ACCSCIENCE PUBLISHING

Journal of Seismic Exploration

ARTICLE

Intelligent universal full-waveform inversion of seismic wavefields based on multi-scale Swin Transformer transposed U-Net

Binghui Zhao[1,2], Liguo Han[3]*, Pan Zhang[3], Laiyu Lu[1,4], Xujia Shang[3], and Tongwei Qin[1,2]

[1]Institute of Geophysics, China Earthquake Administration, Beijing, China

[2]Beijing Baijiatuan Earth Sciences National Observation and Research Station, Beijing, China

[3]College of Geoexploration Science and Technology, Jilin University, Changchun, Jilin, China

[4]State Key Laboratory of Earthquake Dynamics and Forecasting, Institute of Geophysics, China Earthquake Administration, Beijing, China

Abstract

The quality of the wavelet significantly affects the accuracy of full-waveform inversion (FWI). In practice, seismic data are often contaminated by non-ideal wavelets with complex side lobes and ambiguous dominant frequencies. Directly employing such data in conventional FWI workflows, which typically assume an ideal Ricker wavelet, can induce severe cycle-skipping problems and lead to inversion failure. Existing remedies, such as sophisticated source wavelet estimation or the design of wavelet-independent objective functions, often suffer from instability in estimation or the loss of waveform details. To address this issue at its root from a data-driven perspective, this study proposes a novel deep learning-based approach. We introduce a multi-scale Swin Transformer transposed U-Net (MSTU) network to intelligently transform seismic records containing unknown, non-ideal wavelets into universal records composed solely of ideal Ricker wavelets. Through end-to-end learning, our method establishes a direct nonlinear mapping from the complex wavelet domain to the universal wavelet domain, effectively stripping the inherent wavelet ambiguity from the raw data. Numerical examples demonstrate that data universal by the MSTU network can be directly fed into standard FWI routines, yielding superior inversion results. For data originally generated with Klauder or Ormsby wavelets, the inversion results obtained after our standardization process significantly outperform those from direct inversion using the incorrect wavelet or even using the original non-ideal wavelet itself. Moreover, the accuracy approaches the theoretical upper limit achievable by inversion with the ideal Ricker wavelet as the ground truth. Requiring no prior wavelet estimation or modification to the core inversion algorithm, our method serves as an efficient, robust, and "plug-and-play" data preprocessing tool, offering a highly practical solution to enhance the reliability of FWI for field seismic data.

Keywords: Multi-scale Swin Transformer transposed U-Net; Wavelet standardization; Full-waveform inversion; Deep learning; Seismic data processing; Wavefield reconstruction

***Corresponding author:**
Liguo Han
(hanliguo@jlu.edu.cn)

Citation: Zhao B, Han L, Zhang P, Lu L, Shang X, Qin T. Intelligent universal full-waveform inversion of seismic wavefields based on multi-scale Swin Transformer transposed U-Net. *J Seismic Explor*. 2026;35(4):026030010.
doi: 10.36922/JSE026030010

Received: January 16, 2026

Revised: April 27, 2026

Accepted: May 13, 2026

Published online: June 19, 2026

Copyright: © 2026 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Full-waveform inversion (FWI) stands as one of the most promising methods for achieving high-resolution subsurface velocity modeling in geophysical exploration. Its fundamental principle involves numerically simulating seismic wave propagation through subsurface media and iteratively updating the model's parameters. This process ultimately aims to achieve optimal matching between synthetic seismic data and actual field observations across the entire waveform spectrum.[1,2] This method simultaneously utilizes both kinematic (travel time) and dynamic (amplitude, phase) information from seismic wavefields, theoretically enabling modeling accuracy and resolution far surpassing traditional ray tomography or velocity analysis methods.[3,4]

However, the success of FWI in practical applications heavily relies on accurate knowledge of the source wavelet. The source wavelet defines the initial conditions of the seismic wavefield, and its morphology, phase, and frequency band characteristics directly influence the accuracy of the synthetic data. In theoretical studies and many model experiments, the source is typically assumed to be a known, simple ideal wavelet, such as the zero-phase Ricker wavelet.[5] In actual seismic data acquisition, however, the effective source wavelet is typically unknown due to the combined effects of source excitation mechanisms, instrument response, near-surface absorption, and wave propagation paths. It often exhibits complex phase characteristics, limited bandwidth, and irregular amplitude spectrum shapes, such as those seen in the Klauder or Ormsby wavelets.[6] This discrepancy between actual and idealized wavelets constitutes a fundamental challenge in FWI applications.

Inaccurate source wavelets systematically distort the waveforms of synthetic data. The fidelity of synthetic data to observed data underpins the evaluation of FWI performance, and the morphology of these synthetic waveforms is inherently governed by the source wavelet.[7] When simulations employ erroneous wavelets, even with a perfectly accurate velocity model, non-negligible residuals persist between synthetic and observed data. This significantly increases the nonlinearity of the inversion objective function, making FWI methods prone to frequency jumps or convergence to local minima far from the true solution. Ultimately, this leads to failed inversions or severely distorted velocity models.[5,8] Therefore, effectively addressing the uncertainty of source wavelets is crucial for advancing FWI technology toward practical application.

Existing research approaches to this problem fall into two main categories. The first focuses on source wavelet estimation, attempting to directly recover the source function from seismic data. For instance, Song and Williamson[9] integrated wavelet calculations into the velocity model inversion process, but this method relies on a relatively accurate velocity model. Hu *et al.*[10] proposed an adjoint-state inversion method that updates the source wavelet simultaneously with velocity inversion. Liu *et al.*[11] investigated wavelet estimation techniques within a time-domain FWI framework, while Han *et al.*[12] explored the extraction of deep seismic wavelets from real data using generative adversarial networks. These approaches address the root cause of the problem but rely heavily on the accuracy of the initial velocity model and the signal-to-noise ratio of the data. When initial model errors are significant or noise interference is strong, wavelet estimation itself may become ill-posed, with errors propagating directly and compromising the final velocity model. The second approach aims to circumvent wavelet influence by developing "wave-independent" inversion methods. These techniques redesign the objective function to render it insensitive to source wavelet details. Key strategies include: constructing objective functions based on seismic record envelopes to exploit their insensitivity to wavelet phases for background velocity inversion;[8,9,13-20] directing envelope inversion,[21-23] constructing cross-folded target functions to eliminate common wavelet terms through mathematical transformations; or shifting to match the normalized energy spectrum between observed and synthetic data to reduce stringent phase matching requirements.[24] These approaches enhance robustness to wavelet errors to varying degrees, but typically come at a cost such as loss of high-frequency details in waveforms, reduced inversion resolution, or stability challenges when applied to complex wavefields.

Collectively, mainstream approaches, whether using wavelets with inherent estimation errors as inputs or employing inversion algorithms insensitive to waveform details, fail to fully resolve issues arising from non-ideal source wavelets in the input data. Wavelet uncertainty in the data remains a fundamental constraint limiting inversion accuracy and reliability.

To address this, this study proposes a novel technical approach: seismic wavelet standardization. The core concept is not to embed the wavelet problem within complex inversion iterations nor to require inversion algorithms to adapt to the data. Instead, it focuses on fundamentally transforming the wavelet characteristics of the data during preprocessing. We aim to establish a data-driven mapping model that directly converts seismic records containing arbitrary, unknown, non-ideal

wavelets (e.g., Klauder, Ormsby) into "universal" seismic records resembling those generated by a known, ideal Ricker wavelet excitation. This approach decouples the wavelet uncertainty problem from the inversion domain, transforming it into a data-domain transformation problem. It thereby provides pure, consistent, high-quality input for subsequent standard FWI workflows, promising to significantly reduce inversion nonlinearity and enhance solution stability and reliability.

To achieve this goal, we propose the multi-scale Swin Transformer transposed U-Net (MSTU). Conventional convolutional neural networks struggle to capture long-range dependencies in seismic data (e.g., far-offset continuity). The Swin Transformer[25-27] overcomes this via hierarchical shifted-window attention, achieving linear complexity and multi-scale feature pyramids. Building on this, MSTU integrates three core components: the Swin Transformer as the encoder backbone, transposed convolutions for learnable upsampling in the decoder, and the U-Net[28] encoder-decoder framework with skip connections to preserve multi-scale information.

Although Swin U-Net-like architectures have been explored for seismic denoising[29,30] and inversion,[31] none of these works target wavelet standardization. Our MSTU adapts the Swin U-Net framework to this specific task by introducing data-driven designs (anisotropic downsampling, non-square windows, and learnable upsampling) that respect the distinct time–space dimensions of seismic records, enabling efficient and accurate mapping from non-ideal wavelets to the ideal Ricker wavelet.

Through end-to-end supervised learning, MSTU can implicitly learn complex, nonlinear mapping rules from arbitrary non-ideal wavelets to ideal Ricker wavelets directly from massive data pairs. Seismic data processed with this intelligent normalization effectively strips ambiguity and complexity introduced by the original source wavelets, providing a robust and reliable data foundation for FWI. This significantly advances the large-scale application of this technology in processing complex real-world seismic data.

2. Theoretical principles

The accuracy of FWI fundamentally depends on the quality of seismic data, for which the morphology and properties of the source wavelet are the core determinants of the wavefield characteristics. In practical seismic exploration, influenced by the combined effects of different source excitation mechanisms, acquisition-system responses, and propagation-path effects, seismic records are often dominated by source wavelets with diverse physical characteristics. These actual wavelets often exhibit significant differences from the simple, ideal wavelets (e.g., the Ricker wavelet) typically assumed in FWI theory and algorithms, particularly in terms of phase, amplitude spectrum, and side lobe structure.[6]

To systematically clarify the challenges posed by non-ideal wavelets to inversion and to provide a clear physical and mathematical rationale for the standardization method proposed in this study, this section first presents a comparative analysis of three typical source wavelets: the Ricker wavelet, the Klauder wavelet, and the Ormsby wavelet. This analysis covers their mathematical models, time–frequency characteristics, and the resulting seismic wavefields. By dissecting how non-ideal wavelets (such as Klauder and Ormsby) introduce additional phase complexity, side lobe interference, and irregular frequency-band energy distribution into seismic records, we will reveal their specific mechanisms of impact on data consistency, waveform matching accuracy, and ultimately the reliability of the inversion solution. This analysis aims to theoretically demonstrate the necessity and urgency of performing wavelet standardization on raw seismic data.

2.1. Types and morphological characteristics of source wavelets

The mathematical representation of source wavelets is crucial for understanding their physical properties and their impact on seismic data. This study involves three typical source wavelets, whose mathematical expressions and generation principles are described below.

