ACCSCIENCE PUBLISHING

Journal of Seismic Exploratio

ARTICLE

An improved CNN-Transformer network for seismic denoising with local–global feature modeling

Zhaoyang Ge[1], Jingrui Luo[1]*, Guoxin Chen[2], and Naijian Wang[3]

[1]School of Automation and Information Engineering, Xi'an University of Technology, Xi'an, Shaanxi, China

[2]Ocean College, Zhejiang University, Zhoushan, Zhejiang, China

[3]BGP, Inc., China National Petroleum Corporation, Zhuozhou, Hebei, China

Abstract

In seismic exploration, field data are always affected by noise, reducing the quality of subsequent analysis. Striking a balance between signal preservation and noise attenuation effectively is a major challenge. This study adopts a dual-branch framework with convolutional neural networks (CNN) and Transformer architecture that integrates blind-spot convolution (BS-Conv) for local detail capture, rank-augmented linear attention (RALA) for global feature extraction, and the gated CNN block (GCB) for adaptive feature selection to achieve high-fidelity signal reconstruction and noise suppression. Specifically, the CNN branch introduces BS-Conv to strengthen the local feature extraction of seismic data. The Transformer branch adopts RALA to capture long-range dependencies, strengthening global feature extraction. Finally, each skip connection is equipped with GCB to adaptively extract relevant features. Both branches are constructed on the U-Net++ architecture and utilize its multi-scale feature-fusion capability to optimize denoising performance. The network was tested on synthetic data and field seismic data to validate the model performance. We quantitatively evaluated the denoising results using peak signal-to-noise ratio and structural similarity index. Comparative tests prove that this dual-branch denoising framework is superior to other methods in both noise reduction and signal fidelity. These results further confirm its effectiveness and robustness for seismic denoising on both synthetic and field data.

Keywords: Deep learning; Seismic denoising; Blind-spot convolution; Rank-augmented linear attention; Gated convolutional neural network block

***Corresponding author:**
Jingrui Luo
(bluebirdjr@126.com)

Citation: Ge Z, Luo J, Chen G, Wang N. An improved CNN-Transformer network for seismic denoising with local–global feature modeling. *J Seismic Explor.* doi: 10.36922/JSE026270127

Received: July 5, 2026

Revised: August 10, 2026

Accepted: August 11, 2026

Published online: October 26, 2026

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

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

1. Introduction

Seismic data acquired in the field are heavily contaminated by noise from natural phenomena such as ocean waves, wind, traffic, electrical noise, and instrument noise.[1,2] These noises greatly affect the accuracy and effectiveness of subsequent seismic processing, inversion, and migration. In seismic exploration, as a result, how to reduce unwanted noise while preserving the desired signal is an important and challenging topic.[3-7]

To obtain high-quality seismic signals, researchers have proposed various denoising approaches. Among these, predictive filtering methods, including *t-x* predictive

