

Original Paper

Dual-branch wavelet-fusion U-Net for vibroseis harmonic noise attenuation

Yang Liu^{a,b}, Xiao-Tao Wen^{a,b,*}, Qing-Jun Yang^c, Huan Zhu^{a,b}, Yun Zhao^{a,b},
Theo Le Gallais^d

^a Key Laboratory of Earth Exploration and Information Techniques, Ministry of Education, Chengdu University of Technology, Chengdu, 610059, Sichuan, China

^b College of Geophysics, Chengdu University of Technology, Chengdu, 610059, Sichuan, China

^c College of Minerals, Energy and Chemical Engineering, Curtin University, Perth, Western Australia, 6102, Australia

^d College of Earth and Planetary Sciences, Curtin University, Perth, Western Australia, 6102, Australia



ARTICLE INFO

Article history:

Received 26 November 2025

Received in revised form

21 March 2026

Accepted 29 April 2026

Available online 8 May 2026

Edited by Jie Hao

Keywords:

Deep learning

Vibroseis

Harmonic noise attenuation

DW-U-Net

Time-frequency analysis

ABSTRACT

Vibroseis is widely used in land surveys, but source-ground coupling and equipment nonlinearities generate second- and higher-order harmonics as well as narrowband resonance lines. These source-related interferences overlap the target reflection band, markedly degrading the signal-to-noise ratio (SNR) and masking weak reflections. We propose a dual-branch wavelet-fusion U-Net (DW-U-Net) for harmonic noise attenuation. The backbone operates in the time domain, while a wavelet side branch explicitly represents low-frequency background and high-frequency details. A differentiable inverse discrete wavelet transform (IDWT) injects the reconstructed wavelet features into the decoder to structurally suppress harmonic streaks while preserving reflections. Training data are generated by adding higher-order harmonics to a reference sweep and then simulating shot records through forward modeling. On the Marmousi2 model under strong harmonic noise, DW-U-Net increases the whole-shot average SNR from 9.92 dB to 29.29 dB, and achieves residual energy 2.44×10^{-3} , root mean square error (RMSE) 4.51×10^{-5} , and multiscale structural similarity (MS-SSIM) 0.9959. Experiments on the Marmousi2 model and the public Utah FORGE field data demonstrate that DW-U-Net attenuates coherent harmonic streaks and suppresses narrowband resonance lines in shot records. Compared with DnCNN, a standard U-Net, dual residual network (DURN), and multi-input multi-output U-Net (MIMO-U-Net) trained under the same protocol, DW-U-Net delivers consistently higher SNR under strong noise levels. Time-frequency analyses further show that the main reflection band is well preserved in amplitude and bandwidth, while coherent harmonic remnants are substantially reduced.

© 2026 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

High-quality seismic data directly influences the reliability of hydrocarbon targeting and subsurface interpretation. In practice, interference from the source and the near surface reduces the SNR, masks weak reflections, and destabilizes the effective wavelet. Vibroseis is widely used in land acquisition due to its flexible deployment and controllable output. However, it is susceptible to nonlinearities in the source mechanism and coupling with the ground. These effects distort an ideal linear sweep and produce higher-order harmonics and narrowband resonance lines. After

correlation with the pilot, these distortions appear as sidelobes flanking the main lobe. These artifacts degrade velocity model building, impair elastic parameter inversion, and hinder attribute analysis. As a result, imaging and reservoir characterization are compromised (Vedanti et al., 2021). Both simulations and field data show that source-related harmonics form banded and comb-shaped coherent patterns that substantially overlap the main reflection band. This strong overlap creates a trade-off for directional filters and conventional deconvolution between harmonic attenuation and primary signal preservation (Abd El-Aal, 2010). Transform-domain modeling, when combined with learning-based inversion, has recently shown strong potential for suppressing structured interference. This progress motivates the development of methods that specifically target harmonic structures while preserving reflections (Si et al., 2023). At low SNR, simple thresholding or linear filtering often fails by either

* Corresponding author.

E-mail address: wenxiaotao@cdut.cn (X.-T. Wen).

Peer review under the responsibility of China University of Petroleum (Beijing).

insufficiently reducing noise or excessively attenuating it. This motivates the use of more robust, adaptive, and multistage decompositions (Kong et al., 2017; Yuan et al., 2022). Furthermore, harmonic and resonance components can pass through multiple processing stages. In some cases, they are even amplified, leading to cumulative degradation of data quality (Walden and White, 1998).

On the acquisition and processing fronts, several strategies have been explored. For simultaneous source and high productivity designs, such as slip sweep, anticorrelation or cancellation across prestack records can suppress crosstalk streaks while preserving reflections (Zhang et al., 2012). This line of work addresses crosstalk, which is related to but distinct from the source-induced harmonics considered here. In the correlation domain, stepwise harmonic removal using a pipeline of synchrosqueezing transform, filtering, and inverse synchrosqueezing transform remains compatible with truncated correlation records while maintaining operational efficiency (Qu et al., 2015). From the source control perspective, optimizing servo parameters and regulating ground force reduces harmonic distortion and enhances the usability of low-frequency output (Wei and Phillips, 2013). Mixed sweep designs, including those based on orthogonal frequency-division multiplexing, jointly control amplitude and phase and redistribute energy across the band to balance penetration against interference suppression, thereby improving the field SNR (Gao et al., 2025). For encoded imaging, source-independent sinusoidal encoding uses trigonometric orthogonality to facilitate multiscale inversion and debrending while reducing sensitivity to the unknown source signature (Zhang et al., 2020).

Harmonic noise suppression in seismic data has progressed from explicit modeling through adaptive representations to the enforcement of phase consistency. Early studies operated in the correlation domain and emphasized explicit identification followed by progressive subtraction of harmonic components. One representative approach identifies harmonics in increasing order and removes them sequentially in the correlation domain, thereby directly weakening striped interference (Abd El-Aal, 2011). Building on this idea, the double phase shift technique applies a 90° phase rotation to spectral peaks and constructs linear combinations that cancel harmonics from the first through the fourth order by exploiting the spectral peak characteristics of correlated sweeps, thereby improving the efficiency of high-order suppression (Li et al., 2022). Subsequently, researchers introduced parametric models with multiple sinusoids for harmonic cancellation. Iterative estimation and subtraction produced a general framework for harmonic modeling that maps all harmonic orders into an estimable parameter space and improves transferability across acquisition conditions (Butler and Russell, 2003). In parallel, multiple-filter decomposition uses a bank of narrowband filters to separate the fundamental sweep component and its harmonics in the joint time-frequency domain. It also enforces amplitude and phase templates, thereby strengthening the connection between interpretation and processing (Li, 1997).

As non-stationarity in field records and strong source-related interference have become more pronounced, the research focus has shifted toward more robust representations, including adaptive thresholding, and low-rank and sparse modeling. Adaptive time-frequency thresholding constructs a dynamic threshold curve via exponential fitting and applies soft thresholding in the frequency domain to track the slow drift of harmonics and useful reflections in both time and frequency, thereby reducing the risks of over-suppression and under-suppression inherent to fixed thresholds (Zhang et al., 2024). Multiband low rank methods project the data into a tensor space formed by multiple bandpass components and, when combined with sparse regularization,

jointly attenuate random noise and anomalous coherent events, providing an amplitude-preserving and robust background-cleaning mechanism for harmonic mitigation (Lin and Zhang, 2022). Centroid-oriented transforms are tailored to the geometry of harmonic components and delineate the principal direction of energy dispersion in the time-frequency domain. This enhances the separability between harmonics and transient reflections at the representation level and enables nearly lossless reconstruction (Chen et al., 2024). Multiscale phase-only correlation prioritizes phase consistency. By reinforcing the phase coherence of weak reflections in a generalized time-frequency transform domain, it maintains stable performance in the presence of residual harmonics and strong background noise (Fang and Zhang, 2025). Meanwhile, to support engineering deployment and parameter control, studies on dictionary-learning parameters systematically quantify how patch size, dictionary cardinality, and sparsity level affect SNR improvement and computational cost. The results provide actionable tuning guidelines for diverse targets and acquisition settings (Kuruguntla et al., 2022).

In recent years, deep learning has substantially advanced seismic denoising, offering improved fidelity and adaptability over many classical filtering pipelines. Time-frequency masking jointly estimates signal and noise masks to better preserve amplitudes when reflections and interference spectrally overlap (Zhu et al., 2019). Wavelet-based U-Net variants replace standard pooling with wavelet operators and exploit subband sparsity to maintain stable performance under both regular and irregular sampling conditions (Dodda et al., 2023). Cross-scale feature interaction aggregates multi-resolution structural cues and improves weak-reflection preservation in convolutional pipelines (Liu et al., 2020). Dilated receptive-field expansion further broadens contextual perception without sacrificing resolution, thereby enhancing robustness to spatially extended interference (Hu et al., 2025). Nevertheless, purely convolutional denoisers are often limited in their ability to model long-range dependencies in complex seismic data. To address this issue, Transformer-based architectures have been introduced to improve global contextual modeling. Shifted-window Transformer designs capture cross-window interactions and have demonstrated improved random-noise attenuation while preserving structural details (Li et al., 2024). Incorporating Swin-Transformer blocks into U-Net backbones further strengthens multiscale contextual mixing for joint denoising and resolution enhancement (Sun et al., 2025). Global-local fusion frameworks combine Transformer-derived global context with DenseNet-style local detail modeling to address more challenging reconstruction tasks, such as internal multiple suppression (Zhang et al., 2025). Joint super-resolution and denoising architectures further integrate Swin-Transformer-enhanced residual blocks with attention-guided feature weighting to improve stratigraphic continuity and fine-detail recovery under noisy conditions (Wu et al., 2026). Beyond these task-specific network designs, recent seismic denoising studies have also begun to explore large pre-trained models and foundation-model-style frameworks. Transformer-based pretraining schemes, such as StorSeismic, suggest that shared representations learned from large seismic datasets can be transferred to downstream processing tasks, including denoising (Harsuko and Alkhalifah, 2022). More recently, diffusion-based generative models and other large-capacity architectures have also shown promise for seismic denoising, indicating that large-model priors may improve robustness and generalization under complex noise conditions (Durall et al., 2023). In parallel, seismic foundation models trained with large-scale self-supervised learning have further demonstrated the potential of transferable feature extraction for denoising, interpolation, and inversion (Sheng et al., 2025). These

methods are generally developed for broader seismic processing requirements and may require substantial computational resources. For source-related harmonic attenuation in vibroseis data, efficient task-oriented architectures therefore remain of considerable practical importance.

At the same time, network architecture alone does not fully resolve the problem when interference is structured and coherent, as in vibroseis source-related harmonics. In such cases, the learning paradigm itself becomes a key constraint. Self-supervised learning exploits intrinsic data redundancy to enable label-free processing in tasks such as seismic interpolation without ground truth (Birnie et al., 2021). Masking-based training strategies provide surrogate supervision from partial observations when paired clean and noisy data are unavailable (Liu et al., 2022). Many widely used self-supervised objectives implicitly rely on pixel-wise independence assumptions, as exemplified by Noise2Void-style masked prediction (Krull et al., 2019). Coherent and correlated interference, however, violates the noise assumptions underlying blind-spot self-supervision and thus increases the risk of signal leakage (Birnie and Alkhalifah, 2022). When coherent ridges cannot be reliably disentangled from reflection energy, ambiguous noise modeling may also leave residual striping artifacts or produce over-smoothed textures (Liu et al., 2023). Therefore, when paired data are available, fully supervised learning remains a practical and reliable framework for precise harmonic suppression while preserving amplitude and phase fidelity.

Despite recent progress, it remains challenging to jointly address the computational bottleneck of global modeling and the distinctive long-range structure of coherent harmonic noise: CNNs may not provide sufficient long-range context to follow sweep-aligned harmonic trajectories, whereas full Transformers can be prohibitively expensive in practice. To meet the need for an efficient yet globally aware and strictly supervised solution, we propose a Dual-Branch Wavelet Fusion U-Net (DW-U-Net). DW-U-Net integrates multiscale wavelet subband representations with a lightweight context-mixing module (ContMix1D) that captures long-range dependencies without full self-attention. The wavelet branch facilitates signal-noise separation across subbands, and ContMix1D enhances trajectory-consistent suppression of large-scale harmonic artifacts while preserving weak reflections within a robust fully supervised framework.

We evaluate it on the Marmousi2 model and field data from the Utah FORGE (Miller, 2018). Because harmonic noise in single traces is both prominent and quasi-periodic, we use full-length 1D single-trace records as training instances. Avoiding patch extraction preserves absolute temporal localization and eliminates patch boundary artifacts, thereby enabling stronger harmonic suppression while maintaining spectral fidelity and temporal continuity. We summarize our main contributions as follows:

- 1) We propose DW-U-Net: a U-Net backbone augmented with a wavelet side branch. Multiscale features are injected into the decoder via a differentiable IDWT, enabling directional suppression of higher-order harmonics and narrowband resonance lines, while preserving amplitude, bandwidth, and temporal continuity.
- 2) We design joint objectives in the time and wavelet domains. Time-domain supervision preserves amplitude and phase fidelity, while wavelet-domain supervision guides the learning of approximation and detail coefficients. In addition, a synthesis-domain consistency term is introduced to improve cooperative stability between the wavelet branch and the backbone.
- 3) We train the network on full-length single-trace records to preserve absolute timing and avoid block artifacts. On the Marmousi2 model and Utah FORGE field data, DW-U-Net

consistently suppresses second through fourth harmonics and narrowband resonances, outperforming other networks in terms of SNR and spectral fidelity.

The remainder of the paper is organized as follows. Section 2 reviews vibroseis sweeps and harmonic-generation mechanisms, summarizes how harmonics manifest in the correlation domain, and formulates the problem and technical challenges addressed in this work. It then introduces the proposed DW-U-Net, including the dual-branch architecture, the differentiable IDWT-based multiscale fusion strategy, the ContMix1D module for long-range context modeling, and the training objectives and implementation details. Section 3 describes the experimental setup and evaluation protocol, including Marmousi2-based synthetic data construction, train-validation splits, comparison baselines, and quantitative/qualitative metrics, followed by results on both synthetic and Utah FORGE field data. Section 4 provides an in-depth discussion covering computational cost and efficiency analysis, module- and loss-level ablations, wavelet-basis sensitivity, and generalization under different validation-set ratios and additional unseen field shots. Section 5 concludes the paper by summarizing the main findings, limitations, and directions for future work.

2. Methodology

2.1. Vibroseis signal

In vibroseis exploration, the linear sweep signal is commonly used: the excitation bandwidth is prescribed so that the frequency sweeps from f_1 to f_2 within a specified duration T (Aldridge, 1992). The signal can be written as

$$S(t) = A(t) \sin \left(2\pi \left(f_1 + \frac{qt}{2} \right) t + \phi_0 \right) \quad 0 \leq t \leq T \quad (1)$$

where $A(t)$ is an amplitude taper, ϕ_0 is the initial phase, f_1 is the start frequency of the sweep signal, f_2 is the cutoff frequency of the sweep signal, and $q = (f_2 - f_1)/T$ is the sweep rate.

We apply a smooth amplitude envelope to the linear sweep. The envelope ramps up at the start, stays at unity in the middle, and ramps down at the end. This suppresses start and stop transients and reduces spectral leakage after pilot correlation. A convenient choice is a raised-cosine taper with edge fraction $\alpha \in [0, 1]$. Let $\tau = \alpha T/2$. Then

$$A(t) = \begin{cases} \frac{1}{2} \left[1 + \cos \pi \left(\frac{t}{\tau} + 1 \right) \right], & 0 \leq t < \tau \\ 1, & \tau \leq t \leq T - \tau \\ \frac{1}{2} \left[1 + \cos \pi \left(1 - \frac{T-t}{\tau} \right) \right], & T - \tau < t \leq T \end{cases} \quad (2)$$

with $0 \leq A(t) \leq 1$, $A(0) = A(T) = 0$, and continuous first derivatives at the junctions. Special cases include a rectangular envelope $\alpha = 0$ and a full Hann envelope ($\alpha = 1$).

2.2. Harmonic signal

Two classes of harmonic noise are considered. The first arises from distortions in the vibrator mechanism: nonlinearities replicate the instantaneous frequency of the linear sweep signal at integer multiples, so the ground-force measurement contains harmonic components at orders $m = 2, 3, \dots$. A convenient representation is

$$h_1(t) = \sum_{m=2}^M a_m \sin\left(2\pi m\left(f_1 + \frac{qt}{2}\right)t + \phi_m\right) \quad (3)$$

where a_m is the amplitude of the m -th harmonic. M is the maximum harmonic order, and ϕ_m is the phase offset of the m -th order.

