a systematic search across 24 backbone-branch network configurations. The backbone network consists of 2- and 3-layer CNNs with 64, 128, 256, and 512 channels. The branch networks include 1-, 2-, and 3-layer CNNs with channel settings of {1}, {32, 1}, and {64, 32, 1}, respectively. For comparison, we also evaluate eight independent CNN configurations (i.e., multiple independent CNNs) comprising 2-layer and 3-layer CNNs with 32, 64, 128, and 256 channels.

Fig. 11(a)–(c), and (e) display the best inversion results for v_p, v_s, and ρ using the backbone-branch reparameterization strategy, while Fig. 11(b)–(d), and (f) show the best inversion results using three independent CNNs. Fig. 12 illustrates the evolution of MAPE during the iterative process for different parameterization strategies, where the solid lines represent the mean inversion results for each strategy across varying network configurations, and the shaded areas indicate the variation in inversion results among different network structures. The comparative results demonstrate that both reparameterization strategies significantly outperform traditional FWI methods. Among them, the backbone-branch network exhibits the best performance in recovering different physical parameters especially in the case of density, where the improvement is most pronounced.

Based on a systematic evaluation of various backbone-branch configurations (Fig. 13), we observe that the backbone-branch structure consistently outperforms multiple independent CNNs using the same backbone. At the same time, the performance of this structure is sensitive to architectural choices: excessively deep branches can degrade inversion accuracy, whereas backbones with fewer layers but wider channels generally achieve better results. Specifically, increasing channel width tends to improve performance for a fixed number of layers, whereas merely adding more layers without increasing channels does not necessarily yield gains. These observations provide practical guidance for selecting appropriate backbone and branch configurations, balancing the extraction of shared features with parameter-specific differentiation, and ensuring stable and accurate inversion results. Overall, our findings indicate that network architecture plays a crucial role in inversion performance, and that the effectiveness of architectural and reparameterization strategies may vary depending on the inversion scenario. Future work will focus on developing adaptive schemes that can automatically adjust network structures and parameterization strategies to different inversion settings, thereby further enhancing robustness and generalization.

4.3. Structure ablation of multiparameter reparameterization

Extensive ablation studies conducted in this work demonstrate that DR-FWI significantly enhances both the accuracy and stability of inversion results. Therefore, DR-FWI can be regarded as a general strategy for full waveform inversion, comparable to widely adopted techniques such as total variation regularization, multi-scale frequency continuation, and gradient post-processing. Prior studies have shown that deep reparameterization networks are highly flexible and can be integrated into both conventional

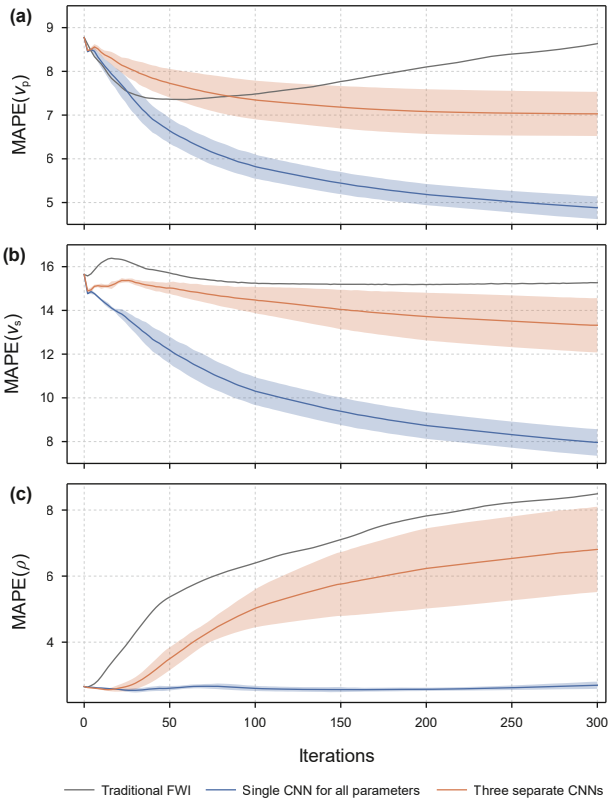

Fig. 12. Evolution of MAPE during inversion for **(a)** v_p, **(b)** v_s, and **(c)** ρ using two reparameterization strategies in multiparameter FWI. Solid lines denote the mean MAPE across different network configurations, while shaded regions represent the variability in inversion performance under each strategy.