filtering[8,9] and *f-x* deconvolution,[10,11] enhance signal quality by exploiting spatial-temporal correlations. Nonstationary filtering[12,13] adjusts itself in response to how signal features vary. Bilateral filtering (BF) retains structural edges by using structure-oriented smoothing.[14] However, their reliance on predefined models and parameters often limits the performance on field data. Common sparse transforms such as the short-time Fourier transform,[15] curvelet transform,[16,17] random transform,[18] and wavelet transform (WT)[19,20] have been widely applied. Noise suppression based on these methods is generally implemented by setting a threshold in the transform domain using coefficient differences to separate noise from the effective signal. Therefore, the success of this method lies in a suitable threshold that can accurately separate noise from the effective signal,[21] which is usually a challenging problem. The low-rank decomposition techniques, such as Cadzow filtering,[22,23] multichannel singular spectrum analysis,[24] singular spectrum analysis,[24,25] and robust principal component analysis,[26,27] exploit the property that clean seismic data possess a low-rank trait that is disrupted by noise, and thus implement denoising via matrix rank reduction. In addition, adaptive decomposition methods such as empirical mode decomposition[28,29] and variational mode decomposition[30] offer advantages for nonstationary signals without predefined basis functions. However, these methods are often computationally expensive and sensitive to parameter settings. In general, these traditional methods rely heavily on manual prior knowledge and extensive parameter tuning, limiting their applicability to automated and robust seismic data processing.[31,32]The advancement of deep learning has provided fresh vitality to seismic data denoising.[33-35] Many classic architectures have been widely applied in this field, such as generative adversarial networks (GANs),[36] U-Net,[37] and denoising convolutional neural networks (DnCNNs).[38] Following these foundational works, researchers began adapting these architectures specifically for seismic challenges. Xie *et al.*[39] applied GANs to onshore seismic data. Zhao *et al.*[40] adjusted the hyperparameters of DnCNN to better conform to the traits of low-frequency desert noise. Luo *et al.*[41] introduced residual blocks into the U-Net model to handle random noise, improving the generalization and signal-to-noise ratio (SNR). Subsequently, Yuan *et al.*[42] modified the network structure of GANs according to the characteristics of ground-roll, improving the SNR of seismic data. For low-frequency desert data, Wang *et al.*[43] adjusted the loss function, enabling the modified GANs to achieve higher denoising quality. At the same time, Yao *et al.*[44] integrated skip connections, residual learning, and group convolution into the DnCNN architecture, enhancing the denoising capability of the DnCNN for desert seismic data. Zhong *et al.*[45] modified the down-sampling and up-sampling layers to increase the denoising accuracy for seismic random noise. Dodda *et al.*[46] combined the WT with U-Net and achieved better effectiveness in synthetic data and in field data. Wang *et al.*[47] proposed a dual-branch dense network (DBD-Net) that employs multi-scale feature extraction branches within a single CNN framework, achieving superior denoising performance. In addition, lightweight Transformers such as Swin Transformer[48] and linear attention mechanisms[49] have brought new vitality to the field of deep learning. Wang *et al.*[50] combined the multi-head attention of Transformers with channel attention. This design exploits long-range features in seismic records, enhancing denoising performance and adaptability to complex records. Li *et al.*[51] introduced the Swin Transformer into seismic denoising and verified its effectiveness in random noise interference. Chen *et al.*[52] proposed the multi-scale channel attention Swin-Conv U-Net (MCA-SCU), which integrates convolution and the Swin Transformer into a U-Net framework. With testing on synthetic seismic data and field data, their approach achieved superior denoising capability. Li *et al.*[53] developed an SNR-aware dual-branch denoising network (SNR-ADDNet) that dynamically switches between CNN and Transformer branches based on SNR levels, achieving effective feature extraction tailored to regional signal quality. In addition, Mamba-based architectures such as dual-domain Mamba (DDMamba)[54] and DnMamba[55] have been introduced for seismic random noise suppression, demonstrating promising performance.

However, recent studies have indicated that combining convolutional neural networks (CNNs) and Transformers may lead to optimization conflicts between local inductive biases and global statistical modeling.[56] To avoid mutual interference between the two mechanisms, we design a more independent dual-branch denoising network that consists of a CNN branch with blind-spot convolution (BS-Conv),[57] a Transformer branch with rank-augmented linear attention (RALA)[58] (which enhances the key-value [KV] buffer and output feature via dynamic coefficients and Hadamard product), and skip connections with a gated CNN block (GCB).[59] Specifically, the introduction of BS-Conv enables the CNN branch to focus on extracting local details from seismic data. The introduction of RALA allows the Transformer branch to capture long-range information in seismic events. Furthermore, the GCB ensures that the skip connections do not passively receive all features but instead adaptively select features relevant to the seismic signals. This parallel dual-branch design avoids mutual interference between the CNN and Transformer via feature decoupling, and both branches are built on U-Net++[60] to exploit its multi-scale feature fusion capability for optimizing denoising performance.

The main objective of this work is to attenuate random noise in field data. We first validate the elementary denoising capability of the method with Gaussian random noise. Next, we construct a synthetic dataset based on field noise to test the model's effectiveness. Finally, we use field data to validate the model's generalization. The results show that our network achieves better noise suppression while preserving signal details.

2. Methods and architecture

Figure 1 presents the complete network architecture. The network consists of two branches, i.e., a CNN branch and a Transformer branch, both based on U-Net++. The CNN branch comprises GCB and hybrid convolutional blocks. These blocks integrate BS-Conv, which enhances the capacity of this branch to capture local details. Meanwhile, the GCB can adaptively select signal-related information and mitigate harmful noise interference in subsequent layers. The Transformer branch stacks multiple Transformer blocks and GCBs. The Transformer blocks integrate the RALA module, enabling this branch to extract extensive structural features and long-range spatial correlations. Within this design, the input data are fed separately into each branch. The final output is obtained by fusing the local details from the CNN branch with global structural information from the Transformer branch. Through the synergy of these two independent branches, the architecture can suppress noise in seismic data with higher effectiveness.