The second class is induced by nonlinearities in the near-surface response. When the instantaneous frequency of the sweep meets the site's natural frequency f_0 , a resonance is excited and a short, narrowband ringing is superimposed (Chopra, 2020). Let t_r be the time at which $f_i(t_r) = f_0$. A standard damped-sinusoid model is

$$h_2(t) = C e^{-\zeta\omega_0(t-t_r)} \sin(2\pi f_0(t-t_r) + \phi_0) u(t-t_r) \quad (4)$$

where $\omega_0 = 2\pi f_0$ and $u(\cdot)$ is the Heaviside step function. The resonance-onset time is

$$t_r = \frac{f_0 - f_1}{q}. \quad (5)$$

Here, C is the initial amplitude constant, ζ is the dimensionless damping ratio, f_0 is the near-surface natural frequency, and ϕ_0 is the initial phase. (4) captures the exponentially decaying, narrowband ringing centered at f_0 . The decay time constant is $1/(\zeta\omega_0)$.

Autocorrelation of the linear sweep signal and the two harmonic components yield distinct wavelet signatures (Fig. 1). Fig. 1(a) shows the Klauder wavelet obtained from the linear sweep signal. Fig. 1(b) illustrates the wavelet associated with the first class of harmonic interference, whose energy is concentrated at negative time lags. Under shot-by-shot acquisition these components can be largely removed because they fall outside the time window of the current shot, whereas under slip-sweep acquisition they contaminate the record of the preceding shot. Fig. 1(c) shows the wavelet for the second class of harmonic interference, which appears at both positive and negative time lags, with greater amplitude and stronger impact on the positive side.

2.3. Network structure

Fig. 2 illustrates the overall architecture of the proposed DW-U-Net. This 1D dual-branch network comprises four submodules: the

time-domain backbone (T-U-Net), the wavelet-domain feature branch (W-U-Net), the multiscale differentiable inverse wavelet fusion (MS-IF), and the reconstruction module (RM).

Specifically, T-U-Net adopts a four-level encoder and decoder, with feature channels increasing from 32 to 512 ($32 \rightarrow 64 \rightarrow 128 \rightarrow 256 \rightarrow 512$), followed by symmetric upsampling on the decoder path. Each scale contains two blocks of 1D convolution, batch normalization, and LeakyReLU. Skip connections link each decoder level to its corresponding encoder level.

W-U-Net takes the first-level discrete wavelet transform coefficients ($\mathbf{c}_{a1}, \mathbf{c}_{d1}$) as input, progressively generating multiscale coefficients ($\mathbf{c}_a, \mathbf{c}_d$) within the branch. A 1×1 pointwise mapping is used for energy and channel alignment.

Context-Mixing Dynamic Convolution integrates global context into each convolution (Lou and Yu, 2025). It uses a top-down prior to compute location-wise affinities to region centers and turns them into spatially varying kernels, which lets a pure CNN capture long-range dependencies while maintaining a strong local inductive bias. To expand the effective receptive field with minimal computational overhead while preserving detail, we adapt it to the present denoising task into ContMix1D. Dual-neighborhood dynamic kernels apply content-adaptive weighting to local segments aligned in temporal position. Learnable Explicit Positional Encoding (LEPE), Squeeze-and-Excitation (SE) channel reweighting, and a gated residual path are combined to implement a local attention operator with linear complexity. ContMix1D blocks are inserted at multiple decoder levels. The structure of this module is shown in Fig. 3.

At each decoding scale, the MS-IF module concatenates the upsampled features, the corresponding skip-connected features, and the multiscale time-domain reconstruction obtained via a differentiable IDWT. The concatenated tensor is then fused by a 1D convolution. This design couples narrowband characterization in the wavelet domain with structural representation in the time domain, improving harmonic noise suppression.

The reconstruction module (RM) adopts a direct regression strategy. The backbone T-U-Net first produces an intermediate estimate, which is refined using multiscale differentiable inverse wavelet fusion and a decoder fusion head to directly regress the final clean signal. The wavelet branch includes auxiliary supervision heads that predict wavelet coefficients. During training, we jointly optimize the main reconstruction loss and auxiliary coefficient losses to preserve amplitude fidelity and improve

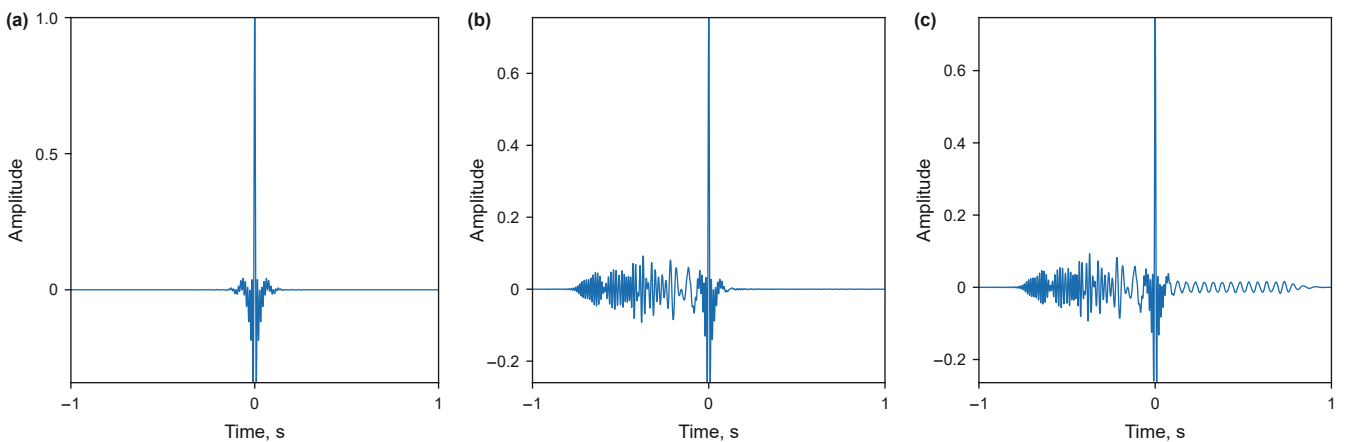


Fig. 1. Three types of waveforms. (a) Klauder wavelet, (b) wavelet of the first class of harmonic interference, (c) wavelet of the second class of harmonic interference.

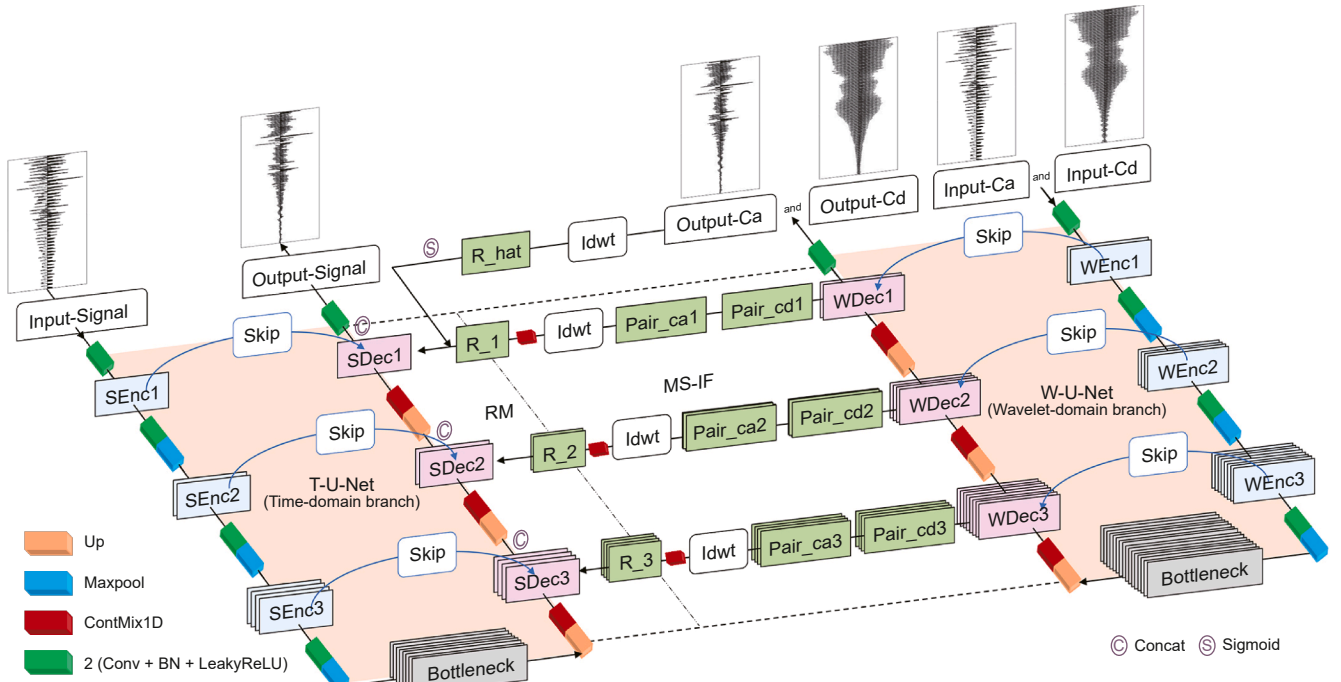


Fig. 2. DW-U-Net structure used for seismic data denoising.

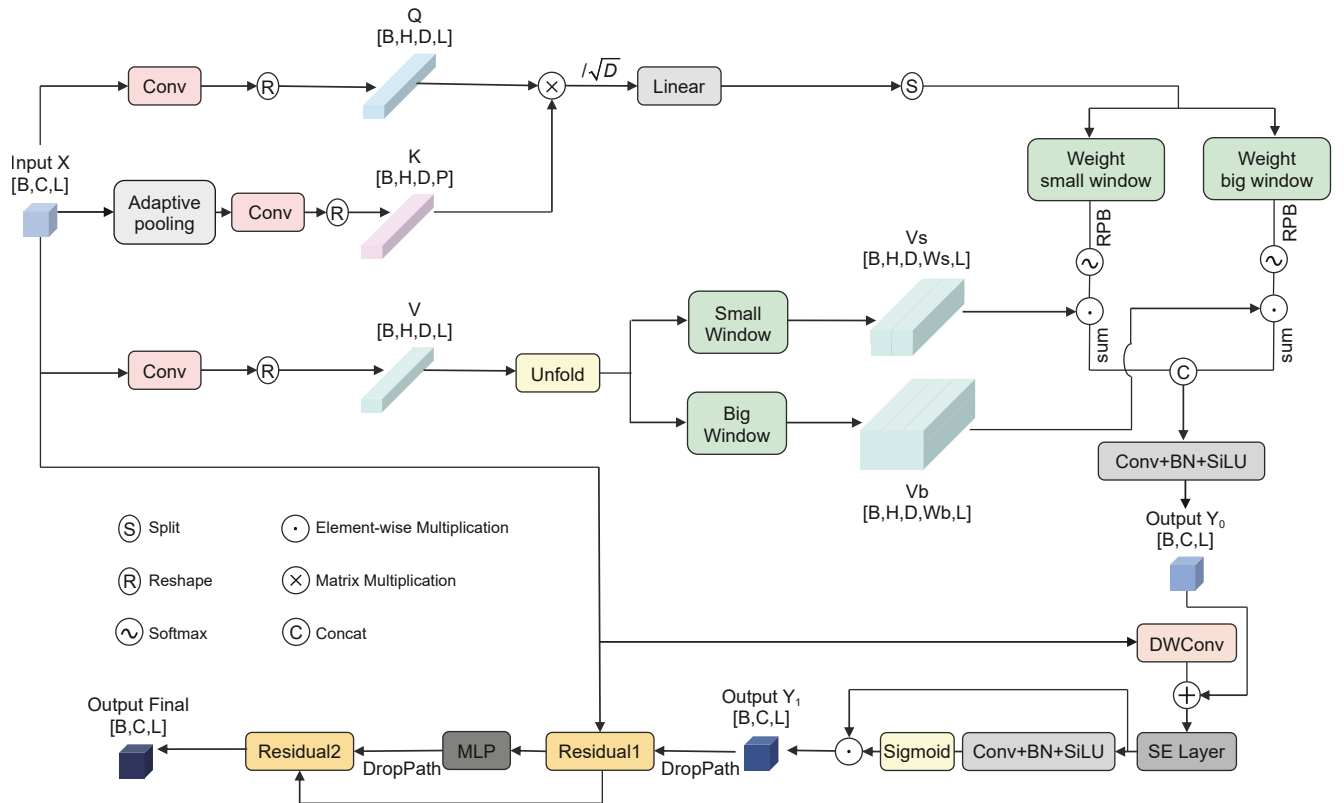


Fig. 3. A schematic diagram of dynamic convolution (ContMix1D).

resolution and SNR. For efficient deployment, ContMix1D supports structural reparameterization, enabling layer merging at inference to accelerate throughput without degrading accuracy.

2.3.1. T-U-Net: time-domain branch

The recorded seismic trace $\mathbf{y}$ is written as

$$\mathbf{y} = \mathbf{x} + \mathbf{h} \quad (6)$$

where $\mathbf{x}$ denotes the desired reflection signal and other components. $\mathbf{h}$ denotes harmonic noise. The time-domain backbone T-U-Net takes $\mathbf{y}$ as input and performs encoding-decoding to extract layered structures and in-phase continuity. Its nonlinear mapping is

$$\hat{\mathbf{y}} = \lambda_t(\mathbf{y}) \quad (7)$$

where $\hat{\mathbf{y}}$ is the backbone prediction and $\lambda_t(\cdot)$ represents the backbone's nonlinear transform. The objective of T-U-Net is to recover the useful signal with correct structure and amplitude and to avoid misclassifying reflections as noise.

2.3.2. W-U-Net: wavelet-domain branch

In the wavelet branch W-U-Net, a one-level discrete wavelet transform (DWT) is applied to $\mathbf{y}$, and the resulting approximation and detail coefficients are fed to the network

$$\begin{aligned} \mathbf{c}_a[n] &= \sum_k h[k] y[2n - k] \\ \mathbf{c}_d[n] &= \sum_k g[k] y[2n - k] \end{aligned} \quad (8)$$

where $\mathbf{c}_a$ and $\mathbf{c}_d$ are the low-frequency approximation and high-frequency detail coefficients. $h[\cdot]$ and $g[\cdot]$ are the fixed low-pass and high-pass analysis filters. Near-surface resonance often concentrates energy in a narrow, nearly constant-frequency band, while harmonic components occupy wavelet detail bands with oscillatory patterns distinct from reflections. W-U-Net performs nonlinear shaping of these coefficients and predicts cleaned coefficients

$$(\hat{\mathbf{c}}_a, \hat{\mathbf{c}}_d) = \lambda_w(\mathbf{c}_a, \mathbf{c}_d) \quad (9)$$

where $\hat{\mathbf{c}}_a$ is the predicted low-frequency background and $\hat{\mathbf{c}}_d$ is the predicted high-frequency detail. $\lambda_w(\cdot)$ denotes the branch's nonlinear mapping. During training, supervision uses the clean signal $\mathbf{x}$ via its wavelet coefficients $(\mathbf{c}_a^x, \mathbf{c}_d^x)$, guiding the branch to learn clean wavelet coefficients and to form an explicit, band-wise separable representation of harmonic components.

2.3.3. ContMix1D: content-adaptive local mixing block

Let the input be $\mathbf{y} \in \mathbb{R}^{B \times C \times L}$. A pointwise 1×1 convolution first projects the C dimensional features at each position into two sets of unnormalized logits for a small and a large neighborhood of lengths k_s and k_l

$$\mathbf{T}(\mathbf{y}) = \text{Conv}_{1 \times 1}(\mathbf{y}) \in \mathbb{R}^{B \times (k_s + k_l) \times L} \quad (10)$$

which we split as $\mathbf{T}_s \in \mathbb{R}^{B \times k_s \times L}$ and $\mathbf{T}_l \in \mathbb{R}^{B \times k_l \times L}$. For each position ℓ , softmax along the kernel dimension yields location-dependent mixing weights

$$\begin{aligned} \alpha_s(\ell) &= \text{softmax}(\mathbf{T}_s[:, :, \ell]) \in \mathbb{R}^{k_s} \\ \alpha_l(\ell) &= \text{softmax}(\mathbf{T}_l[:, :, \ell]) \in \mathbb{R}^{k_l} \end{aligned} \quad (11)$$

Hence at any ℓ , the entries of $\alpha_s(\ell)$ and $\alpha_l(\ell)$ are nonnegative and sum to one. They act as content-adaptive coefficients for the small and large neighborhoods. Using a 1D unfold operator, we extract two local windows centered at each position

$$\mathbf{U}_{k_s}(\mathbf{y}), \mathbf{U}_{k_l}(\mathbf{y}) \in \mathbb{R}^{B \times C \times k_{\ell} \times L} \quad (12)$$

where the 3D indexes kernel positions. The two windows are then combined by weighted sums across the kernel dimension and added together

$$\mathbf{m}(\ell) = \sum_{i=1}^{k_s} \alpha_{s,i}(\ell) \mathbf{U}_{k_s}(\mathbf{y})_{:, :, i, \ell} + \sum_{i=1}^{k_l} \alpha_{l,i}(\ell) \mathbf{U}_{k_l}(\mathbf{y})_{:, :, i, \ell}. \quad (13)$$

This operation performs content-adaptive aggregation at two scales. The small kernel sharpens edges, whereas the large kernel supplies a wider receptive field. The weights α vary with location, which allows the operator to follow dips and breakpoints.