The Ricker wavelet is one of the most commonly used ideal wavelets in theoretical research of seismic exploration.[6] Its time-domain expression can be represented as a function of time (**Equation 1**):

$$w_{\text{Ricker}}(t) = \left[1 - 2\pi^2 f^2 t^2\right] e^{-\pi^2 f^2 t^2} \tag{1}$$

where f represents the peak frequency of the wavelet. This equation is essentially the second derivative of a Gaussian function, resulting in a symmetric main lobe and rapidly decaying side lobes. In this study, Ricker wavelets were generated using this formula with a fixed length, and an algorithm was employed to ensure that their absolute peak is precisely centered within the time sequence. This alignment process is crucial for ensuring timing consistency in subsequent waveform matching and inversion.

The Klauder wavelet is intrinsically linked to vibroseis technology. It is not directly defined by a single analytical expression but is generated from the autocorrelation

function of a linear-sweep signal.[6] The frequency of the sweep signal $s(t)$ varies linearly from f_{min} to f_{max}, and its mathematical expression is shown in **Equation 2**:

$$s(t) = \sin(2\pi(f_{min}t + \frac{(f_{max} - f_{min})t^2)}{2T}, 0 \le t \le T \quad (2)$$

where T is the sweep duration. The Klauder wavelet is then obtained as the autocorrelation of this sweep signal (**Equation 3**):

$$w_{Klauder}(\tau) = \int_0^T s(t)s(t+\tau)dt \quad (3)$$

This process yields a wavelet with a bandpass frequency spectrum corresponding to the sweep frequency band.

Ormsby wavelet is the impulse response, in the time domain, of a trapezoidal bandpass filter.[6] It is uniquely defined by four frequency parameters (f_1, f_2, f_3, f_4, where $f_1 < f_2 < f_3 < f_4$). Its time-domain expression is composed of a combination of squared sinc functions at four different frequencies (**Equation 4**):

$$w_{Ormsby}(t) = \frac{(\pi f_4)^2 \mathrm{sinc}^2(\pi f_4 t)}{\Delta f_{34}} - \frac{(\pi f_3)^2 \mathrm{sinc}^2(\pi f_3 t)}{\Delta f_{34}} - \frac{(\pi f_2)^2 \mathrm{sinc}^2(\pi f_2 t)}{\Delta f_{12}} + \frac{(\pi f_1)^2 \mathrm{sinc}^2(\pi f_1 t)}{\Delta f_{12}} \quad (4)$$

where $\Delta f_{34} = \pi f_4 - \pi f_3$, $\Delta f_{12} = \pi f_2 - \pi f_1$, and $\mathrm{sinc}(x) = \sin(x)/x$. This expression directly corresponds to the inverse Fourier transform of a trapezoidal amplitude spectrum.

The morphological differences can be anticipated directly from the form of their mathematical expressions. The Ricker wavelet features a concise formula and is an ideal wavelet widely used in theoretical research. Its time-domain waveform (Figure 1A) exhibits a symmetric main lobe, characterized by amplitude modulation at a single dominant frequency, with the fastest side lobe decay among the three. This simple morphology grants it superior temporal resolution, making the reflection events it constitutes more readily distinguishable from one another.[6]

In contrast, both the Klauder wavelet (Figure 1B) and the Ormsby wavelet (Figure 1C) display typical non-ideal characteristics. The time-domain waveform of the Klauder wavelet features a broad, flat main lobe flanked by numerous prominent, slowly decaying side lobes. The Ormsby wavelet, defined by four cutoff frequencies that shape its spectrum, presents an even more complex time-domain waveform. Its main lobe is less prominent in energy and is enveloped by a series of side lobes with comparable amplitudes and persistent oscillations.[6]

These two non-ideal wavelets share common detrimental features: wide main lobes and strong side lobes in the time domain, coupled with either a lower dominant frequency or complex bandpass characteristics in the frequency domain. These attributes are the root causes of degraded quality in the seismic wavefields they generate. Consequently, they underscore the necessity for the wavelet standardization processing proposed in this study.

2.2. Analysis of seismic wavefield characteristics generated by different wavelets

The morphological characteristics of the source wavelet, through convolution with the subsurface reflectivity series, directly determine the final wavefield expression in the seismic record. Figure 2 presents the synthetic seismic sections generated using the three aforementioned wavelets, all based on an identical subsurface reflectivity model.

As shown in Figure 2A, the wavefield generated by the ideal Ricker wavelet features sharp, clear reflection events with unambiguous onsets. The waveforms from individual layers are well-separated, resulting in a relatively clean wavefield structure. This high-quality wavefield provides reliable input for subsequent migration imaging or FWI.

In contrast, the wavefields constructed by the Klauder wavelet (Figure 2B) and the Ormsby wavelet (Figure 2C) exhibit significant drawbacks. Due to their inherently broad main lobes and strong side lobe energy, the seismic response from each subsurface interface is significantly temporally stretched and accompanied by pronounced ringing effects. This causes the reflection signals from adjacent thin layers to overlap and interfere, severely degrading the vertical resolution of the seismic data for these layers. Concurrently, the complex side lobe oscillations obscure the primary energy of the reflections, rendering the events blurry and less continuous, thereby hampering precise travel time picking or phase tracking. Furthermore, the strong side lobe energy creates persistent interference before and after major reflections, causing crosstalk between different wave types (e.g., primaries and multiples). This effectively elevates the background noise level of the wavefield, leading to an overall decrease in the signal-to-noise ratio. Figure 2 illustrates typical input–label pairs from our training dataset. The same qualitative characteristics are observed in the test dataset, ensuring a fair evaluation.

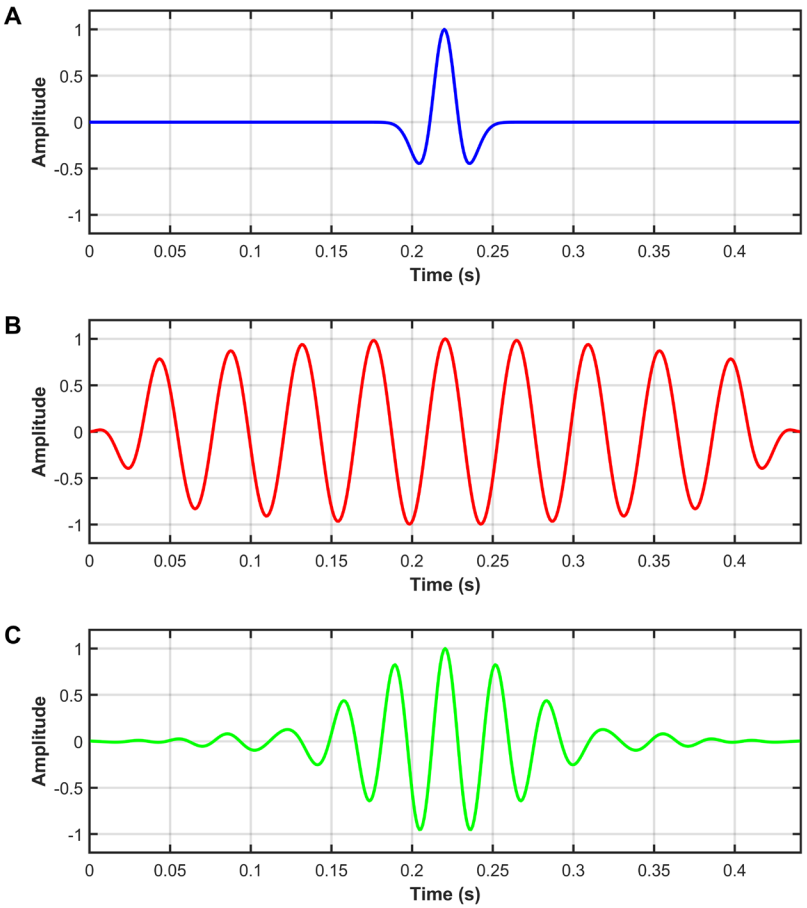


Figure 1. Three types of seismic wavelets. (A) Ricker wavelet ($f = 20\,hz$), (B) Klauder wavelet ($f_{min} = 10\,hz$, $f_{max} = 20\,hz$), and (C) Ormsby wavelet ($f_1 = 5\,hz$, $f_2 = 10\,hz$, $f_3 = 15\,hz$, $f_4 = 20\,hz$).

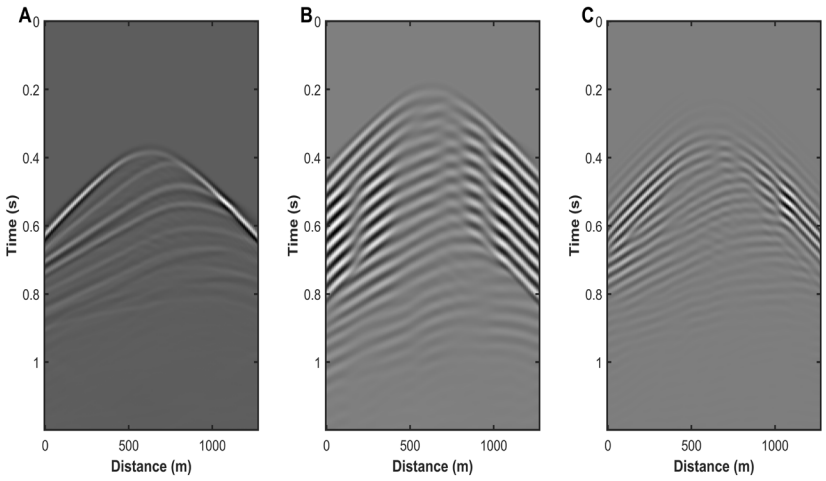


Figure 2. Synthetic seismic sections generated with different wavelets based on the same subsurface model. (A) Ricker wavelet, (B) Klauder wavelet, and (C) Ormsby wavelet.

In summary, non-ideal wavelets, such as Klauder and Ormsby, due to their broad main lobes, strong side lobes, and dominant low-frequency energy, systematically degrade the quality of the seismic wavefield. This manifests as reduced vertical resolution, blurred reflection events, and an overall decline in the signal-to-noise ratio.

When seismic data with such complex wavelet characteristics is directly applied to FWI, which is highly sensitive to waveform details, the systematic waveform distortion introduced by the wavelet morphology nonlinearly couples with errors in the subsurface model parameters. This coupling severely distorts the gradient direction, disrupts the inversion optimization path, and makes the algorithm highly prone to converging to a local minimum far from the true solution, or even failing entirely.