2.1. Local feature extraction via the CNN branch

The CNN branch is mainly used to obtain the local feature structure of effective seismic signals. As shown in Figure 1, this branch comprises three primary components: an encoder stage, a decoder stage, and skip connections. The encoder stage includes four down-sampling layers with corresponding hybrid convolutional blocks. The decoder path consists of 10 up-sampling layers with corresponding hybrid convolutional blocks. In addition, the skip connections are no longer simple direct links, and the GCB is incorporated. In this branch, the input data are first fed into the encoder to produce feature maps at different scales. Then, these characteristics travel through the skip connections to the decoder. Each node in the decoder integrates the up-sampling output from the lower-level node and feature maps from all preceding nodes along the same horizontal skip pathway. This design allows the model to use multi-scale information more effectively, improving denoising accuracy.

Within this branch, we replace the standard units with the hybrid convolutional block, as shown in Figure 2. This improved block places a BS-Conv before a regular convolution. Subsequent to each convolutional layer are batch normalization (BN) and a rectified linear unit (ReLU) activation function. Specifically, the input feature data first pass through a BS-Conv, where the weight at the center of the convolution kernel is forcibly set to zero. As a result, each output pixel is computed solely from the surrounding neighboring pixels. This design compels the network to infer the response at the central point from adjacent seismic traces or from the neighborhood of reflection events. Consequently, the network becomes more sensitive to weak seismic signals and local discontinuous structures. Subsequently, a convolution layer follows immediately. This layer further integrates spatial information based on the neighborhood features extracted by the BS-Conv. It also compensates for potential feature loss caused by the central blind spot. Finally, after convolution, the resulting features pass through BN and a ReLU activation function, which are utilized to ensure training stability and nonlinear expressive capability.

The structure of this hybrid convolutional block ensures that it extracts local information from the neighborhood before the standard spatial transformation, enabling the branch to effectively improve the continuity of events and enhance its ability to capture local features of seismic signals.

2.2. Transformer branch for global feature extraction

The above CNN branch can effectively extract local features. However, effective reflection signals in seismic data often exhibit strong cross-regional correlations along both the time and trace directions. Particularly in the presence of complex background noise, relying solely on local convolutions is insufficient for fully restoring the continuity and overall structural characteristics of reflection events. In other words, a single CNN branch performs at an acceptable level in low-amplitude noise conditions, but it becomes insufficient when the noise intensity becomes high. To overcome this deficiency, the Transformer branch is introduced in parallel to the CNN branch. It is used to obtain long-range spatial correlations and global structural information within effective signals.

As shown in Figure 1, this branch shares the same U-Net++ backbone as the CNN branch. Therefore, we focus only on its core computational units here. Figure 3 illustrates the Transformer block, which serves as the fundamental unit of the branch. Each block comprises a RALA and a feed-forward network (FFN). Layer normalization is applied before the RALA and the FFN to ensure that the distribution of hidden states remains stable across deep layers. Specifically, the data first enter

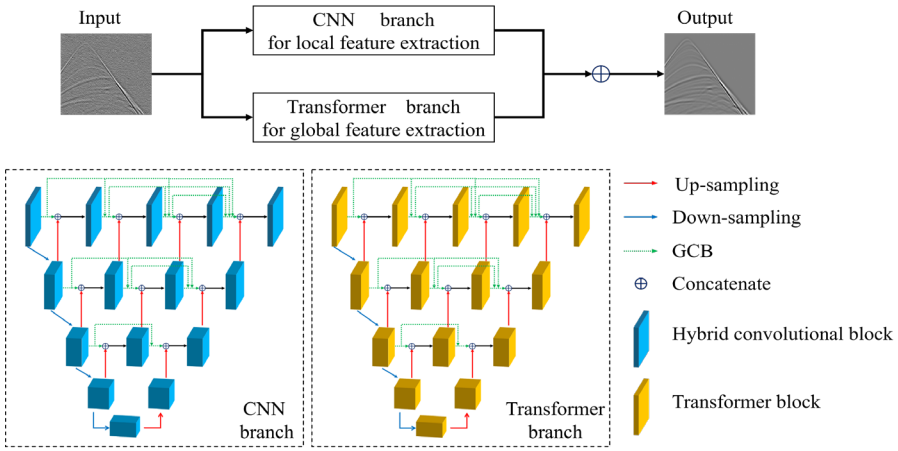


Figure 1. The complete network architecture. The two dashed boxes indicate the CNN branch (left) and the Transformer branch (right).
Abbreviations: CNN: Convolutional neural network; GCB: Gated CNN block.