To inject explicit positional information and stabilize the mixing, we add a multi-dilation depthwise convolutional term as a positional encoding

$$\mathbf{P} = \sum_{d \in \mathcal{D}} \text{DWConv}_{k, p_d, d}(\mathbf{y}), \mathcal{D} = \{1, 2, 3, \dots\} \quad (14)$$

where d is the dilation rate and $p_d = d(k-1)/2$ provides symmetric padding to keep the length. For batch b , channel c , and position ℓ , the d -th depthwise component is

$$\left[\text{DWConv}_{k, p_d, d}(\mathbf{y}) \right]_{b, c, \ell} = \sum_{i=0}^{k-1} \omega_{c, i}^{(d)} \mathbf{y}_{b, c, \ell + id - p_d} + b_c^{(d)} \quad (15)$$

with depthwise kernel $\omega_{c, i}^{(d)}$ and bias $b_c^{(d)}$. The final mixed feature is

$$\mathbf{u} = \mathbf{m} + \mathbf{P}. \quad (16)$$

ContMix1D is implemented by unfold plus weighted summation and by depthwise convolutions. Its computational complexity is $O((k_s + k_l)CL)$, which is comparable to standard local convolutions but provides content adaptivity. Unlike pairwise self-attention with quadratic cost $O(L^2)$, it avoids global similarity computation and remains linear in sequence length, making it well-suited to long 1D seismic signals.

2.3.4. MS-IF: multiscale differentiable inverse wavelet fusion

Within the network, we use a differentiable inverse discrete wavelet transform (IDWT) to project the wavelet-branch coefficients back into the time domain and fuse them with the backbone features. This design enables end-to-end supervision using time-domain losses. The IDWT is a fixed linear operator determined by the chosen synthesis filters, and it is differentiable with respect to both $\mathbf{c}_a$ and $\mathbf{c}_d$. Gradients therefore propagate through the reconstruction map to upstream modules that generate these coefficients, ensuring effective and stable end-to-end training without introducing additional learnable parameters. The one-level IDWT is given as

$$\mathbf{R}[m] = \sum_k \tilde{h}[k] (\uparrow_2 \mathbf{c}_a)[m - k] + \sum_k \tilde{g}[k] (\uparrow_2 \mathbf{c}_d)[m - k] \quad (17)$$

where $\mathbf{R}$ is the reconstructed signal, $\hat{h}$ and $\hat{g}$ are the synthesis low-pass and high-pass filters, and $\uparrow_2$ denotes upsampling by a factor of two via zero insertion.

2.3.5. RM: reconstruction module

Assume the decoder has J scales. Let $\mathbf{S}_j$ denote the skip feature from the j -th encoder level, and let $\mathbf{R}_j$ be the time-domain feature at the same level obtained from the wavelet branch via differentiable IDWT. Define $\phi(\cdot)$ as a Conv $\rightarrow$ BN $\rightarrow$ LeakyReLU block, Up $(\cdot)$ as upsampling, and $[\cdot]$ as channel concatenation. The RM module propagates from the bottleneck to higher resolutions, progressively fusing backbone and side-branch information. The fusion at each level is written as

$$\begin{aligned} \mathbf{z}_{j+1} &= \mathbf{x}_{\text{mid}} \\ \mathbf{z}_j &= \phi\left([\text{Up}(\mathbf{z}_{j+1}), \mathbf{S}_j, \mathbf{R}_j]\right), j = J, J-1, \dots, 1 \end{aligned} \quad (18)$$

where $\mathbf{x}_{\text{mid}}$ is the bottleneck feature.

At the topmost fusion, in addition to $\mathbf{z}_1$ and $\mathbf{R}_1$, we introduce a direct time-domain reconstruction

$$\hat{\mathbf{R}} = \text{IDWT}(\hat{\mathbf{c}}_a, \hat{\mathbf{c}}_d) \quad (19)$$

from the wavelet-branch predictions and modulate its strength by a learnable gate $g = \sigma(\omega_{\text{gate}}) \in (0, 1)$. ω_{gate} is a learnable scalar parameter. The final fused feature is

$$\mathbf{z}_{\text{final}} = \text{concat}(\mathbf{z}_1, g \cdot \hat{\mathbf{R}}). \quad (20)$$

A reconstruction head directly regresses the clean trace from this fused feature

$$\hat{\mathbf{x}} = \mathbf{H}(\mathbf{z}_{\text{final}}) \quad (21)$$

where $\mathbf{H}(\cdot)$ is the decoder prediction head (3×3 Conv $\rightarrow$ BN $\rightarrow$ LeakyReLU $\rightarrow 1 \times 1$ Conv) that combines the backbone and side-branch predictions.

This design lets the side branch provide physically interpretable, band-wise priors at every scale to enhance semantic recovery in the backbone. Because IDWT is a fixed linear operator, the entire reconstruction path is end-to-end differentiable and time-domain losses back-propagate effectively to $(\hat{\mathbf{c}}_a, \hat{\mathbf{c}}_d)$ and upstream parameters.

2.3.6. Loss function

Training minimizes a weighted sum of three terms: a time-domain reconstruction term, wavelet-coefficient supervision, and a synthesis-domain consistency term

$$\begin{aligned} L &= L_{\text{main}}(\hat{\mathbf{y}}, \mathbf{x}) + \omega_{\text{ca}} L_{\text{coef}}(\hat{\mathbf{c}}_a, \mathbf{c}_a(\mathbf{x})) + \omega_{\text{cd}} L_{\text{coef}}(\hat{\mathbf{c}}_d, \mathbf{c}_d(\mathbf{x})) \\ &\quad + \omega_{\text{rec}} L_{\text{rec}}(\text{IDWT}(\hat{\mathbf{c}}_a, \hat{\mathbf{c}}_d), \mathbf{x}) \end{aligned} \quad (22)$$

where ω_{ca} , ω_{cd} , and ω_{rec} are the balancing hyperparameters for the corresponding loss terms.

The backbone is constrained by an MSE term to preserve amplitude fidelity and in-phase continuity

$$L_{\text{main}}(\hat{\mathbf{y}}, \mathbf{x}) = \|\hat{\mathbf{y}} - \mathbf{x}\|_2^2 \quad (23)$$

where $\|\cdot\|_2$ denotes the L_2 norm.

To align the W-U-Net band behavior with the clean coefficients and suppress harmonic leakage into detail sub-bands, we use a robust and insensitive L_1 penalty

$$L_{\text{coef}}(\hat{\mathbf{c}}, \mathbf{c}) = \frac{1}{N} \|\max(|\hat{\mathbf{c}} - \mathbf{c}| - \varepsilon, 0)\|_1, \varepsilon = 10^{-3} \quad (24)$$

where $\|\cdot\|_1$ denotes the L_1 norm, N denotes the total number of elements in the coefficient tensor. Finally, the predicted coefficients are synthesized to the time domain through the fixed, differentiable IDWT and compared to the reference via MSE

$$L_{\text{rec}} = \|\text{IDWT}(\hat{\mathbf{c}}_a, \hat{\mathbf{c}}_d) - \mathbf{x}\|_2^2. \quad (25)$$

Beyond direct coefficient supervision, this synthesis-domain term enforces band-wise consistency between the side branch and the backbone, enhancing their cooperative stability.

3. Model testing and field application

3.1. Model testing

To evaluate the harmonic noise suppression of DW-U-Net, we conduct numerical experiments on the Marmousi2 velocity model. The model spans $11 \text{ km} \times 4 \text{ km}$ (Fig. 4). A symmetric split-spread geometry is used, with the shot at the center of the spread and receivers deployed on both sides. Each shot uses 400 receivers, and 28 shots are simulated. Two source signatures are considered: an ideal linear sweep signal and a sweep contaminated by harmonics. The linear sweep signal has a length of 12 s, a taper of 0.3 s, a bandwidth of 4–84 Hz, and a sampling interval of 1 ms, as shown in Fig. 5(a). Harmonic contamination is constructed by injecting up to the fourth order into the instantaneous phase of the sweep, with amplitude ratios $a_m \in \{0.1, 0.3, 0.6\}$ representing different noise levels. $M = 4$ denotes the fundamental plus the second to fourth harmonics. In addition, a near-surface resonance with amplitude 0.1 and frequency $f_0 = 20 \text{ Hz}$ is superimposed, yielding multi-level harmonically contaminated sweeps (Fig. 5(b)–(d)).

Fig. 6 provides zoomed views at the beginning, middle, and end of the sweep signals, comparing the clean trace with harmonically contaminated cases under different a_m . As a_m increases, the waveform shows progressively stronger high-frequency ripples and larger peak/trough deviations from the clean sweep signal, making the harmonic distortion more visually discernible. The prepared source signatures and velocity model are used in forward modeling to generate 28 single-shot records, each with a recording length of 4 s.

Fig. 7 compares a noisy trace with its clean reference from a representative shot. The nine panels are arranged in a 3×3 grid. The columns, from left to right, correspond to traces 1, 100, and 200, representing far, mid, and near offsets, respectively. The rows, from top to bottom, correspond to three time windows: 0–1333 ms, 1334–2666 ms, and 2667–4000 ms. In each panel, the left curve shows the noisy trace, while the right curve shows the clean reference. Harmonic energy increases markedly as the offset decreases. Near offsets exhibit denser high-frequency oscillations

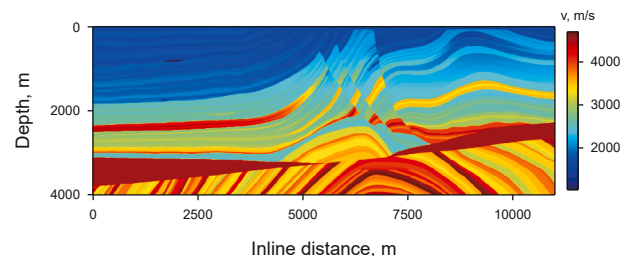


Fig. 4. Marmousi2 velocity model.

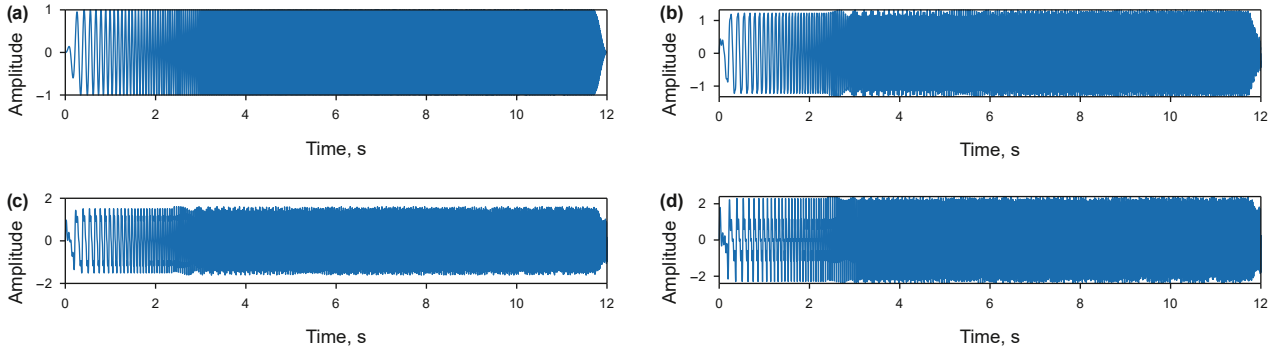


Fig. 5. Time-domain waveforms of sources. (a) Clean linear sweep signal, (b) harmonic-contaminated sweep with $a_m = 0.1$, (c) harmonic-contaminated sweep with $a_m = 0.3$, (d) harmonic-contaminated sweep with $a_m = 0.6$.

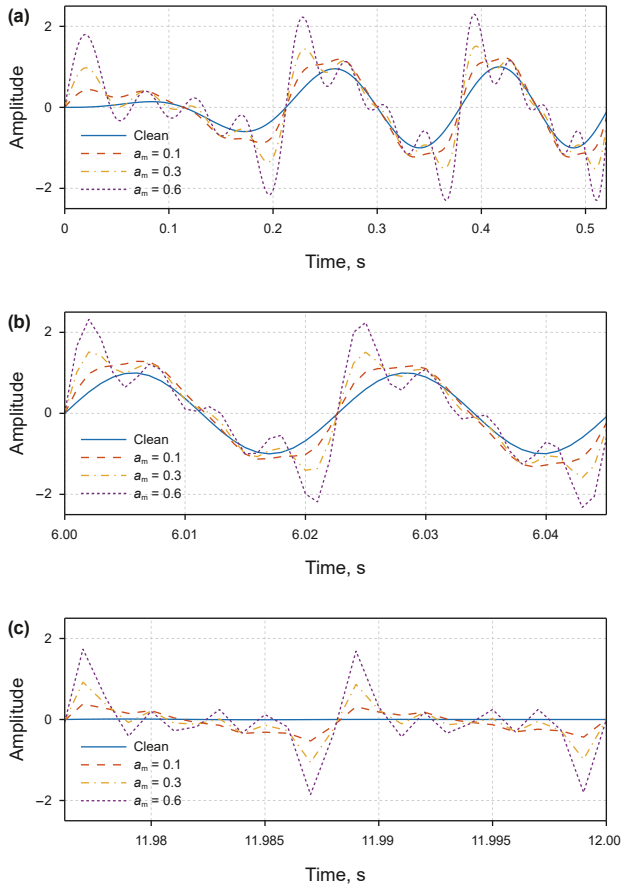


Fig. 6. Clean sweep signal versus harmonically contaminated sweep signals at different noise levels. (a) 0–0.52 s, (b) 6–6.045 s, (c) 11.976–12 s.

and pronounced envelope modulation. Mid offsets are moderately affected. Far offsets are least affected, and their waveforms more closely resemble the reference. For the second row, the segmental SNRs in Fig. 7(d)–(f) are 16.71 dB, 11.40 dB, and –20.12 dB, respectively, indicating that near offsets are the critical range for harmonic suppression and quality control. The SNR for a window is computed as follows

$$\text{SNR} = 10 \log_{10} \frac{\|\mathbf{d}_{\text{clean}}\|_F^2}{\|\mathbf{d}_{\text{clean}} - \mathbf{d}_{\text{observe}}\|_F^2} \quad (26)$$

where $\|\cdot\|_F$ denotes the Frobenius norm. $\mathbf{d}_{\text{clean}}$ is the clean reference trace and $\mathbf{d}_{\text{observe}}$ is the noisy trace.

We feed 11,200 samples into DW-U-Net, each consisting of a standardized noisy signal along with its first-level discrete wavelet transform (DWT) coefficients, using the haar wavelet basis. The training settings are as follows: batch size of 128, 200 epochs, and the Adam optimizer. The model comprises two branches—T-U-Net and W-U-Net—optimized with Adam using initial learning rates of 5×10^{-4} and 3×10^{-4} , respectively, which are further updated with a cosine-annealing schedule during training. A fixed validation set is created at the start by holding out 10% of the training data. Validation is performed every two epochs, and the checkpoint with the lowest validation loss is saved as the final model. The network uses a base channel width of 32 and a depth of four levels.

At test time, standardized noisy single-shot records with varying harmonic levels are passed through the trained model. The normalization parameters are stored during training and later used to inverse-standardize the predictions, yielding the final denoised results.

We select five representative denoising networks for comparison. Classical DnCNN and U-Net are used as CNN baselines. We further introduce DURN with a dual-branch design to evaluate the effectiveness of a dual-branch architecture for this task (Liu et al., 2019). In addition, we include MIMO-U-Net, which adopts multi-scale and multi-input-multi-output mechanisms, to cover methods based on multiscale information fusion (Cho et al., 2021). In this study, MIMO-U-Net is implemented in a 1D manner for harmonic noise attenuation using the same time-domain single-trace training data as DW-U-Net. The multi-input data are formed by generating several 1D scaled versions of the same noisy trace and feeding them into different stages of a shared U-Net encoder. The decoder generates denoised outputs at the corresponding scales, which serve as intermediate restoration results for multi-scale supervision. Through progressive upsampling and propagation of low-scale decoder features to higher-resolution decoder stages, the network finally produces the denoised trace at the original resolution. All baselines are trained under the same data split and the same cosine-annealing optimization schedule for fair comparison.