Therefore, before feeding raw data into the inversion workflow, developing a method that intelligently transforms data containing non-ideal wavelets into data equivalent to that generated by an ideal Ricker wavelet is a fundamental strategy for enhancing the stability, convergence, and ultimate accuracy of FWI at the data source. The intelligent wavelet standardization method proposed in this study, based on the MSTU, is specifically designed to systematically address this core bottleneck.

2.3. Seismic wavelet standardization network architecture based on multi-scale Swin Transformer transposed U-Net

To achieve the transformation of seismic records from non-ideal wavelets (e.g., Klauder, Ormsby) to those corresponding to an ideal Ricker wavelet, this study constructed a Swin Transformer-based U-Net network (Figure 3). This architecture integrates the transformer's global contextual awareness for long-sequence modeling with the U-Net's advantage in multi-scale feature fusion. It is specifically optimized for seismic data characteristics, aiming to achieve high-fidelity, end-to-end wavelet standardization mapping.

The network adopts an overall encoder-decoder structure. The input is a two-dimensional seismic record (time × trace number) containing the effects of non-ideal wavelets, and the output is the corresponding ideal Ricker wavelet record. Given that seismic records typically have long time sequences (e.g., 1,200 samples), directly applying a standard transformer would impose a prohibitive computational burden. Therefore, the network first performs initial downsampling through two specialized convolutional layers: the first layer performs a four downsampling in the time dimension (using a 4 × 1 kernel and stride), and the second layer performs a two downsampling in the spatial (trace) dimension (using a 1 × 2 kernel and stride). This design transforms the input from its original dimensions (e.g., 1,200 × 128 × 1) to a more manageable size (e.g., 300 × 64 × 64), significantly reducing the computational complexity for subsequent transformer blocks while preserving key wavefield features. This rectangular kernel design, tailored to the distinct physical meanings and resolution requirements of the time and spatial dimensions in seismic data, proves more effective for feature extraction compared to conventional square kernels.

The core of the network consists of multiple stacked Swin Transformer blocks. Each block employs alternating window-based multi-head self-attention and shifted window-based multi-head self-attention modules. In window attention, the feature map is partitioned into non-overlapping local windows (this network adopts rectangular 8 × 4 windows suited to the seismic data shape). Attention calculation is confined to each window, making the computational complexity linear rather than quadratic in the sequence length. Shifted window attention then applies a window configuration that is shifted relative to the previous layer within adjacent blocks, breaking the isolation between fixed windows and enabling cross-window information exchange and global context capture. Each attention layer is preceded and followed by layer normalization and incorporates residual connections to ensure training stability, and is then followed by a two-layer multilayer perceptron for nonlinear feature transformation. Notably, the network introduces a learnable relative-position bias table into the attention mechanism. This table captures biases arising from the relative positions of queries and keys, enabling the network to explicitly leverage the prior knowledge of sequentiality and proximity inherent in seismic signals across time and space.[27]

The encoder path comprises four stages, forming a hierarchical feature pyramid. Each stage contains two Swin Transformer blocks (employing regular and shifted window configurations, respectively) followed by a patch-merging operation. Patch merging reduces spatial dimensions by half via 2 × 2 downsampling while doubling the number of channels, progressively expanding the model's receptive field. This enables the network to integrate and abstract local waveform features from shallow layers (e.g., wavelet side lobe shapes) into more global semantic features in deeper layers (e.g., the overall structure of reflection sequences and primary-event trends). A bottleneck layer sits between the encoder and decoder, employing deep transformer blocks to capture the richest contextual information.

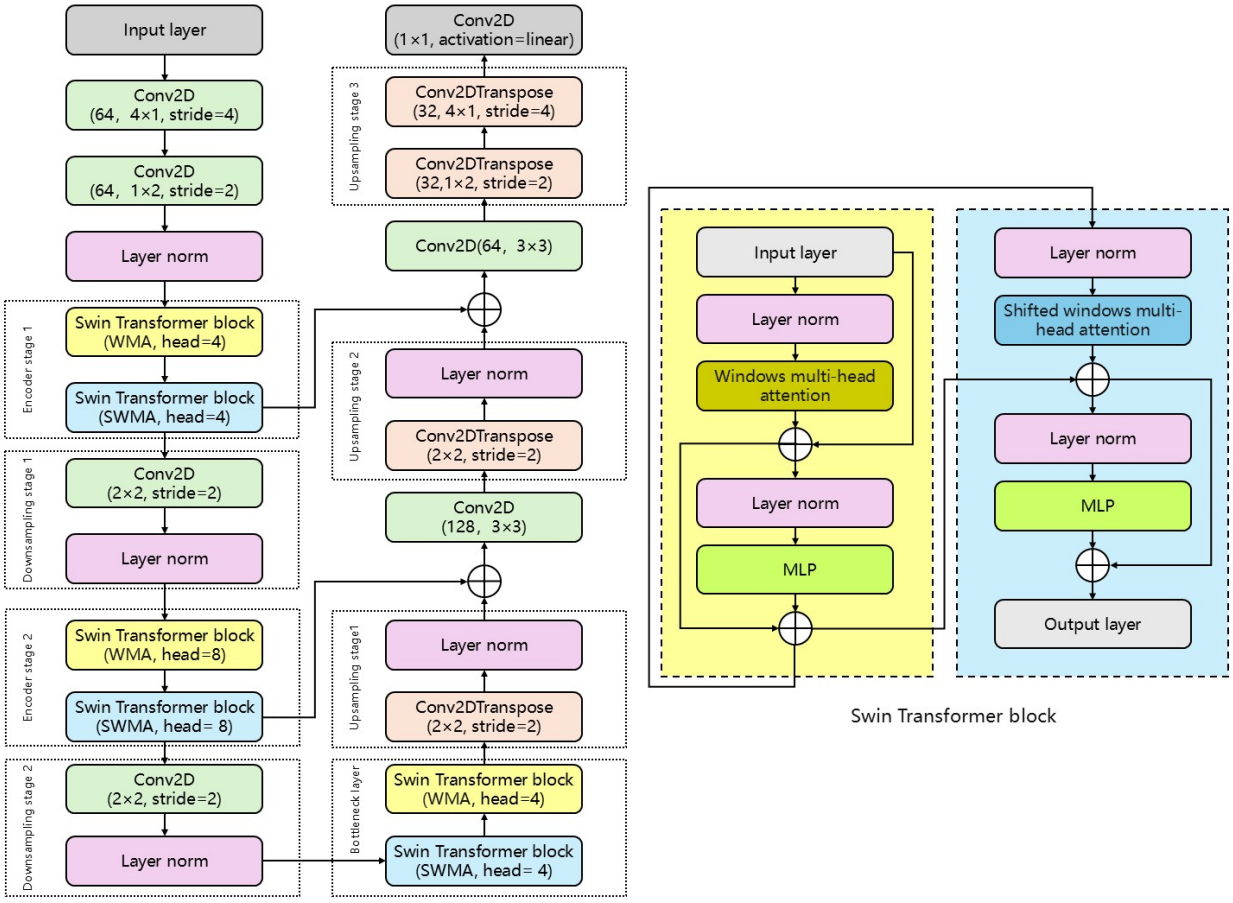


Figure 3. Multi-scale Swin Transformer transposed U-Net (MSTU)
Abbreviations: 2D: Two-dimensional; MLP: Multilayer perceptron; Norm: Normalization; SWMA: Shifted window-based multi-head self-attention; WMA: Window-based multi-head self-attention.

The decoder path symmetrically uses transposed convolutions for upsampling. Crucially, skip connections concatenate and fuse feature maps from corresponding stages of the encoder with those in the decoder, thereby achieving complementary multi-scale information and recovering fine waveform structures.

Regarding parameter configuration, the network employs a progressive design: shallow stages use fewer attention heads (4 heads) and feature channels (64) to focus on local details; as the network deepens, the number of attention heads (increasing to 16 heads) and feature channels (increasing to 256) are increased correspondingly to capture more complex global patterns.[25] Window sizes are dynamically adapted to the feature map dimensions, using an asymmetric rectangular design to match the temporal and spatial characteristics of seismic data. The shift size is set to half of the window dimension to ensure sufficient cross-window interaction. The training process uses the Adam optimizer with an initial learning rate of 1×10^{-4}, coupled with an adaptive learning rate decay strategy based on validation loss plateaus. The loss function employs mean squared error to directly constrain point-wise amplitude consistency between the output waveform and the target Ricker wavelet waveform. Data is randomly shuffled before training to enhance generalization, and early stopping is employed to prevent overfitting.

This network architecture incorporates several specialized designs tailored to the characteristics of seismic data. First, efficient processing of long seismic sequences is achieved through initial downsampling and the windowed attention mechanism. Second, cross-scale feature fusion from local waveform details to the global wavefield structure is achieved via the U-Net's skip-connection architecture. Third, the entire model is trained

end-to-end, requiring no handcrafted features or complex preprocessing pipelines, directly learning the mapping function from the complex wavelet domain to the ideal wavelet domain. Through these designs, the network can effectively avoid the wavelet ambiguity and complexity inherent in raw seismic records, providing stable, purified, high-quality input data for subsequent FWI.

2.4. Full-waveform inversion

Full-waveform inversion is one of the most advanced inversion techniques in geophysical exploration for obtaining high-resolution subsurface models of medium parameters (e.g., velocity, density). Its core principle is to use numerical simulation methods to iteratively update the subsurface model parameters, aiming for a globally optimal match between synthetic and field-observed seismic records across the entire waveform (including amplitude, phase, and travel time).[1,2]

For an acquisition system consisting of N_s sources and N_r receivers, the objective function for time-domain FWI is defined as the L2-norm of the residuals between the synthetic and observed data (**Equation 5**):

$$\Phi(m)=\frac{1}{2}\sum_{i=1}^{N_s}\sum_{j=1}^{N_r}\int_0^T\left\|u\left(m,t;x_s^{(i)},x_r^{(j)}\right)-d\left(t;x_s^{(i)},x_r^{(j)}\right)\right\|^2 dt \quad (5)$$

where m represents the subsurface medium parameter model, u is the seismic wavefield forward-modeled based on the current model m, d is the actual observed data, the summation iterates over all source points x_s and receiver points x_r, and T is the recording duration. The objective of FWI is to find the model m that most closely approximates the true subsurface structure by minimizing the objective function $\Phi(m)$.