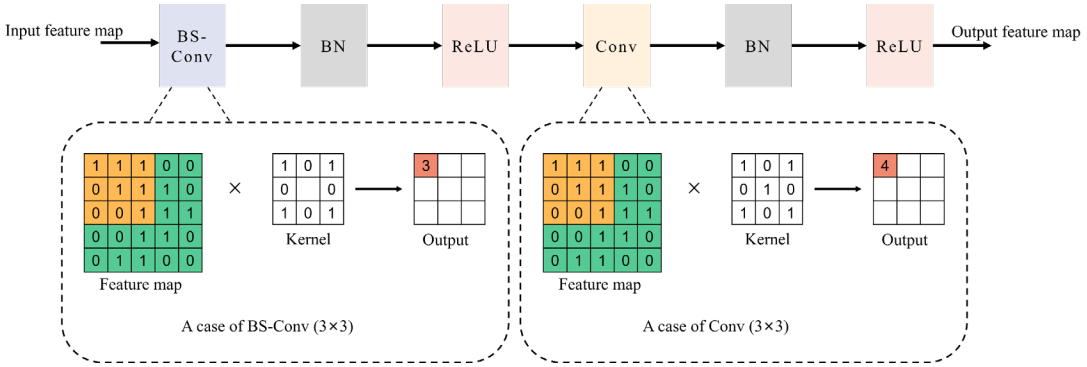


Figure 2. Structure of the hybrid convolutional block. The two dashed boxes from left to right illustrate the operation of BS-Conv (with the central weight excluded from the computation) and the subsequent 3 × 3 regular convolution.
Abbreviations: BS-Conv: Blind-spot convolution; BN: Batch normalization; Conv: Regular convolution; ReLU: Rectified linear unit.

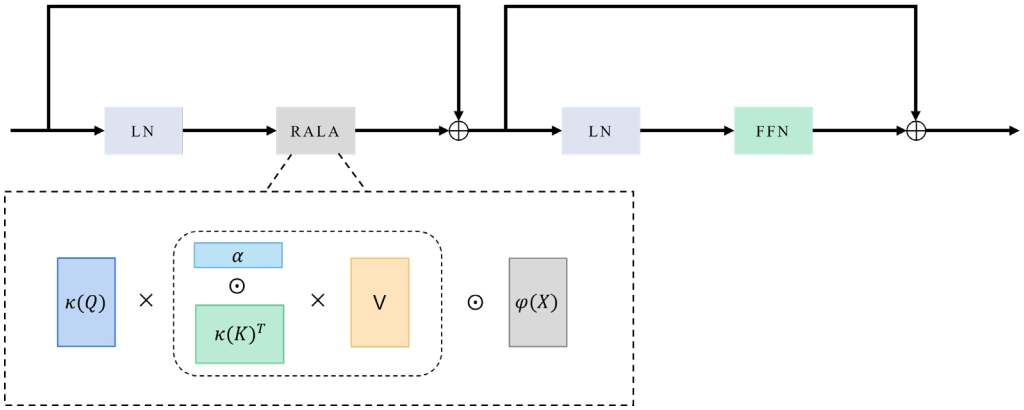


Figure 3. Structure of the Transformer block. The outer dashed box delineates the overall architecture of the RALA module, and the inner dashed box indicates the composition of the KV buffer.
Abbreviations: FFN: Feed-forward network; KV: Key-value; LN: Layer normalization; RALA: Rank-augmented linear attention.

the RALA module to capture long-range dependencies across the seismic shot gather with linear computational complexity. Subsequently, the FFN performs point-wise feature transformation and channel-wise information interaction to strengthen the nonlinear representation of the extracted features. Each block also incorporates residual connections. These connections facilitate the direct flow of original signal components through the network, minimizing loss of structural details and preventing the generation of long-range artifacts. Through the integration of RALA and FFN, the Transformer branch provides a robust global constraint for the reconstruction of seismic shot gathers, thus ensuring high fidelity and structural continuity in the denoised results.