Fig. 8 shows the evolution of the training MSE loss versus epoch for the five networks. All models exhibit a rapid loss decrease in the early training stage, where MIMO-U-Net drops faster within a small number of epochs, indicating a higher initial convergence speed. As training proceeds, all models gradually reach a stable convergence regime. The zoomed-in view of the late-stage curves shows that DW-U-Net maintains a lower training loss after convergence and ultimately achieves the smallest loss among the

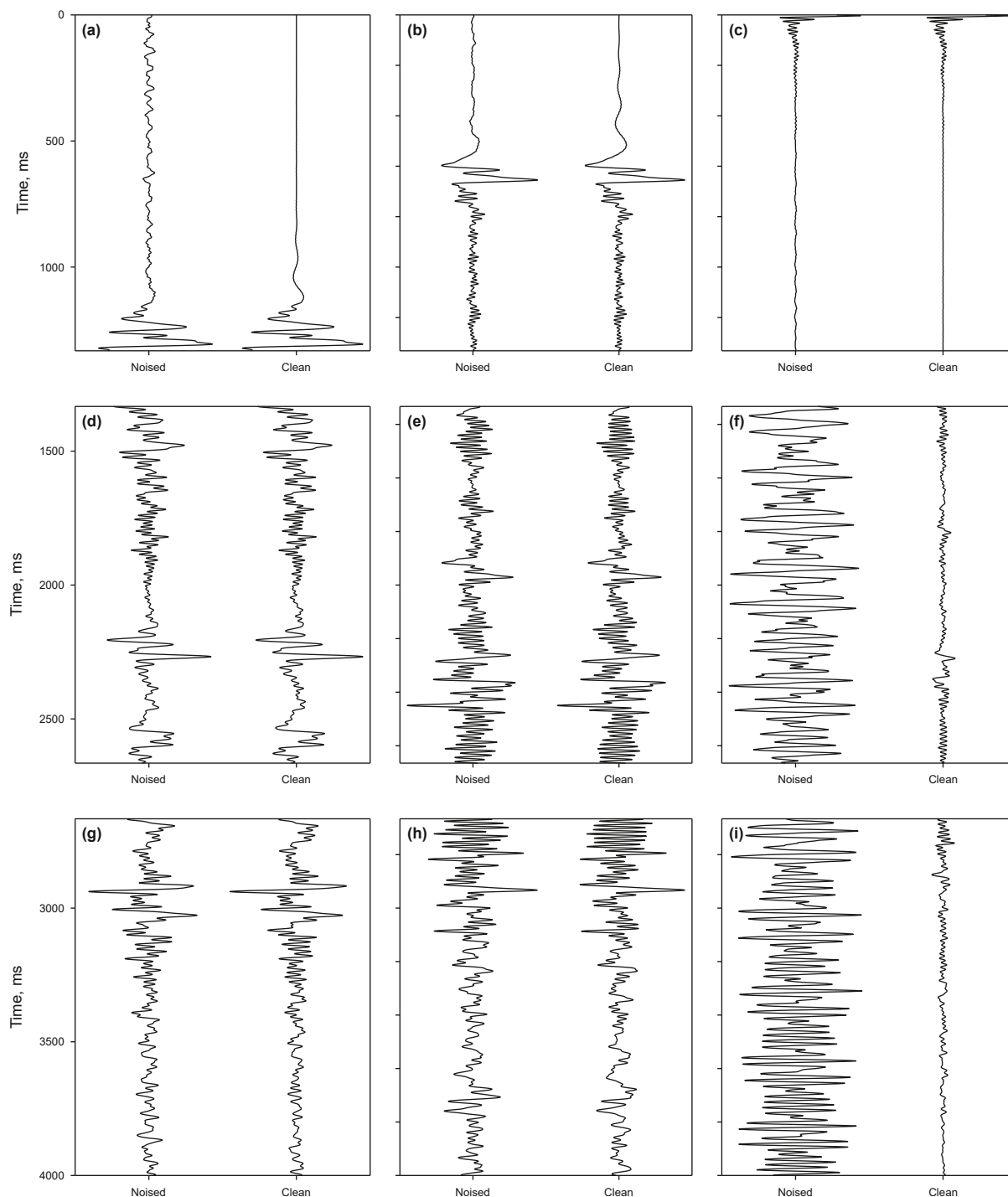


Fig. 7. Noisy-clean trace pairs at three offsets and three time windows. (a) Far offsets, 0–1333 ms, (b) mid offsets, 0–1333 ms, (c) near offsets, 0–1333 ms, (d) far offsets, 1334–2666 ms, (e) mid offsets, 1334–2666 ms, (f) near offsets, 1334–2666 ms, (g) far offsets, 2667–4001 ms, (h) mid offsets, 2667–4001 ms, (i) near offsets, 2667–4001 ms.

five. These results suggest that, under the same training configuration, DW-U-Net has stronger fitting capability for harmonic noise suppression, providing training-level evidence supporting its subsequent denoising performance advantages.

Fig. 9 shows the single-shot records used in our experiments. Fig. 9(a) is the clean single-shot record, while Fig. 9(b)–(d) are the harmonically contaminated inputs with $a_m = 0.1, 0.3$, and 0.6 , corresponding to low, medium, and strong noise levels, respectively. As a_m increases, the coherent harmonic streaks become

stronger and denser, with more pronounced contamination on small-offset traces. Such source-related interference persists over the full recording time window and increasingly overlaps the reflection band, masking weak events.

Fig. 10 presents single-shot denoised results under three noise levels for DnCNN, U-Net, DURN, MIMO-U-Net, and the proposed DW-U-Net. Table 1 reports the whole-shot average SNRs for the contaminated inputs and the denoised outputs under three noise levels. Overall, all learning-based baselines improve SNR

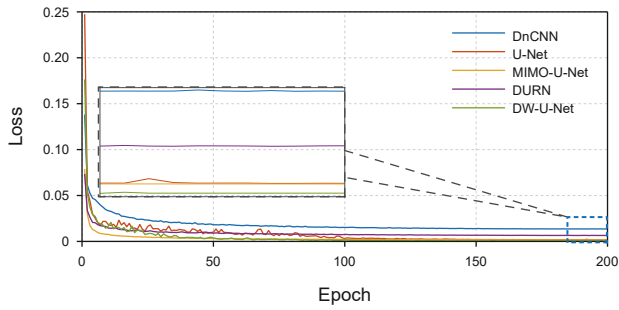


Fig. 8. Comparison of training MSE loss among the five networks.

compared with the input, while the proposed DW-U-Net achieves the best global SNR under the medium and strong noise settings. To evaluate performance in a more challenging region, Table 2 presents the SNR computed on a near-offset trace window (traces 150–250), where harmonic streaks strongly overlap the reflections and can overwhelm the useful signal. In this window, the input SNR becomes negative, indicating that the interference energy exceeds the signal energy. Despite this, DW-U-Net consistently delivers the highest SNR across all three noise levels, demonstrating stronger robustness in heavily contaminated near-offset data.

Under the low-noise level (input average SNR = 18.75 dB; Fig. 9(b)), DnCNN (Fig. 10(a)) attenuates a large portion of the harmonic interference but still leaves visible residual streaks. U-Net (Fig. 10(b)) and DURN (Fig. 10(c)) further reduce these artifacts, while the multiscale MIMO-U-Net (Fig. 10(d)) and the proposed DW-U-Net (Fig. 10(e)) produce the cleanest records in the global view. Quantitatively (Table 1), MIMO-U-Net achieves the highest whole-shot SNR of 28.39 dB, closely followed by DW-U-Net at 27.85 dB, corresponding to SNR gains of +9.64 dB and +9.10 dB over the noisy input, respectively.

Under the medium noise level (input average SNR = 15.20 dB; Fig. 9(c)), the overall appearance in Fig. 10(f)–(j) remains visually similar to the low-noise case from a global perspective, making the performance gap less obvious by inspection alone. Nevertheless, Table 1 shows a clear quantitative advantage: DW-U-Net reaches 30.18 dB, surpassing MIMO-U-Net (26.03 dB) by 4.15 dB, U-Net (22.64 dB) by 7.54 dB, DURN (21.10 dB) by 9.08 dB, and DnCNN (18.05 dB) by 12.13 dB, yielding an overall improvement of +14.98 dB over the input. This advantage becomes more pronounced in the challenging near-offset region (Table 2, traces

150–250), where the input SNR is –8.39 dB due to severe harmonic dominance, yet DW-U-Net increases it to 15.76 dB, substantially higher than MIMO-U-Net (8.76 dB) and U-Net (7.57 dB).

Under strong harmonic noise (input average SNR = 9.92 dB; Fig. 9(d)), the red box in Fig. 10(k)–(o) marks a representative region with particularly intense harmonic streaks, and the corresponding enlargements are shown in Fig. 11(a)–(f). In this window, the noisy input (Fig. 11(a)) exhibits dense, high-amplitude stripe-like events that obscure underlying reflections. After DnCNN (Figs. 10(k) and 11(b)), many stripes remain and some appear broken or distorted, indicating limited suppression under heavy contamination. DURN (Figs. 10(m) and 11(d)) mitigates part of the streaking but still leaves noticeable coherent remnants. MIMO-U-Net (Figs. 10(n) and 11(e)) further attenuates the harmonics, yet residual streaks and local amplitude inconsistency are still observable. U-Net (Figs. 10(l) and 11(c)) reduces the stripes more strongly, but reflection events become less continuous with evident fragmentation in places. In contrast, DW-U-Net (Figs. 10(o) and 11(f)) suppresses most harmonic streaks while better preserving reflection continuity and more consistent amplitudes.

To further assess whether the networks suppress harmonic interference while preserving useful reflections, we visualize the removed components under the strong-noise case in Fig. 12. The removed component represents the portion extracted from the noisy input by each method; ideally, it should be dominated by the stripe harmonic patterns and contain as little coherent reflection-like energy as possible. The reference removed components in Fig. 12(a) provide a benchmark pattern of the harmonic-dominated energy. Compared with this reference, DnCNN (Fig. 12(b)) removes substantial harmonic energy but still exhibits noticeable segments of coherent, in-phase structures, suggesting that part of the primary signal is also removed. DURN (Fig. 12(d)) produces a more noise-like removed component overall, yet structured energy persists around strong events and small-scale features, implying mild signal leakage. U-Net (Fig. 12(c)) further reduces such leakage, while MIMO-U-Net (Fig. 12(e)) concentrates the removed energy more tightly along the harmonic trends. In contrast, the DW-U-Net removed component (Fig. 12(f)) most closely resembles a near zero-mean, high-frequency texture: the extracted energy aligns mainly with the harmonic streak trajectories, and laterally coherent reflection-like events are rarely visible, indicating minimal leakage of useful signal.

The blue dashed boxes in Fig. 12 highlight a representative far-offset region, and their enlargements are shown in Fig. 13. In this zoomed view, competing methods still retain mixed patterns

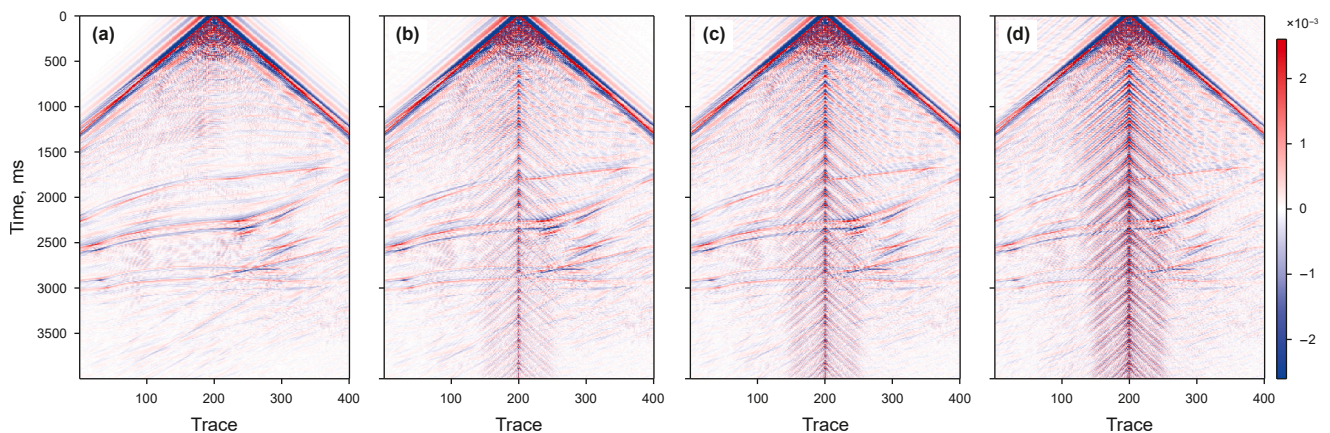


Fig. 9. Clean single-shot record and single-shot records contaminated by harmonic interference at different amplitudes. (a) Clean single-shot record; (b) harmonic-contaminated single-shot record with $a_m = 0.1$, (c) harmonic-contaminated single-shot record with $a_m = 0.3$, (d) harmonic-contaminated single-shot record with $a_m = 0.6$.

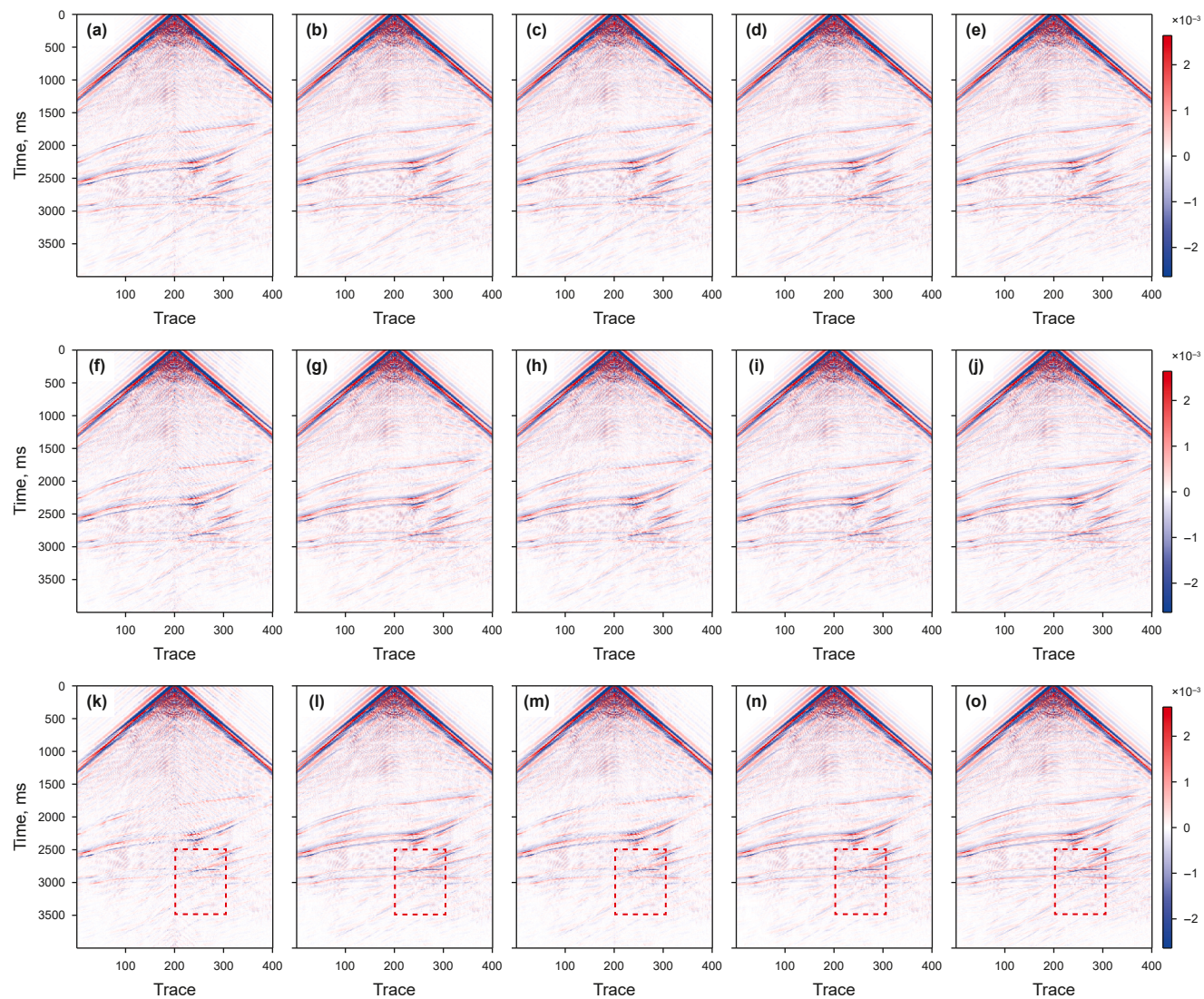


Fig. 10. Denoised results under different noise levels. (a) DnCNN at low noise, (b) U-Net at low noise, (c) DURN at low noise, (d) MIMO-U-Net at low noise, (e) DW-U-Net at low noise, (f) DnCNN at medium noise, (g) U-Net at medium noise, (h) DURN at medium noise, (i) MIMO-U-Net at medium noise, (j) DW-U-Net at medium noise, (k) DnCNN at strong noise, (l) U-Net at strong noise, (m) DURN at strong noise, (n) MIMO-U-Net at strong noise, (o) DW-U-Net at strong noise.