To update the model, it is necessary to compute the gradient of the objective function with respect to the model parameters, $\nabla_m\Phi$. Direct calculation of the Fréchet derivative (i.e., the sensitivity kernel of the wavefield to model parameters) is computationally prohibitive.[1,2] Therefore, the adjoint-state method is widely employed for efficient gradient computation. The essence of this method involves performing one forward propagation (from the source to compute the synthetic wavefield) and one backward propagation (using the data residual as a virtual source to propagate the adjoint wavefield). The final gradient is then obtained via the zero-lag cross-correlation of the forward wavefield and the adjoint wavefield (**Equation 6**):

$$\nabla_m\Phi\approx-\frac{2}{v^3}\int_0^T\frac{\partial^2 u}{\partial t^2}\cdot\bar{u}dt \quad (6)$$

where $\bar{u}$ represents the adjoint wavefield. Once the gradient is obtained, nonlinear optimization algorithms (such as the conjugate gradient method or limited-memory Broyden–Fletcher–Goldfarb–Shanno) are used to determine the direction and step size of the model update for iterative refinement.

However, the successful application of FWI faces two major challenges: initial model dependence and cycle-skipping.[1,2] Due to the highly non-convex nature of the objective function, it contains multiple local minima. If the initial velocity model deviates significantly from the true model, inversion based on local optimization strategies is highly prone to converging to an erroneous local solution, leading to cycle-skipping: a phenomenon where the synthetic waveform incorrectly matches an adjacent cycle of the observed waveform.[1,2]

A key strategy to mitigate this problem is multi-scale inversion, which involves progressing from low-frequency to high-frequency data in a stepwise manner. Low-frequency data correspond to the long wavelengths of seismic waves, which are sensitive to the background velocity (macro-structure). Moreover, the objective function associated with low frequencies is smoother, facilitating convergence toward the vicinity of the global minimum. Building upon this recovered background model, higher-frequency data are then incorporated to recover the short-wavelength details (fine-scale structures) of the velocity model.[1,2] The FWI implementation in this study adopts this strategy.

The FWI conducted in this study is based on the two-dimensional acoustic wave equation and encompasses the core components of modeling, inversion, and optimization. The inversion workflow begins with constructing an initial velocity model, typically a linearly increasing gradient model from shallow to deep, as the starting point. The crucial multi-scale inversion strategy is implemented through frequency band segmentation: first, low-frequency band data (e.g., around 2 Hz) is selected for iterative inversion to recover the large-scale velocity background; subsequently, the process transitions to a high-frequency band (e.g., around 25 Hz) to further update the model and delineate finer structural details. A fixed number of iterations is allocated to each frequency band, and the process proceeds sequentially.

During each iteration, the following steps are performed for each source gather. First, forward modeling is performed using the current velocity model to generate synthetic seismic records. Next, the residual between the synthetic record and the observed data within the corresponding frequency band is computed. Finally, the contribution of this source's data to the velocity model

gradient is calculated using the adjoint-state method. The gradients from all sources are accumulated and then processed with preprocessing steps, such as illumination compensation. Subsequently, the model is updated using an adaptive step-size strategy that is progressively adjusted throughout the iterations. To maintain the stability of the inversion process, the model updates are performed within the logarithmic velocity domain.

Throughout the inversion, a time-aligned Ricker wavelet is employed as the source wavelet to ensure temporal consistency of the main wavelet peak, thereby reducing errors introduced by wavelet phase misalignment.

In summary, FWI is an effective method for achieving detailed subsurface imaging through iterative optimization that progressively minimizes the discrepancy between synthetic and observed data. The quality of its inversion results depends critically on the quality of low-frequency data, the reasonableness of the initial model, and the accuracy of the source wavelet. The "seismic wavelet standardization" preprocessing method proposed in this study aims to provide FWI with observed data composed of ideal Ricker wavelets, offering higher consistency. By doing so, it mitigates the interference from non-ideal wavelets at the data source, thereby enhancing the stability and accuracy of multi-scale inversion.

3. Methodology

To validate the effectiveness of the proposed seismic wavelet standardization network based on the MSTU, systematic numerical experiments were designed and conducted. The experiments primarily consist of two parts: first, to evaluate the network's capability to standardize seismic records with different non-ideal wavelets; second, to apply the universal data to FWI to examine its practical contribution to improving inversion accuracy.

3.1. Wavefield processing

Training deep learning models typically requires massive amounts of data, whereas high-fidelity numerical simulation in geophysics is computationally expensive. To construct a dataset that, while limited in size, is highly relevant, we selected the classic overthrust velocity model as the foundation. By slicing the model horizontally at fixed intervals (every 15 grid points), we obtained a series of two-dimensional velocity profiles that are structurally similar but differ in detail. The first sliced velocity model was designated as the test set and kept completely isolated from the training process. The remaining velocity models were used to generate the training set. This data construction strategy ensures that the test model shares macro-geological features with the training models while maintaining independence, effectively testing the model's generalization capability under a "small-sample" condition—its ability to handle unseen velocity structures similar to those in the training data.

Our computing hardware consists of a Quadro RTX 5000 with 16 GB of memory. We used 17 speed models for training, resulting in a training dataset of 2,176 samples. We used one test model and a test dataset of 128 samples. During training, the batch size was set to 4, and each iteration took approximately 80 seconds. The total number of iterations was 200. Finally, the wavefield transformation of the test data took only three seconds.

Using the velocity models described above, we employed Klauder and Ormsby wavelets as the sources for forward modeling, generating synthetic seismic records containing complex, non-ideal wavelet effects as input to the network. Feeding the test set data into the trained MSTU network yielded the standardization results, as shown in Figures 4 and 5.

Figures 4A and 5A display the single-shot records generated by the original Klauder wavelet and Ormsby wavelet, respectively, on the test velocity model. It can be observed that due to the intrinsic broad main lobe and strong side lobes of these wavelets, the reflection events in the wavefield appear relatively blurred, with reflections from different layers intertwining, resulting in a lower signal-to-noise ratio.

In contrast, Figures 4B and 5B show the corresponding network-processed results. After standardization, the waveforms in the seismic records are significantly sharpened, exhibiting a morphology similar to that produced by an ideal Ricker wavelet excitation: the reflection events become clear and sharp, the onset times of reflection signals are well-defined, and the wavefield structure appears much cleaner, making it more suitable for velocity model building.

To more finely contrast the waveform changes, we extracted single-trace records (as shown in Figures 6A, 6B, 7A, and 7B). Figure 6A compares the waveforms of the original Klauder wavelet record and the ideal Ricker wavelet record at the same receiver trace. It is evident that the Klauder wavelet exhibited complex oscillations with a non-prominent main peak, which differs significantly from the Ricker wavelet. Figure 6B, in contrast, compares the network-processed waveform with the Ricker wavelet waveform. The two demonstrated high consistency in their primary phase and amplitude characteristics. The same phenomenon was observed in the comparison for the Ormsby wavelet (Figure 7A and 7B). This visually and quantitatively demonstrates that our network can

effectively learn and strip away the complex characteristics of Klauder and Ormsby wavelets, transforming them into high-quality, universal wavefields that closely approximate the Ricker wavelet. This process provides a superior data foundation for subsequent inversion.

3.2. Comparison of full-waveform inversion

To quantitatively assess the practical improvement in FWI accuracy afforded by wavelet standardization, a comparative experiment was designed. The true velocity model (i.e., the test set model) and the initial velocity model used for the inversion are shown in Figure 8. The initial model was constructed as a linear gradient model: the shallow velocity was set to a constant (consistent with the average shallow velocity of the true model), and the deep velocity increased linearly from a minimum to a maximum. This model contains only a rough velocity trend, devoid of any specific structural information, thereby providing a strict test of FWI's ability to recover details from the data. The grid spacing was set to 10 m, the sampling interval to 0.001 seconds, with a total of 22 shots participating in the inversion.

To address the limitations imposed by the initial model on the inversion results,[2,32] a multi-scale iterative strategy was employed.[5,33] First, low-frequency (2 Hz) data with a Ricker wavelet were used uniformly to perform 15 iterations on the initial model, establishing a relatively reasonable macro-velocity background. Building upon this, we compared high-frequency (25 Hz) inversion results under four different data conditions, each undergoing 10 iterations. Scenario A (known wavelet): Using the original data containing non-ideal wavelets (Klauder or Ormsby), with the FWI forward simulation accurately employing the corresponding true wavelet. This represents a theoretically ideal but often impractical scenario. Scenario B (wavelet mismatch): Using the original data containing non-ideal wavelets, but with FWI forward simulation incorrectly using a Ricker wavelet of similar dominant frequency. This simulates the common "wavelet mismatch" scenario in practical exploration when wavelet complexity is ignored. Scenario C (ideal standard): Using synthetic data directly generated with the ideal Ricker wavelet (and inverting with the same Ricker wavelet). This serves as the theoretical benchmark for inversion. Scenario D (our method): Using the wavefield universal by the neural network as the observed data, with FWI forward simulation employing the standard Ricker wavelet.

Figures 9 and 10 show the FWI result comparisons for the Klauder and Ormsby wavelet datasets, respectively. The comparative analysis reveals that Scenario A (known wavelet) yielded relatively poor inversion results, largely due to the inherent complexity of the non-ideal source wavelet. Scenario B (wavelet mismatch) produced the worst results, exhibiting significant structural distortions, spurious artifacts, and severe cycle-skipping phenomena. This verifies the detrimental impact of using an incorrect wavelet in FWI. Scenario C (ideal standard) served as the benchmark and achieved high-quality inversion results.

Crucially, the inversion results of Scenario D (our method) were visually and quantitatively very close to those of Scenario C (ideal standard) and were significantly better than both Scenarios A and B. This indicates that our wavelet standardization network effectively serves as a wavefield reconstruction front end. It does not require prior knowledge of the precise wavelet morphology of the input data, yet it successfully transforms the data into a high-quality form that can be accurately inverted by a standard FWI workflow assuming a Ricker wavelet.

The experimental results strongly demonstrate that the proposed method can significantly mitigate FWI instability caused by inaccurate or non-ideal source wavelets, without requiring complex wavelet estimation procedures or specialized inversion algorithms. In practical scenarios where the source wavelet is unknown, this approach substantially enhances both the accuracy and reliability of the inversion.

3.3. Tests on noise robustness and model generalization

To comprehensively evaluate the network's robustness and applicability, we conducted two additional tests: processing noisy data and generalization across models. The tests employed the Ormsby wavelet and a completely new velocity model that was not involved in any previous network training (Figure 11A). To ensure a fair evaluation, the network was retrained using sliced data from all other velocity models, excluding this test model.