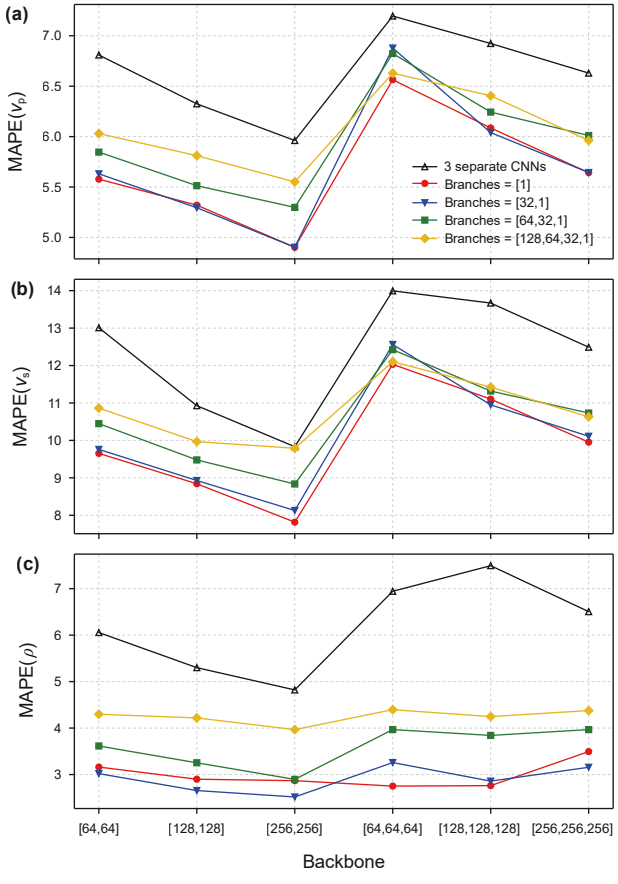

Fig. 13. Comparison of inversion performance across different backbone-branch network configurations. Each panel presents the finally MAPE for **(a)** v_p, **(b)** v_s, and **(c)** ρ, respectively. The horizontal axis denotes various backbone and branch settings, while the vertical axis indicates the corresponding MAPE.

adjoint-state-method-based FWI frameworks (He and Wang, 2021; Dhara and Sen, 2022) and automatic differentiation-based deep learning frameworks (Zhu et al., 2022; Liu et al., 2025). Owing to its high compatibility with various FWI paradigms, DR-FWI exhibits strong scalability and plug-and-play capability, facilitating seamless integration into existing inversion workflows with minimal modification.

Nonetheless, several practical limitations should be acknowledged. When incorporated into traditional FWI frameworks, the optimization of the reparameterization network typically requires GPU acceleration, whereas conventional FWI implementations primally rely on multi-core CPU clusters. The resulting data exchange between heterogeneous computing architectures may introduce additional computational overhead. In contrast, integrating DR-FWI into automatic differentiation-based frameworks allows for unified GPU computation, which is more efficient. However, the substantial memory consumption inherent to ADFWI remains a significant bottleneck, especially when applied to large-scale 3-D modeling or field data.

In terms of applicability, this study has extended the use of DR-FWI beyond its original focus on noise robustness, demonstrating its effectiveness under more challenging conditions, including sparse acquisition and multiparameter inversion. Future efforts will focus on applying DR-FWI to more complex and realistic scenarios, such as low-frequency-missing field data (Li and Demanet, 2016) and large-scale 3D FWI tasks (Ben-Hadj-Ali et al., 2008). More broadly, deep reparameterization strategies hold promises for a wide range of inverse problems, including surface wave inversion, gravity and magnetic inversion, medical imaging, and atmospheric parameter estimation. Furthermore, the demonstrated capability of DR-FWI to suppress crosstalk in multiparameter inversion opens new avenues for research in joint multiparameter and multiphysics inversion, providing valuable guidance for future theoretical and algorithmic developments.