Table 1
Whole-shot average SNR of different denoised models under three harmonic noise levels.

Noisy, dB	DnCNN	U-Net	DURN	MIMO-U-Net	DW-U-Net (ours)
18.75 ($a_m = 0.1$)	20.19	24.28	23.40	28.39	27.85
15.20 ($a_m = 0.3$)	18.05	22.64	21.10	26.03	30.18
9.92 ($a_m = 0.6$)	14.49	20.71	17.43	22.25	29.29

Table 2
Near-offset SNR evaluated on traces 150–250 under three harmonic noise levels.

Noisy, dB	DnCNN	U-Net	DURN	MIMO-U-Net	DW-U-Net (ours)
−5.45 ($a_m = 0.1$)	3.26	10.74	5.74	9.56	12.99
−8.39 ($a_m = 0.3$)	2.32	7.57	4.86	8.76	15.76
−12.33 ($a_m = 0.6$)	1.12	7.29	3.15	6.59	13.37

where harmonic stripes coexist with residual coherent structures, whereas DW-U-Net captures the far-offset harmonic streaks more cleanly and leaves fewer reflection-shaped remnants. This observation supports that DW-U-Net primarily removes harmonic

contamination rather than subtracting meaningful reflection energy, consistent with its improved denoising performance in the strong-noise setting.

Fig. 14 presents the absolute prediction error maps under the strong harmonic noise condition. Brighter colors indicate larger deviations from the clean reference. Overall, the error patterns of DnCNN (Fig. 14(a)) are the most widespread, with prominent energy distributed along the harmonic streak trajectories and around high-amplitude events, suggesting both residual harmonics and imperfect recovery of reflections. DURN (Fig. 14(c)) reduces the error magnitude compared with DnCNN, yet noticeable error remains in the mid-to-late time range and along coherent dipping features. U-Net (Fig. 14(b)) yields a cleaner error map than the previous baselines, although residual errors are still visible along the harmonic trends and in some reflection-dominated regions. MIMO-U-Net (Fig. 14(d)) further suppresses part of the streak-related error, but still shows scattered mismatch over reflection bands and locally enhanced error around complex structures. In contrast, DW-U-Net (Fig. 14(e)) exhibits the lowest overall error level, with most areas remaining near the dark background and only weak, sparse error traces, indicating that its predictions are

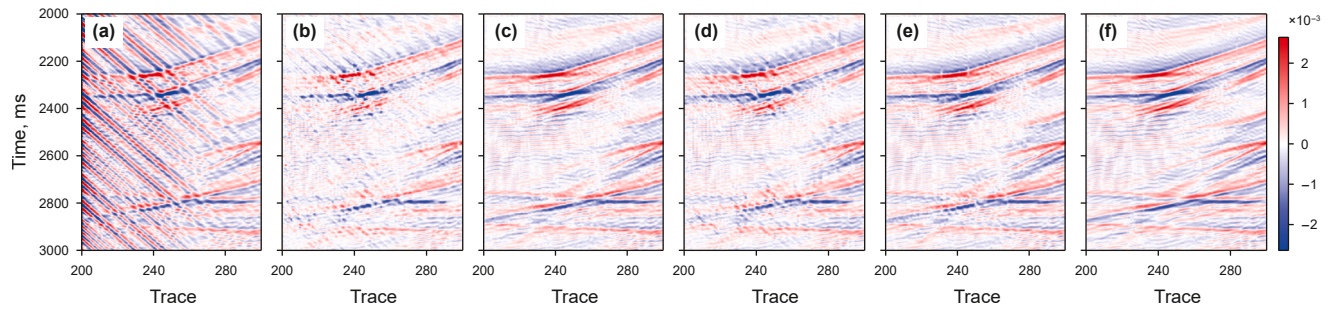


Fig. 11. (a) Zoomed view of Fig. 9(d), (b) zoomed view of Fig. 10(k), (c) zoomed view of Fig. 10(l), (d) zoomed view of Fig. 10(m), (e) zoomed view of Fig. 10(n), (f) zoomed view of Fig. 10(o).

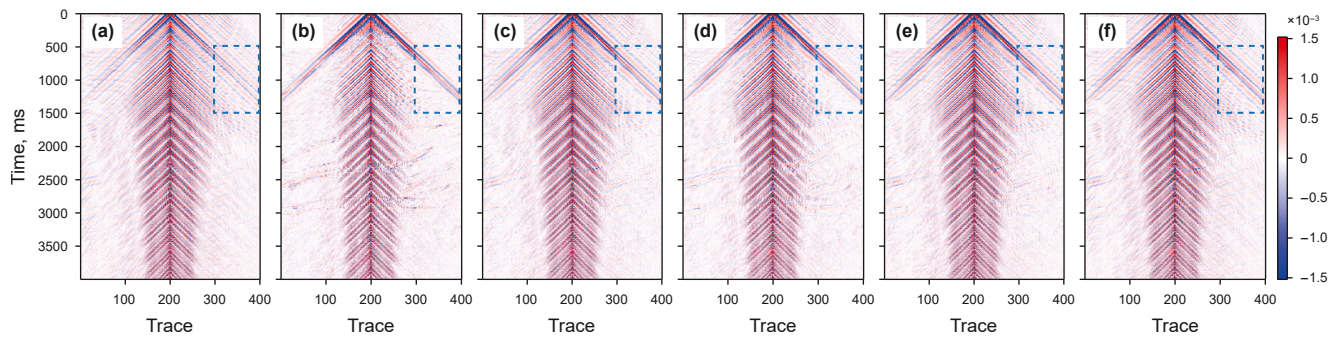


Fig. 12. Removed components. (a) Reference removed components, (b) DnCNN, (c) U-Net, (d) DURN, (e) MIMO-U-Net, (f) DW-U-Net.

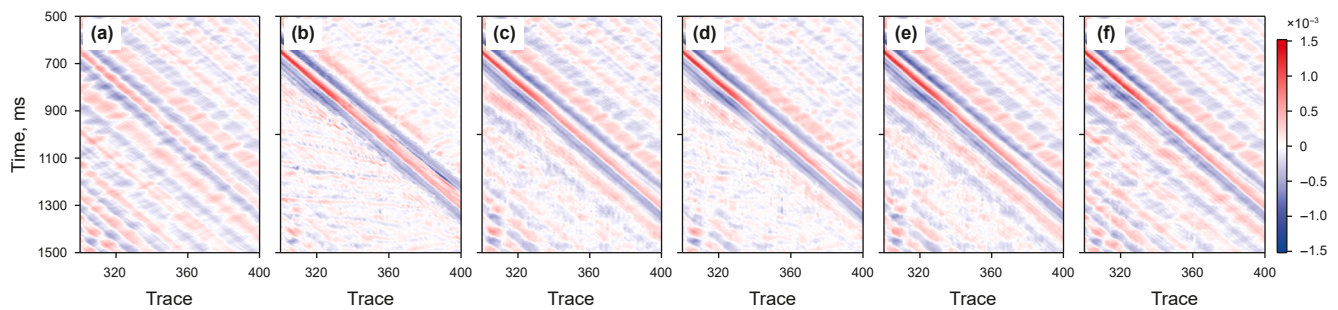


Fig. 13. (a) Zoomed view of Fig. 12(a), (b) zoomed view of Fig. 12(b), (c) zoomed view of Fig. 12(c), (d) zoomed view of Fig. 12(d), (e) zoomed view of Fig. 12(e), (f) zoomed view of Fig. 12(f).

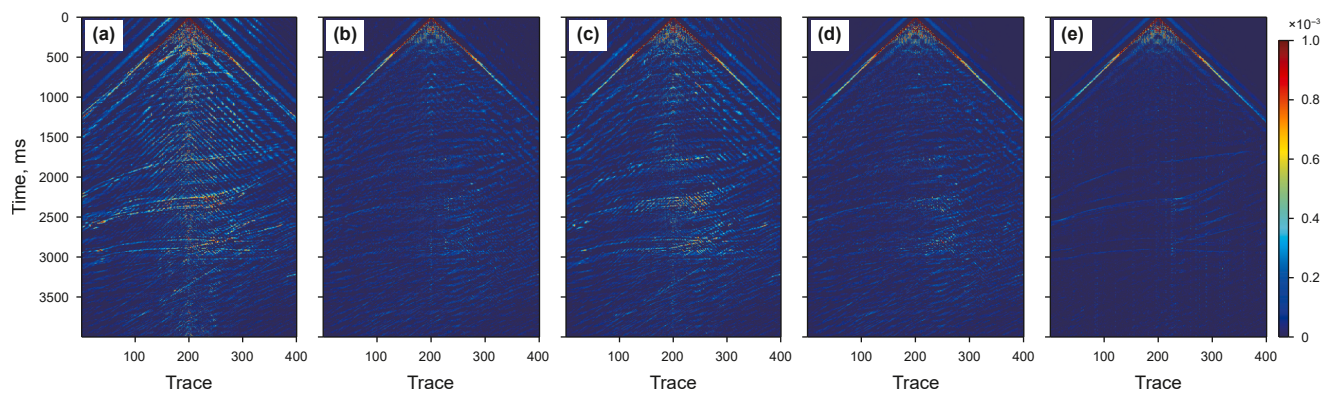


Fig. 14. Absolute prediction error maps under strong harmonic noise. (a) DnCNN, (b) U-Net, (c) DURN, (d) MIMO-U-Net, (e) DW-U-Net.

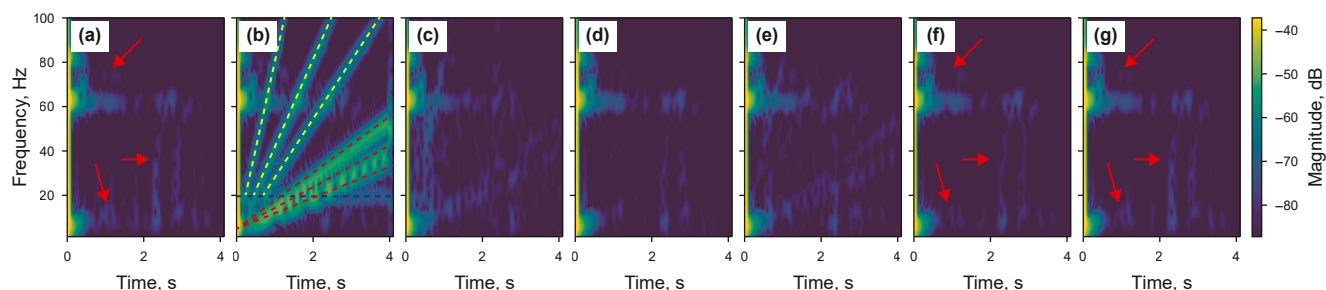


Fig. 15. Time-frequency spectra comparison. (a) Ground truth, (b) raw observation with harmonics and resonance, (c) DnCNN, (d) U-Net, (e) DURN, (f) MIMO-U-Net, (g) DW-U-Net.

closest to the ground truth (GT) under strong harmonic contamination.

Time-frequency evaluation of harmonic suppression and spectral fidelity under strong harmonic interference is shown in Fig. 15. The figure presents the STFT of the central trace. The horizontal axis is time, the vertical axis is frequency, and the color scale encodes amplitude in dB. Because the spectrogram is computed on the trace after correlation with the pilot, the harmonic components are mixed with the fundamental and spread across lags, so they do not appear as three isolated lines. In addition, the spectrogram is quadratic in the signal, so when a linear sweep and a constant-frequency resonance fall within the same analysis window, cross-terms emerge. These cross-terms generate companion energy ridges between the sweep ridge and the constant line, revealing interference.

Fig. 15(a) shows the clean reference, where energy is mainly confined to the reflection band (approximately 10–80 Hz) with good temporal continuity; no evident high-order harmonic ridges or persistent narrowband buildup is observed.

Fig. 15(b) shows the raw observation with harmonics and resonance. Multiple oblique ridges appear along the sweep trajectory. The red lines mark the second-, third-, and fourth-order harmonics, from top to bottom. The yellow lines indicate the companion ridges produced by cross-terms. A narrowband line at 20 Hz traverses the entire window, evidencing resonance. Together, harmonics and resonance elevate the background and obscure details within the main band.

Fig. 15(c)–(g) compare the denoised spectra. DnCNN (Fig. 15(c)) partially suppresses harmonic-related ridges, yet residual streak-like energy and mild leakage remain within the passband, indicating an incomplete compromise between suppression and fidelity. U-Net (Fig. 15(d)) reduces the interference more aggressively, but the main-band texture becomes weaker and less continuous in places, suggesting that part of the useful reflection energy is also attenuated. DURN (Fig. 15(e)) better preserves the reflection band while still leaving noticeable residual interference energy. MIMO-U-Net (Fig. 15(f)) and DW-U-Net (Fig. 15(g)) provide a clearer balance. The harmonic-related ridge structures and resonance components are more strongly attenuated, while the key reflection-band energy patches are retained with improved temporal continuity. Moreover, as indicated by the arrows, the DW-U-Net spectrogram is closer to that in Fig. 15(a), preserving more frequency-energy components of the useful signal. Among all methods, DW-U-Net yields the closest match to the clean reference, exhibiting the weakest residual ridge structures and the cleanest background in the time-frequency domain, which indicates improved spectral fidelity under strong harmonic contamination.

Across the spatial, residual, and time-frequency domains, DW-U-Net achieves the best quantitative and qualitative performance

at all three noise levels, and its advantage widens as noise increases. This robustness stems from wavelet-based multiscale modeling that targets the structural signatures of harmonic streaks at multiple scales. Cross-level feature fusion and fine-grained reconstruction mitigate over-smoothing, preserve event continuity and weak-reflection detail, and sharpen the separation between harmonics and primary reflections, yielding residuals that are noise-like and outputs that are signal-like.

Fig. 16 compares a representative single-trace waveform segment (800–850 ms) among the clean reference and the denoised outputs. Clear differences are observed in phase alignment, amplitude fidelity, and waveform shape preservation. DW-U-Net shows the closest match to the ground truth, with nearly coincident peak-trough positions and the most consistent peak/valley amplitudes, indicating that it suppresses harmonic interference while preserving reflection wavelets. MIMO-U-Net also follows the GT trend well, but exhibits minor amplitude deviations at several extrema. In contrast, U-Net produces visibly smoother curves with compressed extrema, suggesting mild over-smoothing and attenuation of useful signal components. DnCNN shows larger mismatches in both amplitude and local waveform shape, implying less stable reconstruction under strong harmonic contamination. DURN deviates the most, with noticeable overshoot and waveform distortion. Overall, DW-U-Net achieves the best denoising-fidelity trade-off in this strong-noise scenario, retaining key reflection details while effectively reducing interference.

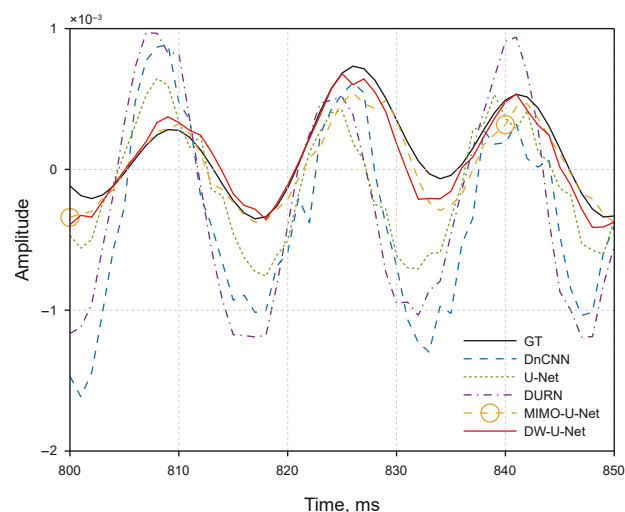


Fig. 16. Single-trace waveform comparison under strong harmonic noise.

Table 3
Residual energy and RMSE of different models under strong harmonic noise.

$a_m = 0.6$	DnCNN	U-Net	DURN	MIMO-U-Net	DW-U-Net (ours)
Residual Energy	3.6462e−2	9.5441e−3	2.1429e−2	9.96e−3	2.4446e−3
RMSE	1.7428e−04	8.9167e−05	1.3361e−04	9.1089e−05	4.5127e−05

Table 4
MS-SSIM of different models under strong harmonic noise.