First, random noise was added to the synthetic data to simulate real-world interference. As shown in Figure 12, the reflection events in the noisy Ormsby wavelet wavefield were severely obscured. The network's output (Figure 12B) demonstrated that it not only converted the wavelet to the standard Ricker wavelet form but also effectively suppressed random noise. The processed wavefield exhibits clear reflection signals with significantly reduced side lobe interference and an improved signal-to-noise ratio.

Subsequently, we systematically compared the outcomes of three FWI strategies based on this noisy test dataset (Figure 11). Specifically, Figure 11B presents the result of directly inverting the noisy original Ormsby wavelet data using an erroneous Ricker wavelet; Figure 11C shows the benchmark result obtained by inverting clean, noise-free

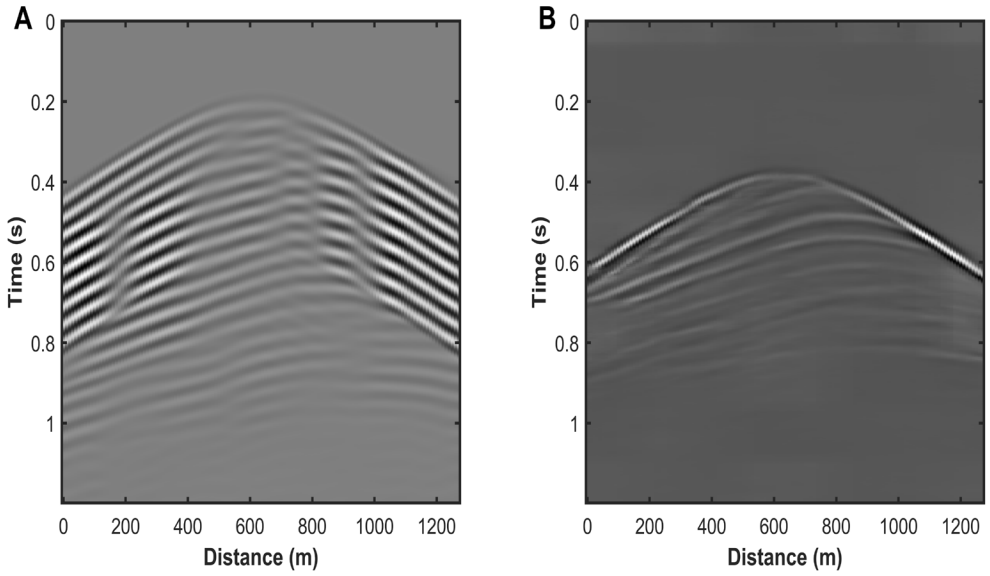


Figure 4. Standardization effect of Klauder wavelet data based on multi-scale Swin Transformer transposed U-Net (MSTU). (A) Original single-shot record with Klauder wavelet. (B) Single-shot record after MSTU standardization.

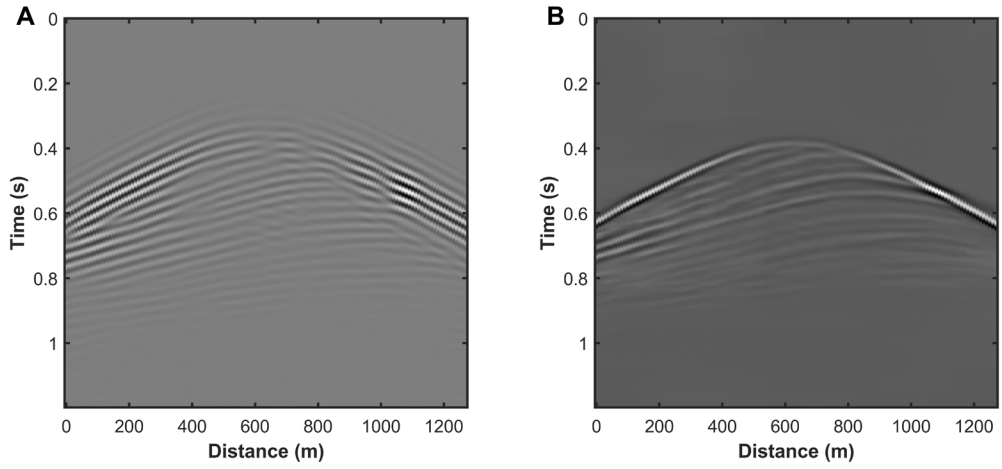


Figure 5. Standardization effect of Ormsby wavelet data based on multi-scale Swin Transformer transposed U-Net (MSTU). (A) Original single-shot record with Ormsby wavelet. (B) Single-shot record after MSTU standardization.

data generated with an ideal Ricker wavelet; and Figure 11D displays the result of inverting the noisy original data after it was universalized by the MSTU network, again using the Ricker wavelet. The inversion results clearly demonstrate that the model in Figure 11B, suffering from severe wavelet mismatch and noise interference, exhibits noticeable layer misalignment and structural distortion. In contrast, the inversion result in Figure 11D (our method) shows high agreement with Figure 11C, the ideal noise-free benchmark, in structural geometry, interface continuity, and background velocity recovery. This comparison not only verifies the effectiveness of our network in performing wavelet standardization under strong noise conditions but, more importantly, demonstrates its ability to maintain stable, reliable generalization performance when processing entirely new, unseen velocity models. This provides compelling experimental support for the potential application of this method in practical, unexplored survey areas.

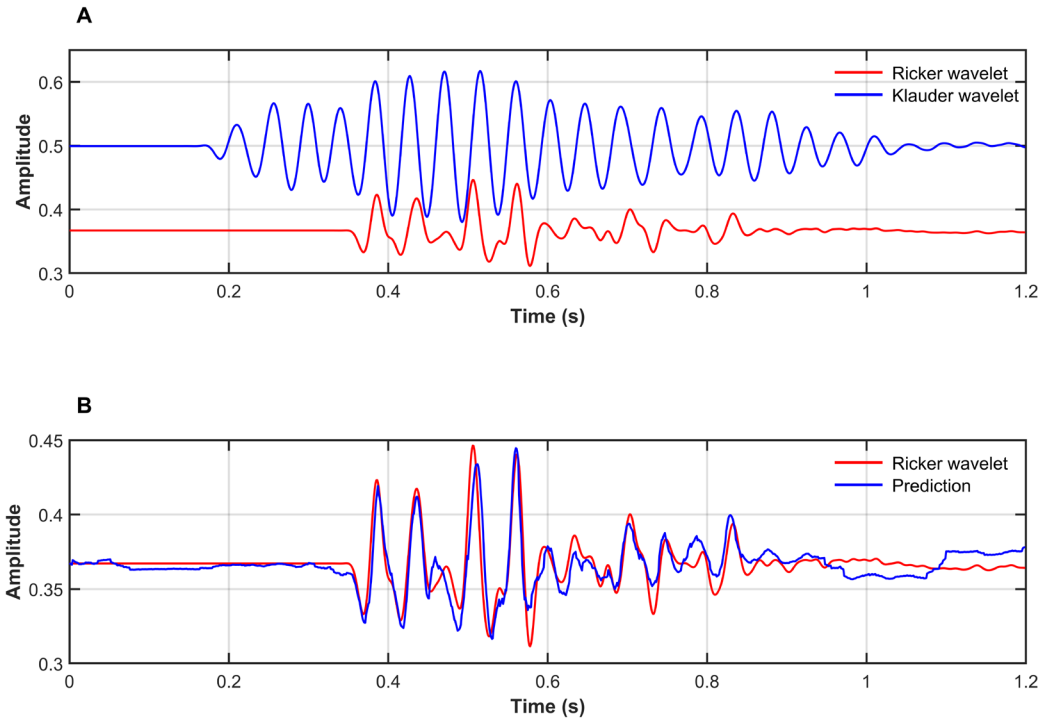


Figure 6. Waveform comparison for Klauder wavelet standardization on trace 64 of shot 64 (test data). (A) Comparison between the original Klauder wavelet and the ideal Ricker wavelet. (B) Comparison between multi-scale Swin Transformer transposed U-Net-universal waveform and ideal Ricker wavelet.

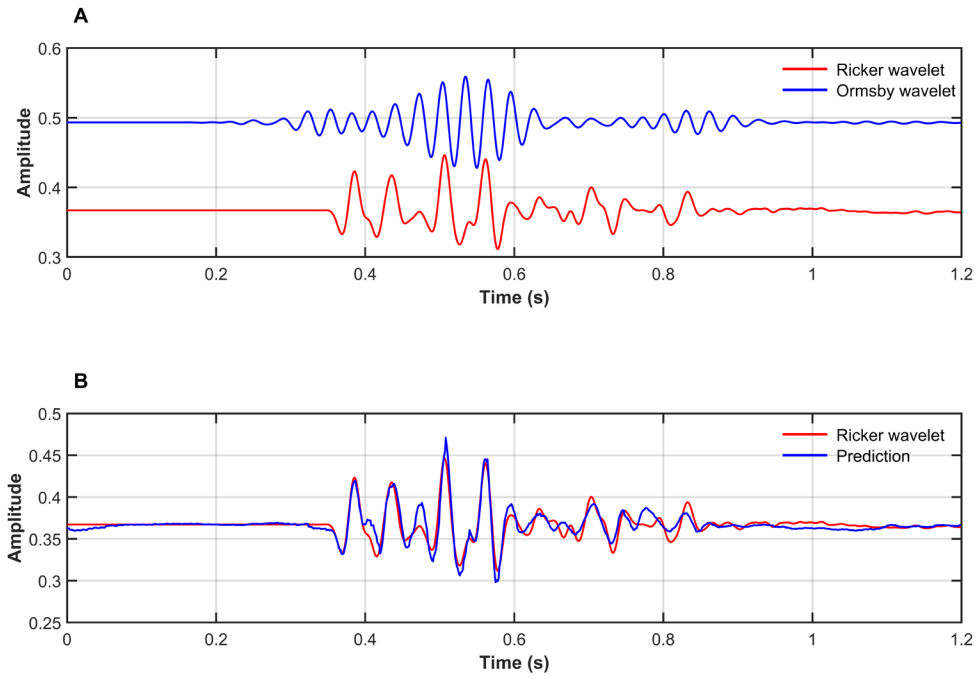


Figure 7. Waveform comparison for Ormsby wavelet standardization on trace 64 of shot 64 (test data). (A) Comparison between the original Ormsby wavelet and the ideal Ricker wavelet. (B) Comparison between multi-scale Swin Transformer transposed U-Net-universal waveform and ideal Ricker wavelet.