5. Conclusions

This study systematically evaluates different deep reparameterization network architectures (UNet, CNNs, MLPs) and initial model embedding schemes (pretraining-based and direct summation) for full waveform inversion. Extensive ablation experiments show that a multi-layer CNN combined with pretraining-based embedding achieves optimal inversion accuracy and stability, providing practical design guidelines. By analyzing the spectral learning behavior of deep neural networks, we reveal that DR-FWI progressively reconstructs velocity models from low- to high-frequency components. This inherent spectral regularization improves noise robustness and enhances interpretability, underpinning DR-FWI's stable inversion performance. From an application standpoint, DR-FWI extends beyond noise robustness to effectively handle sparse observations and multiparameter inversion, reducing data acquisition demands in real-world scenarios. Furthermore, this study extends the single-parameter DR-FWI to multiparameter inversion using a backbone-branch network structure, demonstrating its capability to suppress multiparameter crosstalk, which offers promising avenues for multiphysics joint inversion research and informs future algorithmic developments.

CRediT authorship contribution statement

Feng Liu: Writing – original draft, Software, Methodology, Data curation, Conceptualization. **Ya-Xing Li**: Writing – review & editing, Methodology, Funding acquisition, Conceptualization. **Rui Su**: Writing – review & editing, Supervision, Conceptualization. **Jian-Ping Huang**: Writing – review & editing, Conceptualization. **Lei Bai**: Writing – review & editing, Supervision, Funding acquisition.

Code availability

The open-source implementation of the deep reparameterization framework for full waveform inversion (DR-FWI) used in this study is available in the public repository of the ADFWI project: https://github.com/liufeng2317/ADFWI/tree/bv1.1/examples/DR-FWI. All synthetic tests performed in this study can be reproduced using scripts provided in the examples folder.

Declaration of competing interest

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

Acknowledgements

We sincerely thank the editor, the associate editor, and two anonymous reviewers for their valuable time and for providing constructive and insightful comments that have significantly improved this study. This work was supported by the Shanghai Artificial Intelligence Laboratory, the National Natural Science Foundation of China (grant No. 42504129), the Science for Earthquake Resilience (grant No. XH24012A), and the University-Industry Collaborative Education Program (grant No. 2504244609).

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.petsci.2025.12.027.

References

Aminsadeh, F., 1996. In: 3-D Salt and Overthrust Seismic Models, in: Weimer, P., Davis, T.L. (Eds.). American Association of Petroleum Geologists. https://doi.org/10.1306/St42605C29.

Ben-Hadj-Ali, H., Operto, S., Virieux, J., 2008. Velocity model building by 3D frequency-domain, full-waveform inversion of wide-aperture seismic data. Geophysics 73 (5), VE101–VE117. https://doi.org/10.1190/1.2957948.

Bozdağ, E., Trampert, J., Tromp, J., 2011. Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements: misfit functions for full waveform inversion. Geophys. J. Int. 185 (2), 845–870. https://doi.org/10.1111/j.1365-246X.2011.04970.x.

Brossier, R., Operto, S., Jean, V., 2009. Seismic imaging of complex structures by 2D elastic frequency-domain full-waveform inversion. Geophysics 74 (6), WCC63–WCC76. https://doi.org/10.1190/1.3215771.

Chakrabarty, P., Maji, S., 2019. The Spectral Bias of the Deep Image Prior. https://doi.org/10.48550/ARXIV.1912.08905 arXiv: 1912.08905.

Chen, G., Chen, J., Jensen, K., et al., 2024a. Joint data and model-driven simultaneous inversion of velocity and density. Geophys. J. Int. 237 (3), 1674–1698. https://doi.org/10.1093/gji/ggae128.