$a_m = 0.6$	DnCNN	U-Net	DURN	MIMO-U-Net	DW-U-Net (ours)
MS-SSIM	0.929676	0.974880	0.948302	0.975061	0.995853

Below we quantify the reconstruction error between each denoised output $\hat{\mathbf{y}}$ and the clean reference $\mathbf{x}$. Let the sample-wise residual be

$$\mathbf{E} = \hat{\mathbf{y}} - \mathbf{x}. \quad (27)$$

The residual energy and Root Mean Square Error (RMSE) used in Table 3 are computed as

$$E_{\text{res}} = \sum_{i=1}^N E_i^2, \quad \text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N E_i^2}, \quad (28)$$

where N is the total number of samples in the evaluated traces.

Table 3 shows that DW-U-Net yields the smallest errors under strong harmonic contamination. These results confirm that DW-U-Net produces outputs closest to the ground truth in a squared-error sense, indicating more effective suppression of strong harmonics while better preserving the desired seismic signal.

Table 4 reports the MultiScale Structural Similarity (MS-SSIM) values under strong harmonic noise where a higher score indicates better preservation of multiscale structural information relative to the ground truth. DW-U-Net achieves the highest MS-SSIM of 0.995853. This result is consistent with the visual comparisons: DW-U-Net more effectively suppresses harmonic-related artifacts

while maintaining reflection continuity and fine-scale details, leading to the closest structural match to the clean reference.

3.2. Field application

To assess practical applicability, we evaluate the proposed method on the public Utah FORGE field dataset and select a representative single-shot record from Line 302. The source uses two vibroseis trucks separated by 30 ft, each rated at 62,000 lbf peak force and operated at 70% output. A 12 s linear sweep (4–84 Hz) is repeated four times at each source point. Such vibroseis acquisition is susceptible to source-related nonlinearities and coupling effects, which may generate coherent harmonic streaks and banded interference in single-shot records. This dataset therefore provides a realistic testbed for evaluating the trade-off between harmonic suppression and reflection preservation. We compare five representative denoisers—DnCNN, U-Net, DURN, MIMO-U-Net, and the proposed DW-U-Net. Fig. 17(a)–(f) show the raw record and the corresponding denoised results, while Fig. 17(g)–(l) provide zoomed views of the dashed boxes in Fig. 17(a)–(f) (3000–4000 ms; near-offset traces).

From the global view, the raw record (Fig. 17(a)) contains strong near-source energy, together with coherent oblique banded patterns that are characteristic of harmonic-related streaks. After

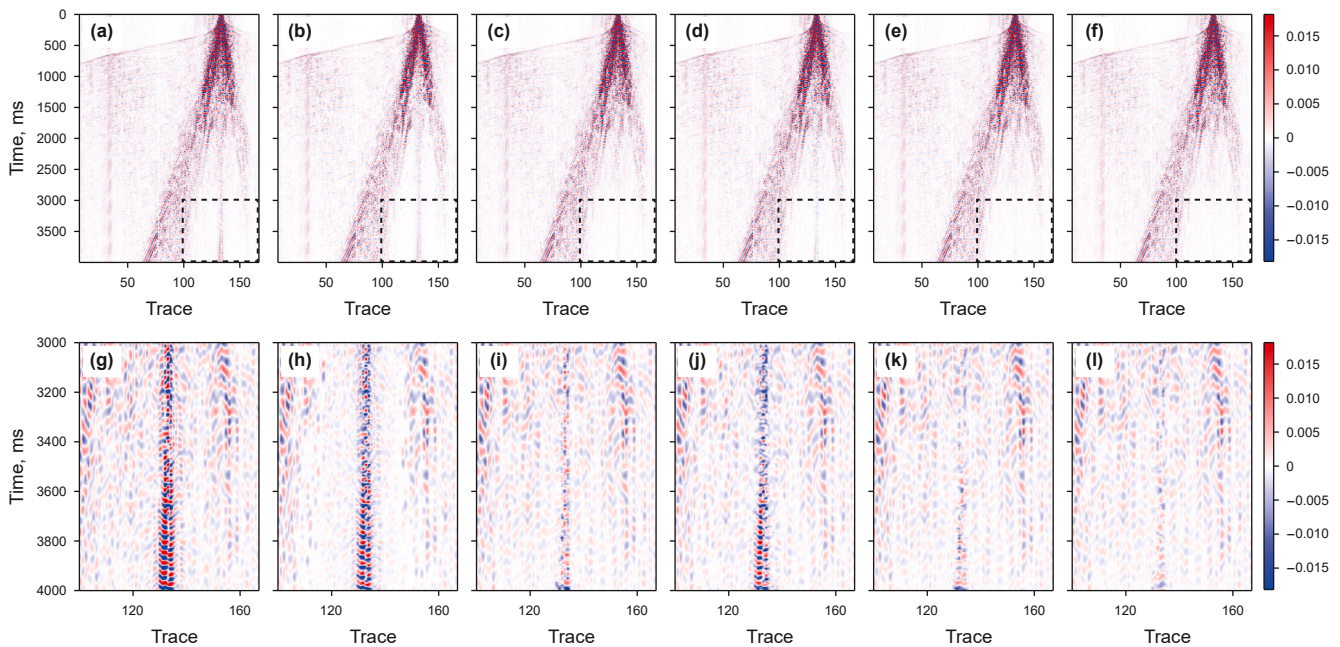


Fig. 17. Field data denoised comparison. (a) Raw single-shot record, (b) DnCNN, (c) U-Net, (d) DURN, (e) MIMO-U-Net, (f) DW-U-Net, (g) zoomed view of (a), (h) zoomed view of (b), (i) zoomed view of (c), (j) zoomed view of (d), (k) zoomed view of (e), (l) zoomed view of (f).

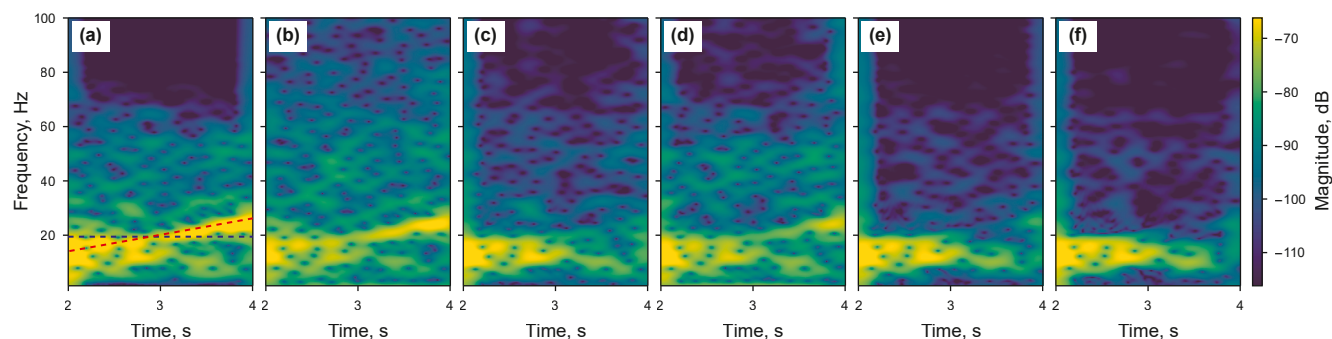


Fig. 18. Time-frequency spectra comparison. (a) Input, (b) DnCNN, (c) U-Net, (d) DURN, (e) MIMO-U-Net, (f) DW-U-Net.

DnCNN denoising (Fig. 17(b)), the background level is reduced and the strongest streaks are weakened, yet faint coherent remnants remain, especially in deeper time windows where weak reflections are easily contaminated. DURN (Fig. 17(d)) further reduces the coherent streaking while maintaining clearer event geometry than DnCNN, though residual banded textures are still visible. U-Net (Fig. 17(c)) suppresses the banded interference more strongly and improves the apparent structural clarity, but it also tends to smooth weak-amplitude regions, indicating a partial sacrifice of reflection energy and fine-scale detail. MIMO-U-Net (Fig. 17(e)) provides stronger suppression with improved continuity of dipping events, but mild smoothing and residual coherent artifacts persist in some regions. In contrast, DW-U-Net (Fig. 17(f)) produces the cleanest record overall, with markedly fewer residual streaks and better preservation of reflection continuity and phase consistency across traces, indicating a more favorable balance between suppression strength and signal fidelity.

The zoomed views in Fig. 17(g)–(l) offer more direct evidence in a challenging deep-time window dominated by weak reflections. The raw record (Fig. 17(g)) exhibits narrow, coherent ridges consistent with harmonic trends, which obscure subtle reflection details. DnCNN (Fig. 17(h)) reduces the ridge amplitude but leaves visible fine streaks. DURN (Fig. 17(j)) suppresses these remnants more aggressively, yet weak events appear slightly softened. U-Net (Fig. 17(i)) and MIMO-U-Net (Fig. 17(k)) further attenuate the coherent interference, but traces of residual banding remain. DW-U-Net (Fig. 17(l)) achieves the most complete suppression in this window while retaining sharper event boundaries and more continuous in-phase reflections, with minimal evidence of amplitude collapse or phase distortion. Overall, the field results demonstrate that DW-U-Net generalizes well to real acquisition conditions and offers a robust compromise between harmonic attenuation and reflection preservation.

Time-frequency evaluation on the Utah FORGE field record is shown in Fig. 18, where we plot the STFT magnitude of trace 135 over the 2.0–4.0 s window (time on the horizontal axis, frequency on the vertical axis, magnitude in dB). In the raw input (Fig. 18(a)), a persistent narrowband component around 20 Hz (blue line) and a slowly rising ridge in the low-frequency band (red line) indicate resonance- and harmonic-related interference. Meanwhile, the 30–100 Hz range contains dense speckled energy and scattered patches, reflecting out-of-band contamination that elevates the spectral background and disturbs the main reflection band. After DnCNN denoising (Fig. 18(b)), the overall background is reduced, but the low-frequency ridge remains evident and many isolated spectral “islands” persist at higher frequencies, especially between 40 and 80 Hz, indicating incomplete suppression. U-Net (Fig. 18(c)) and DURN (Fig. 18(d)) further attenuate the ridge-related energy and remove a larger portion of the high-frequency speckle; however, the dominant low-frequency band becomes slightly less concentrated in places, suggesting a modest trade-off between interference removal and passband energy preservation. MIMO-U-Net (Fig. 18(e)) performs stronger cleaning in the 40–100 Hz band, pushing much of the spurious energy closer to the noise floor; however, a mild ridge-like harmonic component can still be observed locally (as indicated by the arrows). In contrast, DW-U-Net (Fig. 18(f)) delivers the cleanest spectrum overall: the out-of-band speckled energy is largely eliminated, the ridge-like interference is least visible, and the main low-frequency reflection energy remains temporally continuous and smoothly distributed. Overall, DW-U-Net achieves a better compromise between harmonic suppression and continuity of the main band and balances noise reduction with signal fidelity.

Fig. 19 shows residuals for the five methods relative to the raw field record. Ideally, a residual should contain higher-order harmonics and other non-reflective components, with minimal

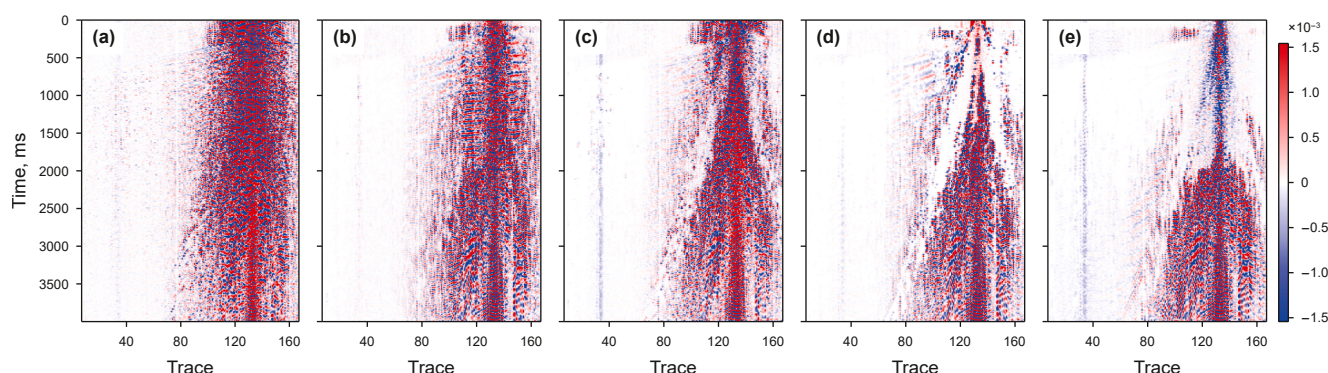


Fig. 19. Comparison of residuals. (a) DnCNN, (b) U-Net, (c) DURN, (d) MIMO-U-Net, (e) DW-U-Net.

coherent in-phase reflections. The DnCNN residual still exhibits continuous streaks and event-like textures in strong-energy zones and along the sweep travel-time trajectory, together with thin vertical bands and granular remnants. This indicates that useful reflections overlapping the harmonic bands are partly removed, causing over-suppression and phase shifts. The U-Net residual has a much lower peak amplitude and further weakens the banded interference, yet continuous reflection remnants remain. This suggests that stronger suppression comes with amplitude roll-off and slight bandwidth narrowing for weak reflections. The DURN residual shows reduced streak continuity compared with DnCNN and a cleaner background overall, but it still retains noticeable coherent remnants in the high-energy region, implying incomplete separation between harmonics and reflections. The MIMO-U-Net residual further concentrates the removed energy along the dominant interference patterns and suppresses part of the granular background, yet faint reflection-aligned textures can still be observed locally, indicating minor leakage. In contrast, the DW-U-Net residual is sparser and concentrated near the main harmonic ridges. Reflection-aligned residuals are effectively broken up, and the background level is the lowest. Overall, DW-U-Net best meets the objective: it leaves harmonics and other non-reflective energy in the residual while retaining useful reflections in the output.

4. Discussion

In this section, we provide an in-depth discussion of the proposed DW-U-Net. We conduct ablation studies to investigate key factors, including different network modules, wavelet bases, and loss-weight settings. We also analyze computational aspects such as training time, inference time, and memory consumption. Finally, we discuss the generalization ability of DW-U-Net based on results obtained with different validation-set ratios and additional tests on real field data.

During training, all experiments are performed on the same GPU platform, a GeForce RTX 5090D. To ensure a fair comparison, the training configuration is kept identical across models: batch size is set to 128, the number of epochs is 200, and the base channel width is 32. We use the Adam optimizer. For DW-U-Net, the T-U-Net and W-U-Net branches are initialized with separate learning rates, which are then updated using a cosine-annealing schedule during training. Baseline models follow the same cosine-annealing optimization protocol for fair comparison.

As the harmonic noise attenuation capability of DW-U-Net has been analyzed above, we now compare the five networks in terms of model size, training time, and inference time. Table 5 reports the end-to-end computational cost of the five denoising networks in terms of parameter count, GFLOPs, inference latency, and total training time under the same benchmarking setup. Lightweight CNN baselines, DnCNN and DURN, contain fewer than 0.2 M parameters (0.19 M and 0.13 M) and require only 1.51/1.05 GFLOPs, leading to fast inference (0.29/0.32 s/shot) and short training time (2185/1606 s). The standard U-Net increases model size to 2.71 M with 2.79 GFLOPs, and remains efficient at 0.34 s/shot, but its training time rises to 3961 s. Although MIMO-U-Net has a

Table 5
Computational cost and runtime comparison of different denoising networks.

Method	DnCNN	U-Net	DURN	MIMO-U-Net	DW-U-Net (ours)
Params, M	0.19	2.71	0.13	2.30	12.58
GFLOPs	1.51	2.79	1.05	6.43	12.57
Training time, s	2185.34	3960.87	1605.66	2360.55	7782.35
Inference time, s/shot	0.29	0.34	0.32	1.02	1.11

Table 6
Efficiency comparison of ContMix1D with MHSA and DWConv under different sequence lengths.

(B= 1, C= 64)	L	Params, M	GFLOPs	ms	MB
ContMix1D	1024	0.064	0.117	1.328	18.8
MHSA	1024	0.017	0.302	0.52	25.4
DWConv	1024	0.001	0.002	0.014	8.9
ContMix1D	2048	0.064	0.234	1.204	30.1
MHSA	2048	0.017	1.141	1.863	74.4
DWConv	2048	0.001	0.003	0.018	9.4
ContMix1D	4096	0.064	0.468	1.234	51.1
MHSA	4096	0.017	4.429	5.738	268.4
DWConv	4096	0.001	0.007	0.028	10.4
ContMix1D	8192	0.064	0.935	2.318	92.2
MHSA	8192	0.017	17.448	22.837	1040.4
DWConv	8192	0.001	0.014	0.054	12.4

comparable parameter count (2.30 M), its multiscale multi-output computation substantially increases the workload to 6.43 GFLOPs, resulting in a much higher inference time (1.02 s/shot). The proposed DW-U-Net is the most expensive model (12.58 M, 12.57 GFLOPs) and shows the longest training time (7782 s), which reflects the additional wavelet branch and multiscale fusion. Importantly, its inference time (1.11 s/shot) is close to MIMO-U-Net (1.02 s/shot), indicating that the added modeling capacity is achieved with a moderate inference-time overhead. Considering the consistent quality gains in both synthetic and field experiments, DW-U-Net provides a practical accuracy-efficiency trade-off for harmonic noise attenuation.