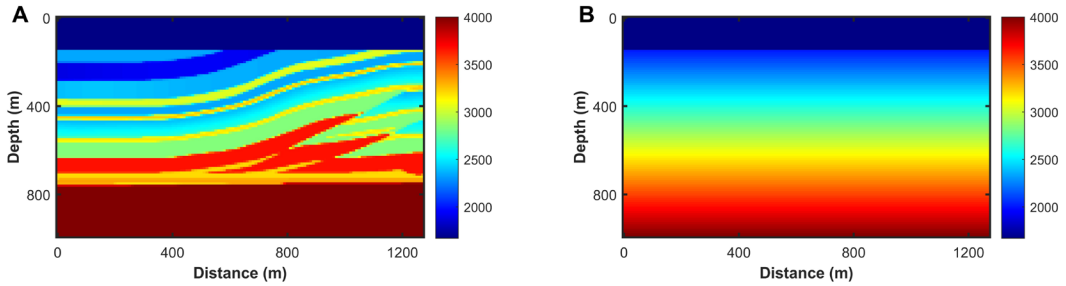


Figure 8. True and initial velocity models used for full-waveform inversion. (A) True velocity model. (B) Initial velocity model (linear gradient model).

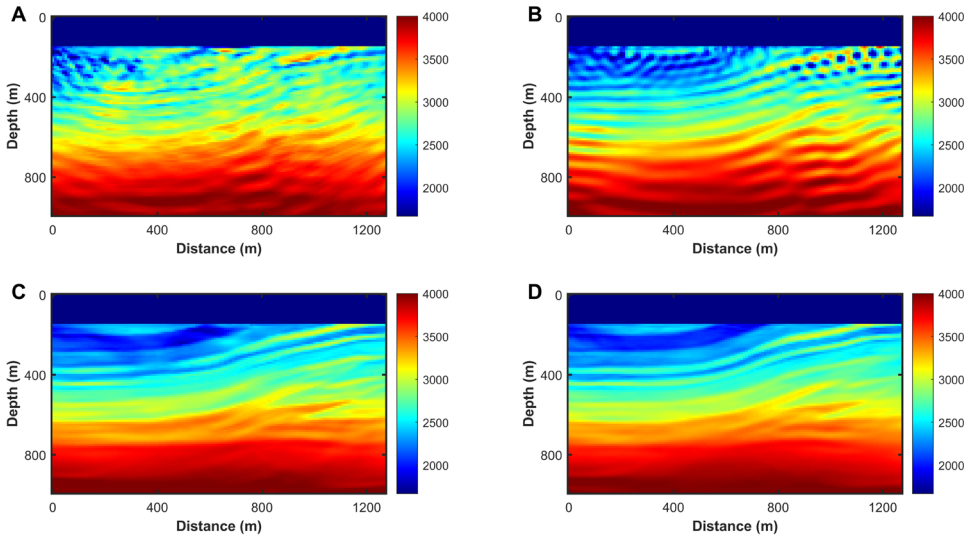


Figure 9. Comparison of full-waveform inversion results using Klauder wavelet data. (A) Scenario A: Inversion using the true Klauder wavelet. (B) Scenario B: Erroneous inversion of Klauder data using a mismatched Ricker wavelet. (C) Scenario C: Inversion using ideal Ricker wavelet data (benchmark). (D) Scenario D: Our method (Inversion of universal data using the Ricker wavelet).

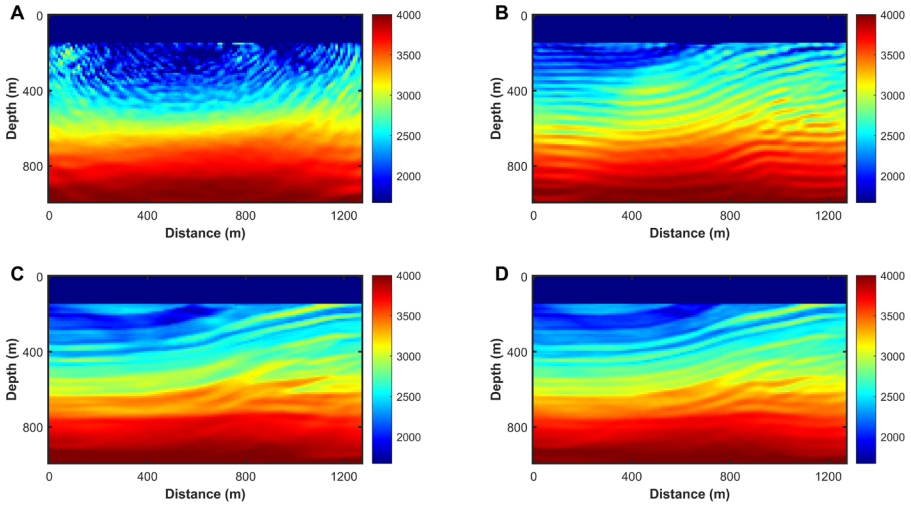


Figure 10. Comparison of full-waveform inversion results using Ormsby wavelet data. (A) Scenario A: Inversion using the true Ormsby wavelet. (B) Scenario B: Erroneous inversion of Ormsby data using a mismatched Ricker wavelet. (C) Scenario C: Inversion using ideal Ricker wavelet data (benchmark). (D) Scenario D: Our method (Inversion of universal data using the Ricker wavelet).

4. Results and discussion

4.1. Seismic-oriented network architecture design

Multi-scale Swin Transformer transposed U-Net introduces rectangular windows and asymmetric convolutional kernels, a design rooted in the inherent physical characteristics of seismic data: the number of time samples is usually much larger than the number of receiver traces. Traditional square windows or square kernels compress both the time and space dimensions simultaneously, leading to loss of vertical continuity information or computational redundancy. In contrast, rectangular windows and asymmetric kernels can independently handle the long time axis and the short spatial axis, preserving the original features of the seismic wavefield to a greater extent. Moreover, the network depth, number of attention heads, and number of channels were optimized through extensive parameter tuning to balance model expressiveness and GPU memory (Quadro RTX 5000, 16 GB), yielding the optimal configuration.

4.2. Noise robustness analysis

To evaluate the robustness of MSTU under noisy conditions, we added Gaussian noise of different intensities to the test inputs and computed the signal-to-noise ratio (**Equation 7**), including both the signal-to-noise ratio of the test data relative to the label and the signal-to-noise ratio of the prediction relative to the label (Table 1). Experiments show that even in the presence of strong noise, the network can effectively suppress noise and recover the main reflection-wavefield morphology, demonstrating MSTU's robustness to noise. This benefits from the implicit suppression of local noise by the Swin Transformer's window attention mechanism and from faithful multi-scale feature transmission via skip connections. Further analysis of noise robustness (Figure 13) reveals that when the noise level reaches 0.1, the input waveform becomes almost invisible; nevertheless, the network can still effectively suppress the noise and successfully standardize the wavefield. However, when the noise level increases to 0.5, although the network still suppresses most of the noise, the underlying waveform is severely damaged and can hardly be used for reliable inversion. This observation defines the practical noise-robustness boundary of our method.

$$Signal-to-noise\ ratio = 10 * log_{10} \frac{\left(u_{label}\right)_{rms}}{\left(u - u_{label}\right)_{rms}} \tag{7}$$

where u represents the wavefield to be calculated, u_{label} represents the target wavefield without noise (composed of Ricker wavelets), and rms stands for root mean square.

4.3. Comparison with a baseline network

To objectively evaluate MSTU's performance, we selected the standard U-Net as the baseline. Both networks were trained under the same training set, test set, and hyperparameters (batch size = 4, epochs = 200, Adam optimizer). The experimental results show that the signal-to-noise ratio after processing by MSTU (13.94 dB) is better than that of U-Net (13.67 dB). Although the inference time per shot of MSTU (3.2 seconds) is slightly higher than that of U-Net (1.7 seconds), both are far lower than the time cost of seismic wavefield forward modeling and fully meet the real-time requirements of the preprocessing stage. More importantly, the U-Net still leaves residual side lobe waveforms after processing and recovers reflected signals poorly at far offsets and deep regions, whereas MSTU performs significantly better in these aspects (Figure 14).

4.4. Data limitation and generalization capability

Constrained by hardware limitations, computational cost, and the difficulty of acquiring field data, this study did not use real data or massive synthetic datasets. Instead, based on the classic Overthrust velocity model, we generated a series of two-dimensional velocity profiles with similar macro-structures but different fine details via equally spaced slicing. This strategy ensures that the training and test sets share similar macro-geological features while maintaining independent velocity details, allowing us to effectively evaluate the model's generalization to unseen velocity structures under a "small-sample" condition. Meanwhile, we employed only noisy synthetic data and two typical non-ideal wavelets to demonstrate training performance. Although the dataset is small, the experimental results show that MSTU can still learn a stable nonlinear mapping, thereby demonstrating the rationality of this data construction strategy. However, open-source field datasets generally lack accurate velocity models and true source wavelets, making it difficult to directly validate our method. Therefore, extensive training on diverse data is required first to achieve strong generalization, and applying the method to real data remains a key direction for future work.

4.5. Applicability boundaries and future extensions

This study mainly discusses two typical "poor" wavelets (Klauder and Ormsby) characterized by an indistinct dominant frequency, strong side lobe energy, and long waveform duration, which lead to low resolution and severe interference in seismic records. For wavelets with extremely narrow bandwidth and a highly concentrated dominant frequency (e.g., the zero-phase Ricker wavelet itself), the proposed method is not necessary. For the envelope method, although it can reduce side lobe

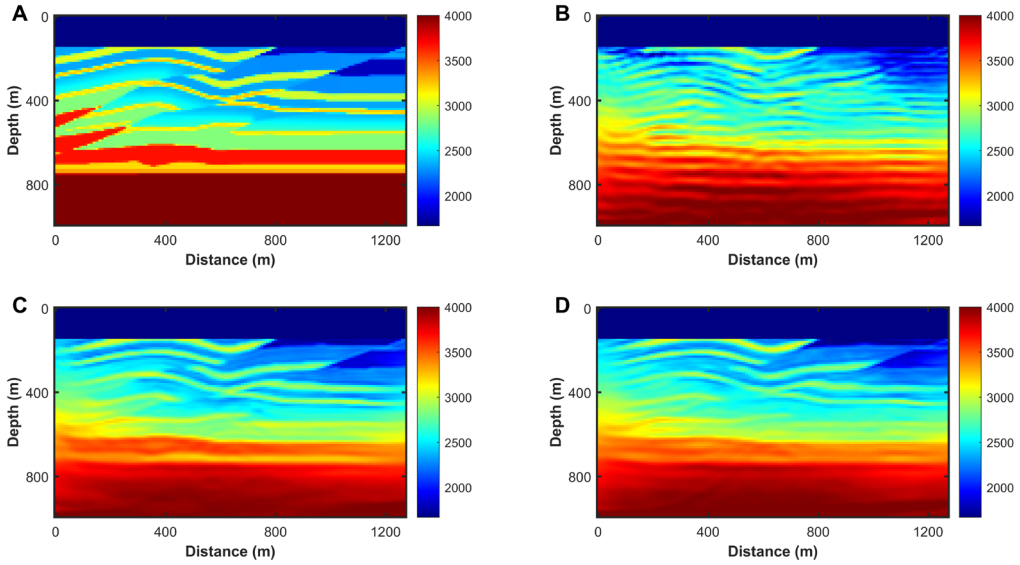


Figure 11. Comparison of inversion results using different strategies on noisy data and a new velocity model. (A) New velocity model used for testing. (B) Inversion result using noisy original Ormsby data with a mismatched Ricker wavelet. (C) Inversion result using ideal noise-free Ricker wavelet data. (D) Inversion result using noisy data universal by the multi-scale Swin Transformer transposed U-Net network with the Ricker wavelet.