Chen, G., Li, J., Chen, J., et al., 2025. High-precision sub-seafloor velocity building based on joint tomography and deep learning on OBS data in the South China Sea. J. Earth Sci. 36 (2), 830–834. https://doi.org/10.1007/s12583-025-0170-0.

Chen, K., Liu, L., Xu, L., et al., 2024b. Linearized waveform inversion for vertical transversely isotropic elastic media: Methodology and multi-parameter crosstalk analysis. Pet. Sci. 21 (1), 252–271. https://doi.org/10.1016/j.petsci.2024.01.002.

Choi, Y., Alkhalifah, T., 2012. Application of multi-source waveform inversion to marine streamer data using the global correlation norm. Geophys. Prospect. 60 (4), 748–758. https://doi.org/10.1111/j.1365-2478.2012.01079.x.

Dhara, A., Sen, M., 2022. Elastic-AdjointNet: A physics-guided deep autoencoder to overcome crosstalk effects in multiparameter full-waveform inversion. In: IMAGE/Technical Program Expanded Abstracts, pp. 882–886. https://doi.org/10.1190/image2022-3745050.1.

Dokter, E., Köhn, D., Wilken, D., et al., 2017. Full waveform inversion of SH- and love-wave data in near-surface prospecting. Geophys. Prospect. 65 (S1), 216–236. https://doi.org/10.1111/1365-2478.12549.

Fichtner, A., Bunge, H.-P., Igel, H., 2006. The adjoint method in seismology. Phys. Earth Planet. Inter. 157 (1–2), 86–104. https://doi.org/10.1016/j.pepi.2006.03.016.