To quantitatively support the efficiency claim of ContMix1D over attention-based context modeling, Table 6 benchmarks ContMix1D against multi-head self-attention (MHSA) and a light-weight depthwise convolution (DWConv) under different sequence lengths (Batch (B) = 1, Channel (C) = 64). As the sequence length increases, MHSA exhibits a pronounced super-linear growth in both compute and memory (e.g., from 0.302 to 17.448 GFLOPs and from 25.4 to 1040.4 MB when length (L) increases from 1024 to 8192), consistent with the near-quadratic scaling of attention with respect to sequence length. In contrast, ContMix1D scales much more gently: its cost grows approximately linearly with L (0.117 → 0.935 GFLOPs and 18.8 → 92.2 MB over the same range). Notably, while MHSA can be faster at short sequences (0.52 ms vs. 1.328 ms at L= 1024), it becomes substantially slower as L grows; at L= 8192, ContMix1D is 18.7 × lower in GFLOPs (0.935 vs. 17.448), 11.3 × lower in memory (92.2 vs. 1040.4 MB), and 9.9 × faster in runtime (2.318 vs. 22.837 ms). DWConv provides a lower-bound cost reference (e.g., 0.014 GFLOPs and 12.4 MB at L= 8192) but lacks long-range context modeling capability. Overall, these results demonstrate that ContMix1D offers a scalable and hardware-feasible alternative to attention for long 1D seismic traces, mitigating the computational and memory bottlenecks of MHSA while retaining the ability to model long-range dependencies.

Table 7 reports a module-wise ablation study that quantifies the independent contributions of key components in DW-U-Net in terms of SNR and training time. Starting from the time-domain baseline T-U-Net (20.01 dB; 3730.09 s), adding ContMix1D yields

Table 7
Module ablation results of DW-U-Net.

Model	Training time, s	SNR
T-U-Net	3730.09	20.01
T-U-Net + ContMix1D	5858.75	20.93
T-U-Net + W-U-Net + ContMix1D	6489.50	22.24
T-U-Net + W-U-Net + MS-IF + ContMix1D	6518.73	26.23
T-U-Net + W-U-Net + MS-IF + ContMix1D + RM (DW-U-Net)	7782.35	29.29

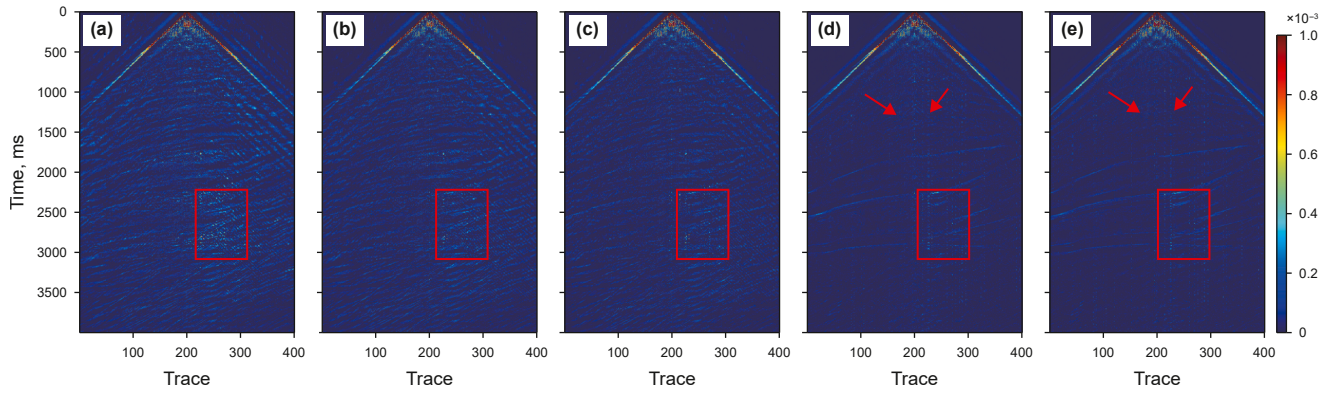


Fig. 20. Ablation study: prediction error maps for different module configurations. (a) T-U-Net, (b) T-U-Net + ContMix1D, (c) T-U-Net + W-U-Net + ContMix1D, (d) T-U-Net + W-U-Net + MS-IF + ContMix1D, (e) DW-U-Net.

a modest SNR improvement to 20.93 dB (+0.92 dB) but increases training time to 5858.75 s, indicating that context mixing helps long-range dependency modeling yet is insufficient alone for strong harmonic suppression. Introducing the wavelet branch on top of ContMix1D further improves SNR to 22.24 dB (+1.31 dB) with a relatively small additional training cost (6489.50 s), confirming the benefit of explicit multiscale decomposition beyond purely time-domain convolution. The most pronounced gain is achieved by enabling the MS-IF, which boosts SNR to 26.23 dB (+3.99 dB) while barely changing training time (6518.73 s), demonstrating that how wavelet-domain information is fused is critical and more effective than simple parallel branching. Finally, the complete DW-U-Net reaches the best SNR of 29.29 dB (+3.06 dB) at 7782.35 s, suggesting that the full configuration provides complementary benefits to maximize harmonic attenuation while preserving reflections. These trends are consistent with the ablation error maps in Fig. 20, where the mismatch progressively weakens and becomes more localized as modules are added, and interference-related structures are increasingly suppressed.

Table 8 presents a one-hot ablation on the loss-term weights, where only one supervision term is activated at a time. The weights ω_{ca} , ω_{cd} , and ω_{rec} correspond to supervision on the wavelet approximation coefficients, the wavelet detail coefficients, and the reconstruction-related term, respectively. The results show that supervising the approximation coefficients alone achieves the highest SNR (27.99 dB), outperforming detail-only supervision by 1.02 dB (26.97 dB) and reconstruction-only supervision by 2.97 dB (25.02 dB). This indicates that, for correlated vibroseis records, harmonic contamination is more effectively constrained through background guidance, which captures the coherent banded structures that overlap the reflection band after correlation. In contrast, relying solely on the reconstruction-related term is insufficient and tends to yield weaker suppression, suggesting that coefficient-level constraints are necessary to prevent residual harmonic leakage. In practice, these trends motivate assigning a dominant weight to ω_{ca} keeping a nonzero ω_{cd} to retain sharp boundaries and weak details, and using ω_{rec} as an auxiliary regularizer; the final weights can be selected by normalizing each loss term to a comparable numerical scale and tuning on the validation set.

Table 8
Ablation study on loss-term weights in DW-U-Net.

Loss-term weights	SNR
$\omega_{ca} = 1, \omega_{cd} = 0, \omega_{rec} = 0$	27.99
$\omega_{ca} = 0, \omega_{cd} = 1, \omega_{rec} = 0$	26.97
$\omega_{ca} = 0, \omega_{cd} = 0, \omega_{rec} = 1$	25.02

Fig. 20 visualizes the prediction error maps for the module ablation study. From panel (a) to (e), the mismatch magnitude progressively decreases and becomes increasingly localized, indicating a monotonic improvement in reconstruction fidelity as additional modules are introduced. The baseline T-U-Net in (a) exhibits widespread speckled errors and ridge-like structures, and the deep-time window highlighted by the red box contains noticeable event-aligned residuals, suggesting that weak primaries are not perfectly preserved and that harmonic-related interference is not fully removed. Adding ContMix1D (b) and further introducing the wavelet branch (c) reduces the overall error level, yet coherent textures and reflection-aligned remnants are still visible in the boxed region, implying residual amplitude deviations in low-SNR areas. After enabling MS-IF (d), the background error collapses toward the noise floor and the remaining mismatch is mainly concentrated along interference-like trajectories rather than continuous in-phase reflections. Finally, the complete DW-U-Net (e) further attenuates the remaining ridge-like errors highlighted by the arrows, showing that the final configuration provides additional suppression of harmonic remnants while reducing leakage of reflection energy into the error map. Overall, Fig. 20 confirms that the proposed modules act complementarily: the error becomes both smaller and less reflection-coherent, consistent with stronger harmonic attenuation and better preservation of useful reflections.

Fig. 21 compares prediction error maps obtained with different DWT wavelet basis types in the wavelet branch, and it shows that the wavelet choice affects the balance between harmonic suppression and signal preservation. All three bases keep the overall error low, but differences are visible in the highlighted regions. In the red box, which is dominated by harmonic-related structures near strong coherent ridges, the haar basis leaves weaker residual interference, indicating slightly stronger suppression of structured harmonics. In the arrow-marked area at deeper time, where weak reflections are more vulnerable, the sym8 basis produces a cleaner background with fewer reflection-aligned mismatches, suggesting better protection of subtle in-phase events. In this example, db4 exhibits a less favorable balance, with comparatively more scattered speckle-like errors in the highlighted zones. These observations confirm that the wavelet basis is not a minor detail. Short-support bases such as haar can favor more effective removal of coherent harmonic streaks, while smoother bases such as sym8 can better preserve weak reflections. In practice, the basis should be selected using validation shots that reflect the target field conditions and the desired priority between maximum suppression and maximum fidelity.

Fig. 22 compares the mean SNR of five denoising networks under different train-validation splits. A consistent trend is that DW-

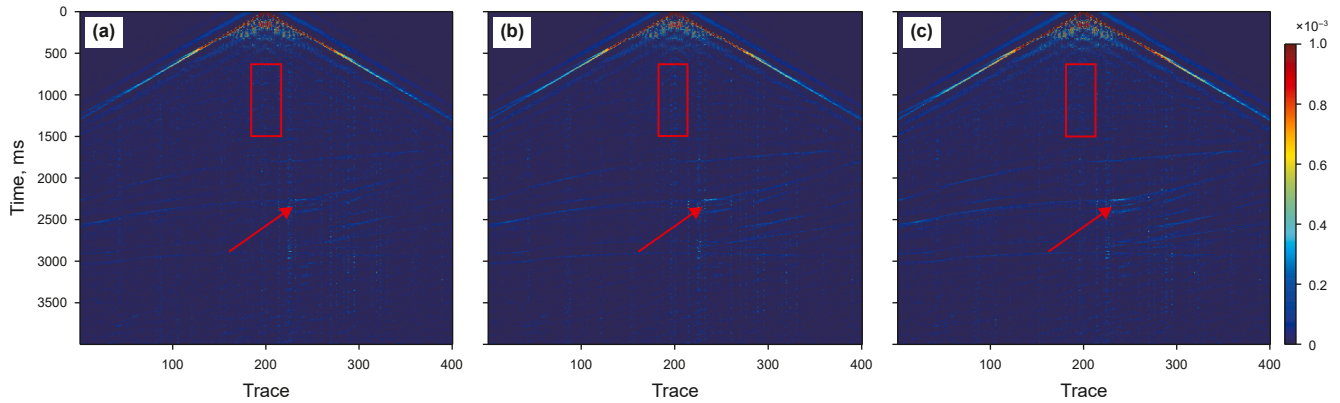


Fig. 21. Wavelet-basis ablation: prediction error maps for different DWT wavelet bases used in the wavelet branch. (a) sym8, (b) db4, (c) haar.

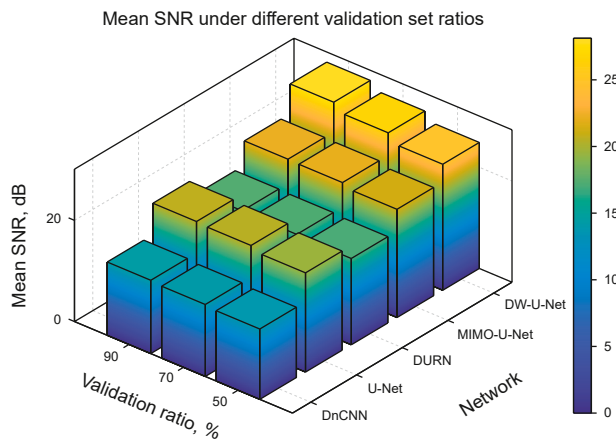


Fig. 22. Mean SNR of five denoising networks with different train-validation splits.

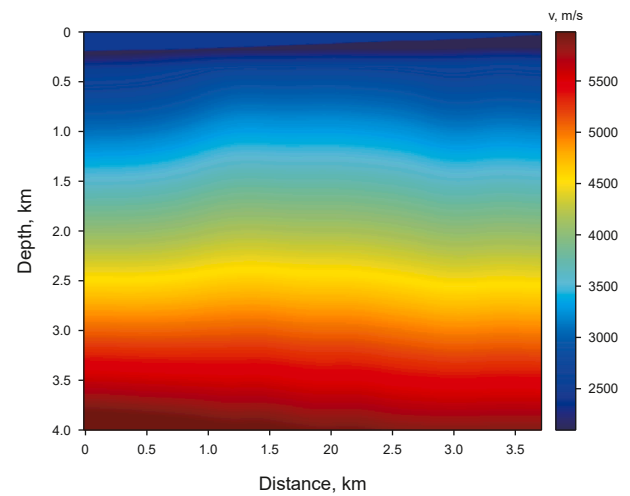


Fig. 23. Smoothed stacking velocity model.

U-Net achieves the highest mean SNR at every split, and its performance degrades only mildly when the training portion is reduced. Notably, even with the smallest training split of 50%, DW-U-Net still delivers a higher mean SNR than the other networks trained with 90% data, indicating stronger sample efficiency and more robust generalization. In contrast, the baseline CNN models and the multi-branch baselines show a clearer dependence on training-set size, with larger SNR drops as fewer training data are available. These results suggest that DW-U-Net's architecture provides a stronger inductive bias for separating coherent harmonic interference from reflections, which helps maintain denoising quality even when the amount of supervised training data is limited.

Before conducting the field generalization tests, we adopt a two-stage, survey-oriented training strategy to mitigate the domain gap between synthetic and field data. First, we perform large-scale synthetic pre-training and simulate diverse harmonic banding and narrowband resonance interference under different offset positions and reflection backgrounds, so that the network learns more transferable, structure-based features in complex reflection settings rather than memorizing a single synthetic appearance. Second, before applying the network to the target survey, survey-consistent paired samples are constructed from a limited amount of field data. As shown in Fig. 23, a smoothed stacking velocity model is first derived from the field data.

Based on this model, the harmonic-only response is simulated by using a pure harmonic signal as the source input. As shown in Fig. 24(a), several regions in the shot gathers that are relatively

unaffected by harmonic interference are selected as target samples. The corresponding corrupted inputs are then generated by superimposing the simulated harmonic response onto these relatively clean traces, as shown in Fig. 24(b). The resulting paired samples are then used in network training. In this stage, the learning rate is set to 5×10^{-4} , and the network is trained for 200 epochs. This adaptation reduces mismatches in amplitude scaling and phase characteristics and improves performance in the target field domain.

To further examine generalization beyond the training and validation shots, we apply the trained DW-U-Net to two additional field records that were not used during model development, as shown in Fig. 25. The raw shots in Fig. 25(a) and (c) exhibit severe coherent interference, characterized by strong stripe- and fan-like banding that extends over a wide time range and partially masks weak events at late times. After denoising, Fig. 25(b) and (d) show a clear reduction of these coherent harmonic-related patterns, with the background becoming noticeably cleaner and the residual banded energy largely suppressed. Importantly, the main signal structure is not collapsed: the dominant event geometry remains continuous across traces, and late-time weak features become easier to follow after interference attenuation. In the second shot, an isolated streak-like artifact visible in the raw record is also strongly reduced after processing, indicating that the network can handle localized abnormal interference in addition to the dominant harmonic banding. The consistent suppression behavior

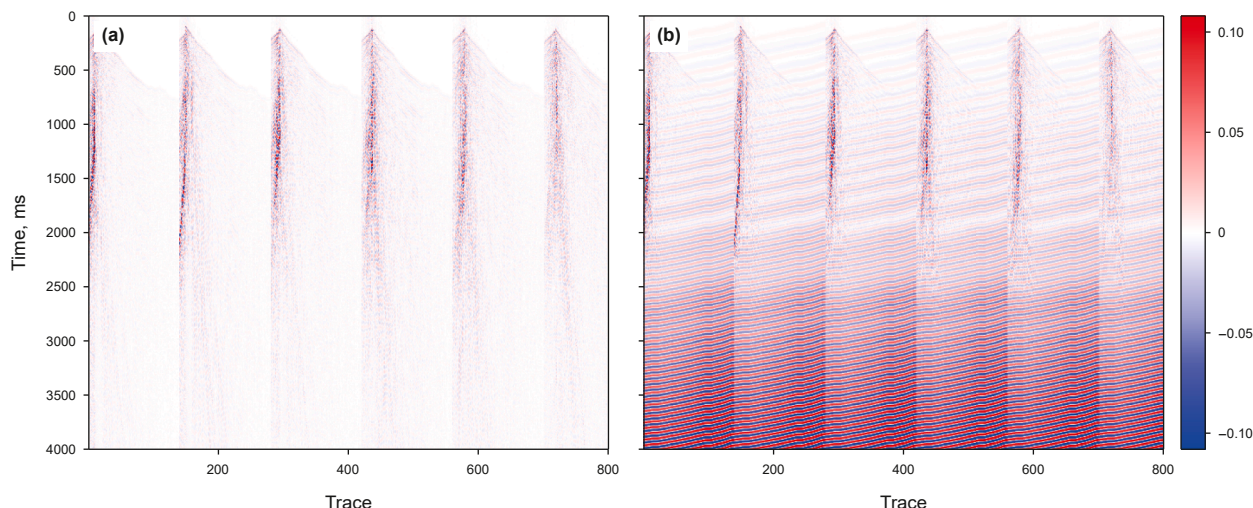


Fig. 24. Training data sets. (a) The field data of the unaffected areas without harmonic influence, (b) the field data of the harmonic interference simulated by superposition method.