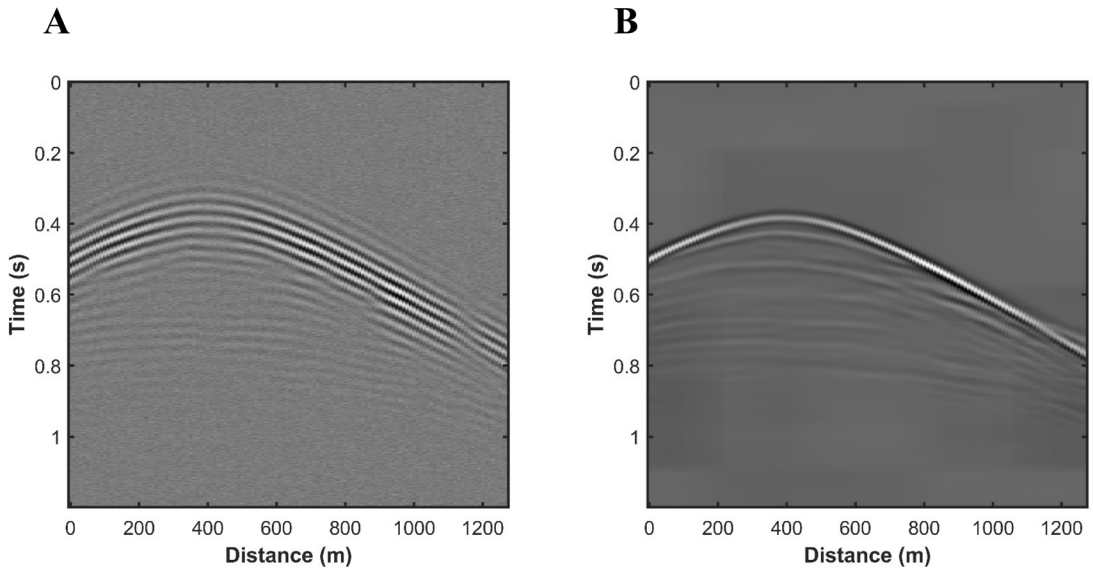


Figure 12. Comparison of standardization processing effects for noisy Ormsby wavelet data. (A) Noisy original single-shot record with Ormsby wavelet. (B) Single-shot record after standardization by the multi-scale Swin Transformer transposed U-Net network.

Table 1. Signal-to-noise ratio in noise immunity testing

Noise percentage	Signal-to-noise ratio of input data	Signal-to-noise ratio of prediction data
0	7.18	13.97
0.05	7.16	13.36
0.1	7.11	12.79
0.5	6.17	11.63
1.0	4.76	11.39

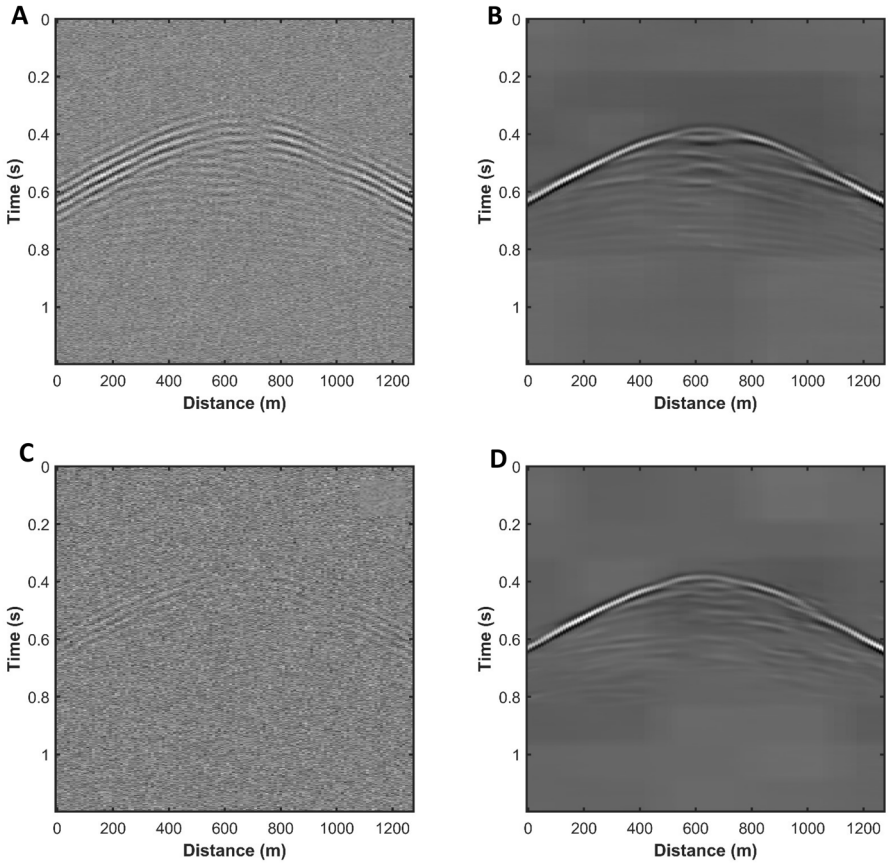


Figure 13. Noise robustness analysis of the multi-scale Swin Transformer transposed U-Net (MSTU) network under different Gaussian noise levels. (A) Input with noise level 0.1 and (B) corresponding standardized output by MSTU (noise level 0.1). (C) Input with noise level 0.5 and (D) corresponding output by MSTU (noise level 0.5).

effects, it also lowers the dominant frequency. In practical exploration, the source wavelet may vary with time, space, or contain multiple mixed types. Therefore, future research could consider: (i) constructing multi-task or adaptive standardization networks that can simultaneously handle multiple wavelet types; (ii) introducing unsupervised or semi-supervised learning frameworks to reduce the dependence on paired label data; (iii) extending the method to three-dimensional seismic data to further exploit the spatial continuity of the wavefield.

4. Conclusion

The accuracy of FWI heavily depends on the precision of the source wavelet. To address waveform distortion and inversion instability caused by non-ideal wavelets (e.g., Klauder and Ormsby wavelets) prevalent in real seismic data, this paper innovatively proposes and implements a deep learning method for seismic wavelets normalization based on the MSTU. The main conclusions of this study are as follows:

(i) Effectiveness of the network architecture: The designed MSTU network efficiently learns the complex nonlinear mapping from non-ideal wavelet seismic records to ideal Ricker wavelet records. Through its hierarchical Swin Transformer encoder, the network precisely models both global and local features in the wavefield. Using a transposed convolutional decoder and U-Net skip connections enables high-fidelity waveform reconstruction by extracting complex wavelet features from the raw data.

(ii) Significant improvement in inversion accuracy through data standardization: Systematic numerical experiments demonstrate that performing FWI directly on raw non-ideal wavelet data (assumed to be Ricker waves) leads to severe structural distortion and frequency jumps. In contrast, data universal by this method exhibits significantly superior structural fidelity and detail resolution during inversion using standard Ricker wavelet procedures compared to mismatched wavelet scenarios. It even outperforms

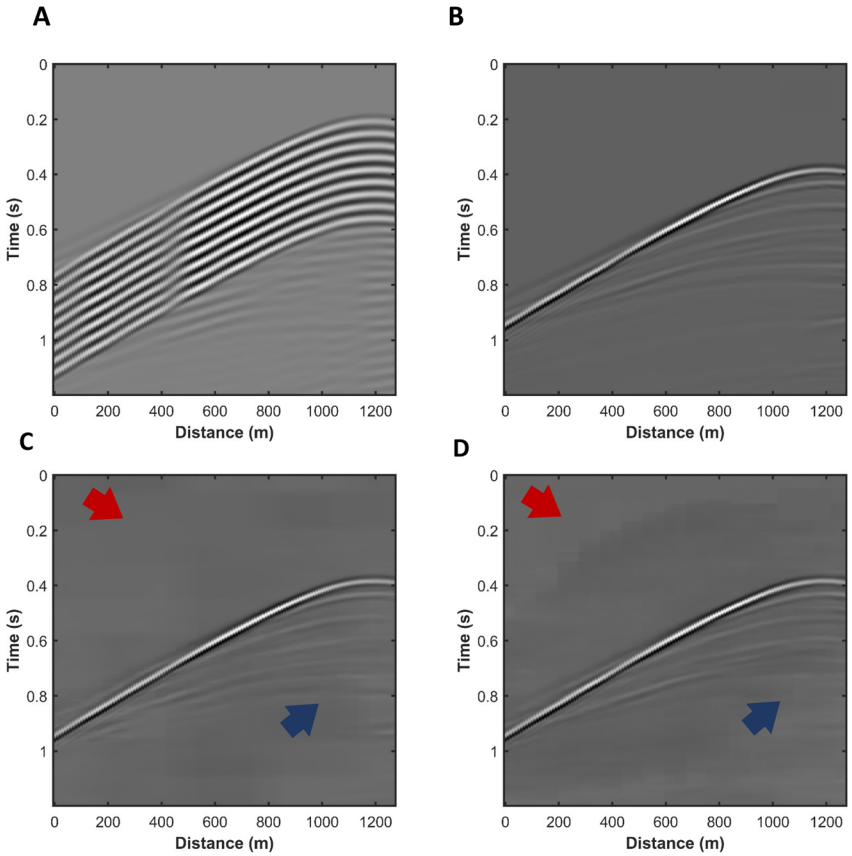


Figure 14. Comparison of wavelet standardization results on a representative test sample. (A) Input seismic record (with non-ideal wavelet). (B) Target label (ideal Ricker wavelet record). (C) Processed result by the proposed multi-scale Swin Transformer transposed U-Net (MSTU) network. (D) Processed result by the baseline U-Net network. The red arrow highlights that the baseline U-Net leaves noticeable residual side lobe energy, whereas MSTU produces cleaner waveforms. The blue arrow highlights that the baseline U-Net poorly recovers deep reflection events (especially at far offsets), whereas MSTU achieves a more complete reconstruction. These comparisons demonstrate the superiority of MSTU in wavelet standardization.

inversion using known yet complex, non-ideal original wavelets, approaching the theoretical optimal standard achieved with true ideal Ricker wavelet data.