Gauthier, O., Virieux, J., Tarantola, A., 1986. Two-dimensional nonlinear inversion of seismic waveforms: numerical results. Geophysics 51 (7), 1387–1403. https://doi.org/10.1190/1.1442188.
Gray, S.H., Marfurt, K.J., 1995. Migration from topography: improving the near-surface image. Can. J. Explor. Geophys. 31 (1–2), 18–24.
He, Q., Wang, Y., 2021. Reparameterized full-waveform inversion using deep neural networks. Geophysics 86 (1), V1–V13. https://doi.org/10.1190/geo2019-0382.1.
Hyndman, R.J., Koehler, A.B., 2006. Another look at measures of forecast accuracy. Int. J. Forecast. 22 (4), 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001.
Kalita, M., Kazei, V., Choi, Y., Alkhalifah, T., 2019. Regularized full-waveform inversion with automated salt flooding. Geophysics 84 (4), R569–R582. https://doi.org/10.1190/geo2018-0146.1.
Kingma, D.P., Ba, J., 2017. Adam: A method for stochastic optimization. arXiv: 1412.6980. https://doi.org/10.48550/arXiv.1412.6980.
Lailly, P., 1983. The seismic inverse problem as a sequence of before stack migrations. In: In: Society for Industrial and Applied Mathematics Expanded Abstracts, pp. 206–220.
Lempitsky, V., Vedaldi, A., Ulyanov, D., 2018. Deep image prior. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 9446–9454. https://doi.org/10.1109/CVPR.2018.00984.
Li, Y.E., Demanet, L., 2016. Full-waveform inversion with extrapolated low-frequency data. Geophysics 81 (6), R339–R348. https://doi.org/10.1190/geo2016-0038.1.
Lin, Y., Theiler, J., Wohlberg, B., 2023. Physics-guided data-driven seismic inversion: Recent progress and future opportunities in full-waveform inversion. IEEE Signal Process. Mag. 40 (1), 115–133. https://doi.org/10.1109/MSP.2022.3217658.
Liu, F., 2024. liufeng2317/ADFWI: Zenodo. Zenodo. https://doi.org/10.5281/ZENODO.14261243.
Liu, F., Li, H., Zou, G., et al., 2025. Automatic differentiation-based full waveform inversion with flexible workflows. J. Geophys. Res.: Mach. Learn. Comput. 2 (1), e2024JH000542. https://doi.org/10.1029/2024JH000542.
Liu, Q., Tromp, J., 2006. Finite-frequency kernels based on adjoint methods. Bull. Seismol. Soc. Am. 96 (6), 2383–2397. https://doi.org/10.1785/0120060041.
Luo, Y., Schuster, G.T., 1991. Wave-equation traveltime inversion. Geophysics 56 (5), 645–653. https://doi.org/10.1190/1.1443081.
Martin, G.S., Wiley, R., Marfurt, K.J., 2006. Marmousi2: An elastic upgrade for marmousi. Lead. Edge 25 (2), 156–166. https://doi.org/10.1190/1.2172306.
Métivier, L., Bretaudeau, F., Brossier, R., et al., 2014. Full waveform inversion and the truncated newton method: Quantitative imaging of complex subsurface structures. Geophys. Prospect. 62 (6), 1353–1375. https://doi.org/10.1111/1365-2478.12136.
Operto, S., Gholami, Y., Prieux, V., et al., 2013. A guided tour of multiparameter full-waveform inversion with multicomponent data: From theory to practice. Lead. Edge 32 (9), 1040–1054. https://doi.org/10.1190/tle32091040.1.
Rahaman, N., Baratin, A., Arpit, D., et al., 2018. On the spectral bias of neural networks. arXiv:1806.08734. https://doi.org/10.48550/ARXIV.1806.08734.
Rasht-Behesht, M., Huber, C., Shukla, K., et al., 2022. Physics-informed neural networks (PINNs) for wave propagation and full waveform inversions. J. Geophys. Res. Solid Earth 127 (5). https://doi.org/10.1029/2021JB023120 e2021JB023120.
Ren, Y., Xu, X., Yang, S., et al., 2020. A physics-based neural-network way to perform seismic full waveform inversion. IEEE Access 8, 112266–112277. https://doi.org/10.1109/ACCESS.2020.2997921.
Routh, P., Neelamani, R., Lu, R., et al., 2017. Impact of high-resolution FWI in the Western Black Sea: Revealing overburden and reservoir complexity. Lead. Edge 36 (1), 60–66. https://doi.org/10.1190/tle36010060.1.
Saadat, M., Hashemi, H., Nabi-Bidhendi, M., 2024. Generalizable data driven full waveform inversion for complex structures and severe topographies. Pet. Sci. 21 (6), 4025–4033. https://doi.org/10.1016/j.petsci.2024.05.002.
Sambridge, M., Rickwood, P., Rawlinson, N., et al., 2007. Automatic differentiation in geophysical inverse problems. Geophys. J. Int. 170 (1), 1–8. https://doi.org/10.1111/j.1365-246X.2007.03400.x.