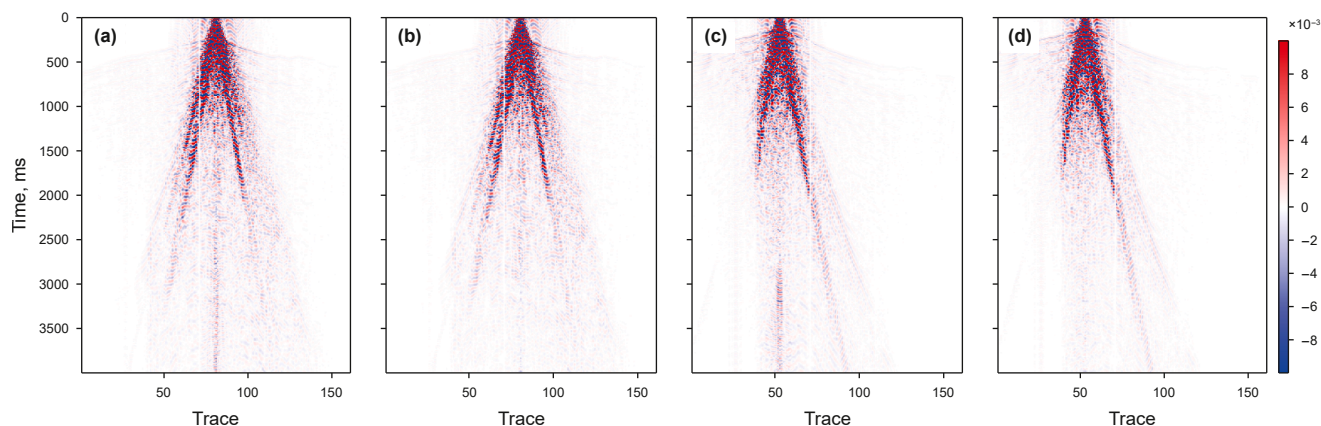


Fig. 25. Generalization evaluation of DW-U-Net on two additional field records. (a) Raw field shot 1, (b) DW-U-Net denoised result for shot 1, (c) raw field shot 2, (d) DW-U-Net denoised result for shot 2.

across two unseen field shots supports that DW-U-Net learns transferable, structure-based cues of vibroseis harmonic contamination rather than overfitting to a single-record, demonstrating practical generalization in complex field conditions.

5. Conclusions

We propose a DW-U-Net to attenuate vibroseis source-related harmonic interference in land seismic acquisition. DW-U-Net couples a time-domain U-Net backbone with a wavelet side branch that decomposes approximation and detail subbands, and injects reconstructed multiscale wavelet features into the decoder via a differentiable IDWT. A lightweight ContMix1D is further used to capture long-range dependencies in 1D traces without full self-attention, enabling trajectory-consistent suppression of harmonic streaks and narrowband resonance lines while preserving reflection continuity. On the Marmousi2 model, DW-U-Net achieves consistent improvements over DnCNN, U-Net, DURN, and MIMO-U-Net across multiple noise levels; under strong harmonic contamination, the whole-shot average SNR increases from 9.92 dB to 29.29 dB, with residual energy 2.44×10^{-3} , RMSE 4.51×10^{-5} , and MS-SSIM 0.9959. Qualitative comparisons on shot records, residual maps, and STFT spectra indicate that coherent

harmonic ridges and resonance components are largely removed, whereas the amplitude, bandwidth, and temporal continuity of the main reflection band are well preserved. Tests on the public Utah FORGE field data further confirm that the proposed model generalizes to real acquisition conditions and provides a favorable balance between suppression strength and signal fidelity.

This study still has several limitations. First, its performance depends on the representativeness of the supervised training data. Large variations in sweep parameters, source-ground coupling conditions, instrument resonances, and interference mechanisms may reduce denoising stability and lead to residual harmonic contamination or over-attenuation of weak reflections. Second, the current implementation is developed and applied at the single-trace level, which limits its ability to exploit spatial information. Future work will therefore focus on extending the current framework to multi-trace and 2D scenarios, so that richer spatial information can be incorporated and data diversity can be improved. Further efforts will also be directed toward improving robustness under varying acquisition conditions, enhancing spatial consistency, and increasing computational efficiency through model compression and optimized implementations. More extensive validation on additional field surveys and equipment configurations will also be carried out.

CRediT authorship contribution statement

Yang Liu: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Data curation, Conceptualization. **Xiao-Tao Wen:** Validation, Supervision, Resources, Funding acquisition. **Qing-Jun Yang:** Supervision, Funding acquisition. **Huan Zhu:** Software, Methodology. **Yun Zhao:** Funding acquisition, Formal analysis. **Theo Le Gallais:** Validation, Supervision, Formal analysis.

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

This study was financially supported by the Deep Earth Probe and Mineral Resources Exploration—National Science and Technology Major Project (Grant No. 2025ZD1004600), and Chengdu University of Technology Postgraduate Innovative Cultivation Program (Grant No. 2025BJCX-DF010). The authors who contributed to the writing and revision of this article are greatly appreciated. The authors thank the editorial board and the reviewers for their valuable and constructive suggestions.

References

- Abd El-Aal, A.E.-A.K., 2010. Eliminating upper harmonic noise in vibroseis data via numerical simulation. *Geophys. J. Int.* 181 (3), 1499–1509. <https://doi.org/10.1111/j.1365-246X.2010.04594.x>.
- Abd El-Aal, A.E.-A.K., 2011. Harmonic by harmonic removal technique for improving vibroseis data quality. *Geophys. Prospect.* 59 (2), 279–294. <https://doi.org/10.1111/j.1365-2478.2010.00918.x>.
- Aldridge, D.F., 1992. *Mathematics of linear sweeps*. *Can. J. Explor. Geophys.* 28 (1), 62–68.
- Birnie, C., Alkhalifah, T., 2022. Transfer learning for self-supervised, blind-spot seismic denoising. *Front. Earth Sci.* 10, 1053279. <https://doi.org/10.3389/feart.2022.1053279>.
- Birnie, C., Ravasi, M., Alkhalifah, T., 2021. Self-Supervised Learning for Random Noise Suppression in Seismic Data. SEG Technical Program Expanded Abstracts, pp. 2869–2873. <https://doi.org/10.1190/segam2021-3583248.1>.
- Butler, K.E., Russell, R.D., 2003. Cancellation of multiple harmonic noise series in geophysical records. *Geophysics* 68 (3), 1083–1090. <https://doi.org/10.1190/1.1581080>.
- Chen, X., Chen, H., Hu, Y., et al., 2024. Centroid-oriented extracting transform and its application in seismic spectral decomposition. *IEEE Trans. Geosci. Rem. Sens.* 62, 1–11. <https://doi.org/10.1109/tgrs.2024.3355250>.
- Cho, S., Ji, S., Hong, J., et al., 2021. Rethinking coarse-to-fine approach in single image deblurring. In: *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, pp. 4641–4650. <https://doi.org/10.1109/ICCV48922.2021.00460>.
- Chopra, A., 2020. In: *Dynamics of Structures in SI Units*, fifth ed. Pearson.
- Dodda, V.C., Kuruguntla, L., Mandpura, A.K., et al., 2023. Simultaneous seismic data denoising and reconstruction with attention-based wavelet-convolutional neural network. *IEEE Trans. Geosci. Rem. Sens.* 61, 1–14. <https://doi.org/10.1109/tgrs.2023.3267037>.
- Durall, R., Ghanim, A., Fernandez, M.R., et al., 2023. Deep diffusion models for seismic processing. *Comput. Geosci.* 177, 105377. <https://doi.org/10.1016/j.cageo.2023.105377>.
- Fang, P., Zhang, J., 2025. Estimating location and polarity of vibroseis reflections using multiscale phase-only correlation. *IEEE Geosci. Remote Sens. Lett.* 22, 1–5. <https://doi.org/10.1109/lgrs.2025.3563168>.
- Gao, J., Xing, X., Zheng, J., 2025. Optimization of the vibroseis to enhance signal-to-noise ratio for seismic exploration. *IEEE Trans. Instrum. Meas.* 74, 1–13. <https://doi.org/10.1109/tim.2025.3533625>.
- Harsuko, R., Alkhalifah, T., 2022. StorSeismic: a new paradigm in deep learning for seismic processing. *IEEE Trans. Geosci. Rem. Sens.* 60, 1–15. <https://doi.org/10.1109/tgrs.2022.3216660>.
- Hu, H., Wei, Y., Chen, H., et al., 2025. Seismic data denoising based on attention dilated CNN. *Sci. Rep.* 15 (1), 28123. <https://doi.org/10.1038/s41598-025-13857-9>.
- Kong, X., Chen, H., Wang, J., et al., 2017. An amplitude suppression method based on the decibel criterion. *Appl. Geophys.* 14 (3), 387–398. <https://doi.org/10.1007/s11770-017-0636-7>.
- Krull, A., Buchholz, T.-O., Jug, F., 2019. Noise2Void: learning denoising from single noisy images. In: *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. CVPR*, pp. 2129–2137. <https://doi.org/10.1109/CVPR.2019.00223>.
- Kuruguntla, L., Dodda, V.C., Elumalai, K., 2022. Study of parameters in dictionary learning method for seismic denoising. *IEEE Trans. Geosci. Rem. Sens.* 60, 1–13. <https://doi.org/10.1109/tgrs.2021.3107541>.
- Li, B., Wang, Y., Yuan, H., et al., 2022. Double-phase-shift filtering method for harmonic elimination based on AR2U-Net. *Appl. Geophys.* 19 (2), 271–283. <https://doi.org/10.1007/s11770-022-0932-8>.
- Li, F., Liu, H., Wang, W., et al., 2024. Swin transformer for seismic denoising. *IEEE Geosci. Remote Sens. Lett.* 21, 1–5. <https://doi.org/10.1109/lgrs.2024.3358234>.
- Li, X., 1997. Decomposition of vibroseis data by the multiple filter technique. *Geophysics* 62 (3), 980–991. <https://doi.org/10.1190/1.1444204>.
- Lin, Y., Zhang, J., 2022. A multispectral denoising framework for seismic random noise attenuation. *IEEE Trans. Geosci. Rem. Sens.* 60, 1–17. <https://doi.org/10.1109/tgrs.2020.3047633>.
- Liu, D., Wang, W., Wang, X., et al., 2020. Poststack seismic data denoising based on 3-D convolutional neural network. *IEEE Trans. Geosci. Rem. Sens.* 58 (3), 1598–1629. <https://doi.org/10.1109/tgrs.2019.2947149>.
- Liu, S., Birnie, C., Alkhalifah, T., 2022. Coherent noise suppression via a self-supervised deep learning scheme. In: *83rd EAGE Annual Conference & Exhibition*, pp. 1–5. <https://doi.org/10.3997/2214-4609.202210382>.
- Liu, S., Birnie, C., Alkhalifah, T., 2023. Trace-wise coherent noise suppression via a self-supervised blind-trace deep-learning scheme. *Geophysics* 88 (2), V101–V113. <https://doi.org/10.1190/geo2022-0371.1>.
- Liu, X., Suganuma, M., Sun, Z., et al., 2019. Dual residual networks leveraging the potential of paired operations for image restoration. In: *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. CVPR*, pp. 7007–7016. <https://doi.org/10.1109/CVPR.2019.00717>.
- Lou, M., Yu, Y., 2025. OverLoCK: an overview-first-look-closely-next ConvNet with context-mixing dynamic kernels. In: *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. CVPR*, pp. 128–138. <https://doi.org/10.1109/CVPR52734.2025.00021>.
- Miller, J., 2018. Utah FORCE: 2D and 3D Seismic Data. DOE Geothermal Data Repository. Energy and Geoscience Institute, University of Utah. <https://doi.org/10.15121/1452746>.
- Qu, S., Zhou, H., Chen, Y., et al., 2015. An effective method for reducing harmonic distortion in correlated vibroseis data. *J. Appl. Geophys.* 115, 120–128. <https://doi.org/10.1016/j.jappgeo.2015.02.002>.
- Sheng, H., Wu, X., Si, X., et al., 2025. Seismic foundation model (SFM): a next-generation deep-learning model in geophysics. *Geophysics* 90 (2), IM59–IM79. <https://doi.org/10.1190/geo2024-0262.1>.
- Si, C., Zhao, B., Huang, L., et al., 2023. A convolutional de-quantization network for harmonics suppression in one-bit SAR imaging. *IEEE Trans. Geosci. Rem. Sens.* 61, 1–16. <https://doi.org/10.1109/tgrs.2023.3330530>.
- Sun, Q., Feng, Y., Du, Q., et al., 2025. Removing random noise and improving the resolution of seismic data using deep-learning transformers. *Geophys. Prospect.* 73 (2), 611–627. <https://doi.org/10.1111/1365-2478.13617>.
- Vedanti, N., Gupta, L., Singh, V., et al., 2021. Problems with the standard vibroseis deconvolution: some practical solutions. *Explor. Geophys.* 52 (3), 308–320. <https://doi.org/10.1080/08123985.2020.1825913>.
- Walden, A.T., White, R.E., 1998. Seismic wavelet estimation: a frequency-domain solution to a geophysical noisy input-output problem. *IEEE Trans. Geosci. Rem. Sens.* 36 (1), 287–297. <https://doi.org/10.1109/36.655337>.
- Wei, Z., Phillips, T.F., 2013. Enhancing the low-frequency amplitude of ground force from a seismic vibrator through reduction of harmonic distortion. *Geophysics* 78 (4), P9–P17. <https://doi.org/10.1190/geo2012-0500.1>.
- Wu, M., Wu, J., Bai, M., et al., 2026. U-STRNet: a unified model integrating swin transformer and residual dense network for seismic image super-resolution and denoising. *J. Seismic Explor.* 35 (1), 025400081. <https://doi.org/10.36922/jse025400081>.
- Yuan, Y., Li, Y., Li, G., et al., 2022. Adaptive ground-roll attenuation using local nonlinear filtering. *Comput. Geosci.* 164, 105124. <https://doi.org/10.1016/j.cageo.2022.105124>.
- Zhang, H., Yang, J., Huang, J., et al., 2024. An adaptive time-frequency denoising method for suppressing source-related seismic strong noise. *IEEE Trans. Geosci. Rem. Sens.* 62, 1–10. <https://doi.org/10.1109/tgrs.2024.3369877>.
- Zhang, H., Zhou, H., Abd El-Aal, A.E.-A.K., et al., 2012. The anti-correlation method for removing harmonic distortion in vibroseis slip-sweep data. *Appl. Geophys.* 9 (2), 159–167. <https://doi.org/10.1007/s11770-012-0325-5>.
- Zhang, M., Hu, T., Duan, W., et al., 2025. Deep learning for efficient prestack seismic internal multiples suppression with global-local feature fusion and Optuna optimization. *J. Geophys. Eng.* gxf142. <https://doi.org/10.1093/jge/gxf142>.
- Zhang, Q., Mao, W., Fang, J., 2020. Elastic full waveform inversion with source-independent crosstalk-free source-encoding algorithm. *IEEE Trans. Geosci. Rem. Sens.* 58 (4), 2915–2927. <https://doi.org/10.1109/tgrs.2019.2957829>.
- Zhu, W., Mousavi, S.M., Beroza, G.C., 2019. Seismic signal denoising and decomposition using deep neural networks. *IEEE Trans. Geosci. Rem. Sens.* 57 (11), 9476–9488. <https://doi.org/10.1109/tgrs.2019.2926772>.