(iii) Practicality and generalizability: This method offers a preprocessing solution that does not rely on precise prior knowledge of source waveform morphology nor require modification of existing mature FWI core algorithms. Test results on a limited dataset constructed based on the classical Overthrust model confirm the method's strong generalizability to unknown models with similar geological backgrounds.

In summary, this study innovatively integrated cutting-edge deep learning architectures into the upstream phase of seismic data quality control, offering a novel and effective technical pathway to address the long-standing challenge of wavelet uncertainty that has constrained the practical application of FWI. By enabling intelligent wavefield reconstruction, this method opens new possibilities for high-precision velocity modeling and detailed reservoir characterization using vast existing seismic archives, demonstrating clear theoretical value and broad application prospects. Future research should focus on enhancing the model's robustness and adaptability to more complex geological conditions, lower signal-to-noise ratio data, and three-dimensional scenarios, thereby advancing this methodology from theoretical validation to large-scale production applications.

Acknowledgments

None.

Funding

This research was supported by the National Natural Science Foundation of China (no.42474081; 42130805; 42374147), the Jilin Provincial Natural Science

Foundation Outstanding Young Scientist Fund Project (no. 20250101038JJ), and the Special Fund of the Institute of Geophysics, China Earthquake Administration (no. DQJB26B29).

Conflict of interest

The authors declare they have no competing interests.

Author contributions

Conceptualization: Binghui Zhao
Formal analysis: Pan Zhang
Investigation: Liguo Han
Methodology: Binghui Zhao, Pan Zhang, Xujia Shang
Writing–original draft: Binghui Zhao, Laiyu Lu
Writing–review & editing: Laiyu Lu, Tongwei Qin

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

All data analyzed have been presented in the paper.

References

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

   doi: 10.1190/1.1441754

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

   doi: 10.1190/1.3238367

3. Yang QY, Hu GH, Wang LX. Quan bo xing fan yan yan jiu xian zhuang ji fa zhan qu shi [Research status and development trend of full waveform inversion]. *Geophys Prospect Pet*. 2014;53(01):77-83. [In Chinese]

   doi: 10.3969/j.issn.1000-1441.2014.01.011

4. Wang XQ, Tan HD, Xu ZL. CPU/GPU xie tong ji suan zai pin lv yu er wei quan bo xing fan yan zhong de ying yong [The Application of CPU/GPU Collaborative Computing in the Frequency Domain 2D Full Waveform Inversion]. *CT Theory Appl*. 2016;25(01):23-32. [In Chinese]

   doi: 10.15953/j.1004-4140.2016.25.01.03

5. Pratt RG. Seismic waveform inversion in the frequency domain, Part 1: Theory and verification in a physical scale model. *Geophysics*. 1999;64(3):888-901.

   doi: 10.1190/1.1444597

6. Ryan H. Ricker, Ormsby, Klander, Butterworth – A Choice of Wavelets. *CSEG Recorder*. 1994;19(7):8-20.

7. Zhang H, Wang X, Wang W, He B, Yang T. Multi-scale elastic full-waveform inversion in a hybrid domain based on deconvolution objective function. *Geophys Prospect*. 2024;72(4): 1222-1235.

   doi: 10.1111/1365-2478.13332

8. Ao RD, Dong LG, Chi BX. Bu yi lai zi bo ji yu bao luo de FWI chu shi mo xing jian li fang fa yan jiu [Source-independent envelope-based FWI to build an initial model]. *Chin J Geophys*. 2015;58(6):1998-2010. [In Chinese]

   doi: 10.6038/cjg20150615

9. Song ZM, Williamson PR. Frequency-domain acoustic-wave modeling and inversion of crosshole data; Part 1, 2.5-D modeling method. *Geophysics*. 1995;60(3):784-795.

   doi: 10.1190/1.1443817

10. Hu Y, Han LG, Xu Z, *et al.* Ji yu jing que zhen yuan han shu de jie tiao bao luo duo chi du quan bo xing fan yan [Demodulation envelope multi-scale full waveform inversion based on precise seismic source function]. *Chin J Geophys*. 2017;60(3):1088-1105. [In Chinese]

    doi: 10.6038/cjg20170321

11. Liu LB, Li ZC, Zhang YY, *et al.* Yi zhong shi jian yu quan bo xing fan yan zhong de di zhen zi bo gu ji fang fa [Seismic wavelet estimation for full-waveform inversion in time domain]. *Prog Geophys*. 2020;35(3):987-994. [In Chinese]

    doi: 10.6038/pg2020DD0138

12. Han HY, Dai YS, Song JG, *et al.* Ji yu sheng cheng dui kang wang luo de ta li mu shen ceng chao shen ceng die qian di zhen zi bo ti qu [Deep prestack seismic wavelets extraction in Tarim based on generative adversarial network]. *Chin J Geophys*. 2022;65(2):763-772. [In Chinese]

    doi: 10.6038/cjg2022P0130

13. Choi Y, Alkhalifah T. Source-independent time-domain waveform inversion using convolved wavefields: Application to the encoded multisource waveform inversion. *Geophysics*. 2011;76(5):R125-R134.

    doi: 10.1190/geo2010-0210.1

14. Wu RS, Luo J, Wu B. Seismic envelope inversion and modulation signal model. *Geophysics*. 2014;79(3):WA13-WA24.

    doi: 10.1190/geo2013-0294.1

15. Chi B, Dong L, Liu Y. Full waveform inversion method using envelope objective function without low frequency data. *J Appl Geophys*. 2014;109:36-46.

    doi: 10.1016/j.jappgeo.2014.07.010

16. Bai T, Tsvankin I. Source-independent waveform inversion for attenuation estimation in anisotropic media. *Geophys. Prospect*. 2019;67(9):2343-2357.

doi: 10.1111/1365-2478.12821

17. Yang T, Zhang HX, Shi CW. Bu yi lai zi bo de tan xing bo hun he yu quan bo xing fan yan [Wavelet-independent elastic wave full waveform inversion in hybrid domain]. *Oil Geophys. Prospect*. 2019;54(2):348-355. [In Chinese]

doi: 10.13810/j.cnki.issn.1000-7210.2019.02.013

18. Liu Y, Tsvankin I. Source-independent time-lapse full-waveform inversion for anisotropic media. *Geophysics*. 2022;87(1):R111-R122.

doi: 10.1190/geo2021-0306.1

19. Shi XY, Yang HC, Zhang JZ. Hai yang tuo lan yu hai di jie dian zi liao lian he quan bo xing fan yan [Full waveform inversion for combined towed-streamer and ocean buttom node seismic data]. *Oil Geophys Prospect*. 2024;59(5):1058-1068. [In Chinese]

doi: 10.13810/j.cnki.issn.1000-7210.2024.05.013

20. Zhang B, Wu GC, Jia ZF, *et al.* Bu yi lai zi bo de ji fen bo chang sheng tan ou he bo xing fan yan fang fa [Wavelet-independent integral wavefield waveform inversion method based on acoustic-elastic coupled equations]. *Oil Geophys Prospect*. 2025;60(6):1492-1503. [In Chinese]

doi: 10.13810/j.cnki.issn.1000-7210.20250053

21. Chen G, Wu R, Chen S. Reflection multi-scale envelope inversion. *Geophysical Prospecting*. 2018;66(7):1258-1271.

doi: 10.1111/1365-2478.12624

22. Chen GX, Wu RS, Wang YQ, Chen SC. Multi-scale signed envelope inversion. *J Appl Geophys.* 2018;153:113-126.

doi: 10.1016/j.jappgeo.2018.04.008

23. Chen G, Yang W, Liu Y, Wang H, Huang X. Salt Structure Elastic Full Waveform Inversion Based on the Multiscale Signed Envelope. *IEEE Trans Geosci Remote Sens.* 2022;60:1-12.

doi: 10.1109/tgrs.2022.3166028

24. Xing ZZ, Han LG, Hu Y, *et al.* Ji yu gui yi hua neng liang pu mu biao han shu de quan bo xing fan yan fang fa [Full waveform inversion based on normalized energy spectrum objective function]. *Chin J Geophys.* 2019;62(7):2645-2659. [In Chinese]

doi: 10.6038/cjg2019M0429

25. Vaswani A, Shazeer N, Parmar N, *et al.* Attention Is All You Need. *arXiv*. 2017.

doi: 10.48550/arXiv.1706.03762

26. Dosovitskiy A, Beyer L, Kolesnikov A, *et al.* An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv*. 2020.

doi: 10.48550/arXiv.2010.11929

27. Liu Z, Lin Y, Cao Y, *et al.* Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. *arXiv*. 2021.

doi: 10.48550/arXiv.2103.14030

28. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science.* International Publishing Springer; 2015;9351:234-241.

doi: 10.1007/978-3-319-24574-4_28

29. Chen J, Chen G, Li J, *et al.* Efficient Seismic Data Denoising via Deep Learning With Improved MCA-SCUNet. *IEEE Trans Geosci Remote Sens.* 2024;62:1-14.

doi: 10.1109/tgrs.2024.3355972

30. Qi Y, Chen G, Chen J, *et al.* Seismic Denoising via Multiround SCU-Net. *IEEE Geosci Remote Sens Lett*. 2026;23:1-5.

doi: 10.1109/lgrs.2025.3640968

31. Chen G, Chen J, Jensen K, *et al.* Joint data and model-driven simultaneous inversion of velocity and density. *Geophys J Int.* 2024;237(3):1674-1698.

doi: 10.1093/gji/ggae128

32. Bunks C, Saleck FM, Zaleski S, *et al.* Multiscale seismic waveform inversion. *Geophysics*. 1995;60(5):1457-1473.

doi: 10.1190/1.1443880

33. Pratt RG, Shin C, Hick GJ. Gauss–Newton and full Newton methods in frequency–space seismic waveform inversion. *Geophys J Int*. 1998;133(2):341-362.

doi: 10.1046/j.1365-246X.1998.00498.x