The core component of this block is the RALA. This decision is due to the fact that RALA has a global receptive field and produces output features with higher diversity than traditional linear attention. In standard linear attention, the introduction of kernel functions to approximate the Softmax operation leads to a low-rank characteristic of the output feature map matrix, resulting in a lack of feature diversity and an inability to model complex spatial information. For traditional linear attention, given an input token $X \in \mathrm{R}^{N\times d}$, output Y_i can be expressed as:

$$Y_i = \frac{\kappa(Q_i)\left(\sum_{j=1}^{N}\kappa\left(K_j\right)^T V_j\right)}{\kappa(Q_i)\left(\sum_{m=1}^{N}\kappa\left(K_m\right)^T\right)} \tag{1}$$

where $Q(\cdot)$, $K(\cdot)$, $V(\cdot)$ is a linear transformation of X, and $\kappa(\cdot)$ is the kernel function equals to $Elu(\cdot) + 1$. To facilitate the analysis of rank properties, we denote the sum of $\kappa\left(K_j\right)^T V_j$ as the KV buffer B:

$$B = \sum_{j=1}^{N}\kappa\left(K_j\right)^T V_j, B \in \mathbb{R}^{d\times \mathrm{d}} \tag{2}$$

Therefore, the rank of the resulting feature map satisfies:

$$rank\left(Y\right) \le min\left(rank\left(\kappa\left(Q\right)\right), rank\left(B\right)\right) \le d \tag{3}$$

Due to the introduction of the kernel function, the rank of the resulting feature map is theoretically limited to within d. However, it is far less than d in practice, and this situation has caused the low-rank dilemma.

As illustrated in the dashed box in Figure 3 (where spatial indices i, j are omitted for visual clarity), RALA enhances the rank of the feature matrix through two aspects. First, it improves the KV buffer by introducing an input-dependent dynamic coefficient α_j (depicted in light blue), which increases the rank of B. Second, it performs an element-wise Hadamard product $\odot$ between the output result and the linear projection of the original tokens $\varphi(X_i)$ (depicted in gray). This Hadamard product operation increases the rank of the final output feature matrix and prevents detail loss. With the dynamic coefficient and the Hadamard product operations, the rank of the feature map can theoretically reach d (full rank). The improvement and output of RALA are expressed as follows:

$$B_{RALA} = \sum_{j=1}^{N}\alpha_j\kappa\left(K_j\right)^T V_j \tag{4}$$

$$Y_i = \varphi\left(X_i\right)\odot\left(\kappa\left(Q_i\right)B_{RALA}\right) \tag{5}$$

The above operations enable RALA to have a global receptive field and to provide output features with higher diversity. Thus, this branch can effectively separate weak seismic signals from strong ambient noise. To fully leverage the complementary strengths of the two branches, the input seismic data are processed through the CNN and Transformer branches in parallel. The outputs of the two branches, denoted as Y_{CNN} and Y_{trans}, are then fused via channel-wise concatenation, followed by a 3 × 3 convolutional layer that reweights the fused features and compresses the channel dimension. This fusion process can be expressed as follows:

$$Y = Conv(Concat(Y_{CNN}, Y_{trans})) \tag{6}$$

The cooperative improvement of the two branches permits the network to extract local and global contexts simultaneously. Thus, the model improves denoising performance in seismic data.

2.3. The gated CNN block operation

The conventional U-Net++ performs direct feature concatenation along the channel dimension at its skip connections. As a result, the decoder receives all feature information from each encoder layer. This leads to two problems. First, noise features from a previous layer propagate to subsequent layers and interfere with signal reconstruction. Second, numerous skip connections cause the same features to be transmitted repeatedly, resulting in parameter redundancy and wasted computational resources. Therefore, we introduce a GCB at the skip connections, a lightweight variant of the Mamba architecture that removes the selective state space model for computational efficiency while retaining input-dependent gating to adaptively select useful features and suppress noise transmission.