Schmelzbach, C., Greenhalgh, S., Reiser, F., et al., 2016. Advanced seismic processing/imaging techniques and their potential for geothermal exploration. Interpretation 4 (4), SR1–SR18. https://doi.org/10.1190/INT-2016-0017.1.
Shi, Z., Mettes, P., Maji, S., et al., 2022. On measuring and controlling the spectral bias of the deep image prior. Int. J. Comput. Vis. 130 (4), 885–908. https://doi.org/10.1007/s11263-021-01572-7.
Sun, J., Innanen, K., Zhang, T., et al., 2023a. Implicit seismic full waveform inversion with deep neural representation. J. Geophys. Res. Solid Earth 128 (3), e2022JB025964. https://doi.org/10.1029/2022JB025964.
Sun, J., Innanen, K.A., Huang, C., 2021. Physics-guided deep learning for seismic inversion with hybrid training and uncertainty analysis. Geophysics 86 (3), R303–R317. https://doi.org/10.1190/geo2020-0312.1.
Sun, P., Yang, F., Liang, H., et al., 2023b. Full-waveform inversion using a learned regularization. IEEE Trans. Geosci. Rem. Sens. 61, 1–15. https://doi.org/10.1109/TGRS.2023.3322964.
Tarantola, A., 1984. Linearized inversion of seismic reflection data. Geophys. Prospect. 32 (6), 998–1015. https://doi.org/10.1111/j.1365-2478.1984.tb00751.x.
Tarantola, A., 2005. Inverse Problem Theory and Methods for Model Parameter Estimation. Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9780898717921.
Taufik, M.H., Wang, F., Alkhalifah, T., 2024. Learned regularizations for multi-parameter elastic full waveform inversion using diffusion models. J. Geophys. Res.: Mach. Learn. Comput. 1 (1), e2024JH000125. https://doi.org/10.1029/2024JH000125.
Tromp, J., 2019. Seismic wavefield imaging of earth's interior across scales. Nat. Rev. Earth Environ. 1 (1), 40–53. https://doi.org/10.1038/s43017-019-0003-8.
Virieux, J., Operto, S., 2009. An overview of full-waveform inversion in exploration geophysics. Geophysics 74 (6), WCC1–WCC26. https://doi.org/10.1190/1.3238367.
Wang, F., Huang, X., Alkhalifah, T.A., 2023. A prior regularized full waveform inversion using generative diffusion models. IEEE Trans. Geosci. Rem. Sens. 61, 1–11. https://doi.org/10.1109/tgrs.2023.3337014.
Wang, Z., Bovik, A.C., Sheikh, H.R., et al., 2004. Image quality assessment: From error visibility to structural similarity. IEEE Trans. Image Process. 13 (4), 600–612. https://doi.org/10.1109/TIP.2003.819861.
Wu, Y., Lin, Y., 2020. InversionNet: an efficient and accurate data-driven full waveform inversion. IEEE Trans. Comput. Imaging (6), 419–433. https://doi.org/10.1109/TCI.2019.2956866.
Wu, Y., McMechan, G.A., 2019. Parametric convolutional neural network-domain full-waveform inversion. Geophysics 84 (6), R881–R896. https://doi.org/10.1190/geo2018-0224.1.
Xie, C., Qin, Z.L., Wang, J.H., et al., 2024. Full waveform inversion based on hybrid gradient. Pet. Sci. 21 (3), 1660–1670. https://doi.org/10.1016/j.petsci.2024.01.013.
Yang, J., Liu, Y., Dong, L., 2014. A multi-parameter full waveform inversion strategy for Acoustic media with variable density. Chin. J. Geophys. 57 (2), 628–643. https://doi.org/10.6038/cjg20140226.
Yu, S., Ma, J., 2021. Deep learning for geophysics: Current and future trends. Rev. Geophys. 59 (3), e2021RG000742. https://doi.org/10.1029/2021RG000742.
Zhang, P., Wu, R.-S., Han, L.-G., et al., 2022. Elastic direct envelope inversion based on wave mode decomposition for multi-parameter reconstruction of strong-scattering media. Pet. Sci. 19 (5), 2046–2063. https://doi.org/10.1016/j.petsci.2022.05.007.
Zhang, Z., Lin, Y., 2020. Data-driven seismic waveform inversion: a study on the robustness and generalization. IEEE Trans. Geosci. Rem. Sens. 58 (10), 6900–6913. https://doi.org/10.1109/TGRS.2020.2977635.
Zhao, D., Zhao, F., Gan, Y., 2020. Reference-driven compressed sensing MR image reconstruction using deep convolutional neural networks without pre-training. Sens. 20 (1), 308. https://doi.org/10.3390/s20010308.
Zhu, W., Xu, K., Darve, E., et al., 2021. A general approach to seismic inversion with automatic differentiation. Comput. Geosci. 151, 104751. https://doi.org/10.1016/j.cageo.2021.104751.
Zhu, W., Xu, K., Darve, E., et al., 2022. Integrating deep neural networks with full-waveform inversion: reparameterization, regularization, and uncertainty quantification. Geophysics 87 (1), R93–R109. https://doi.org/10.1190/geo2020-0933.1.