Figure 4 illustrates the structure of the GCB. The input feature map first passes through a linear layer (Linear) to expand its channel dimension to twice the original size. One part uses a Gaussian error linear unit (GELU) activation function to generate the gate coefficient, while the other part uses a 7 × 7 separable convolution to extract spatial

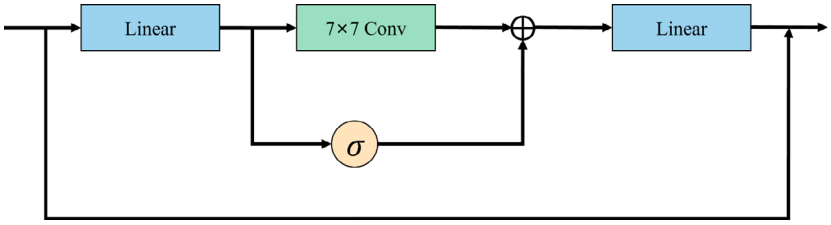


Figure 4. Structure of the GCB
Abbreviations: 7 × 7 Conv: 7 × 7 separable convolution; GCB: Gated conventional neural network block; Linear: Linear layer.

features and ensure large-scale spatial continuity of signals. The activated gate coefficient then multiplies element-wise with the feature map generated by the convolution. Then, the result passes through a second linear layer to project the channel dimension back to the size before processing. Finally, a residual connection adds the original input to the projected result. This allows the module to preserve the original information while preventing gradient vanishing. Through this mechanism, the network can adaptively select useful signal information.

2.4. Loss functions and evaluation indicators

We use the classical mean squared error (MSE) for network training. This loss function computes the squared deviation of the denoised data from the clean data on a pixel-by-pixel basis. Through MSE, the model can effectively learn to separate noise from the effective signal in the seismic data. MSE is expressed as:

$$MSE\left(x,\hat{x}\right)=\frac{1}{N}\sum_{i=1}^{N}\left(x_i-\hat{x}_i\right)^2 \tag{7}$$

where $x,\hat{x}$ represents the clean input data and the denoised output data, and N is the total number of pixels.

We adopt the structural similarity index (SSIM) and the peak signal-to-noise ratio (PSNR) as evaluation indicators to validate the denoising quality of the network. PSNR is based on MSE. A higher PSNR indicates more accurate signal restoration. SSIM is a commonly adopted evaluation indicator that measures the degree of structural similarity between data. A higher SSIM indicates better preservation of the structural details in seismic signals. Their definitions are as follows:

$$SSIM\left(x,\hat{x}\right)=\frac{\left(2\mu_x\mu_{\hat{x}}+C_1\right)\left(2\sigma_{x\hat{x}}+C_2\right)}{\left(\mu_x^{\,2}+\mu_{\hat{x}}^{\,2}+C_1\right)\left(\sigma_x^{\,2}+\sigma_{\hat{x}}^{\,2}+C_2\right)} \tag{8}$$

$$SNR\left(x,\hat{x}\right)=10\,log_{10}\left(\frac{MAX_f}{\sqrt{MSE}}\right) \tag{9}$$

where μ is the average of samples, σ is the variance, C_1, C_2 are stability constants, and MAX_f is the maximum magnitude of the seismic data.

3. Results

We validated the proposed network model using synthetic and field data. All computations were executed using a system equipped with an NVIDIA RTX 5080 graphics card. The network model was trained on the PyTorch-1.13.0 platform. To test the noise reduction capability of the proposed method, we compared it with several other methods, including BF, WT, DnCNN, U-Net, SwinIR,[61] Swin-Conv U-Net (SCU),[62] and MCA-SCU. For the non-learning methods, we used WT (db4, level = 3) and BF (spatial kernel width = 12 samples, structure tensor scale = 2). The detailed parameter configurations for all deep learning methods are summarized in Table 1.

3.1. Training on synthetic data with random noise

We first validated the performance of the proposed network model with random noise.

3.1.1. Construction of synthetic datasets with random noise

We first built a training set dedicated to random-noise suppression. The clean seismic data were generated from the Marmousi model. Using this model, we produced 100 synthetic shot gathers, each with a shape of 2,000 time samples × 200 traces. Random noise was subsequently added to the shot gathers, setting the PSNR to 5, 10, 15, and 20 dB, respectively. Figure 5 shows examples of the clean data and noisy data. To enhance computational efficiency during network training and to increase the count of training samples, a sliding window was applied to partition each shot gather, where the window size was 224, and the stride was 28. Among all the shot gathers, 90 shot gathers were used to create the training set, while the other 10 shot gathers were reserved for testing, ensuring that no samples from the test set were exposed during training. This process ultimately yielded 5,850 clean seismic samples and noisy samples.

3.1.2. Comparison of denoising results

First, we evaluated the denoising capability for different network models at various noise levels. Figures 6–9

Table 1. Detailed parameter configurations for deep learning methods

Method	Optimizer	Learning rate	Decay/epoch	Batch size	Epochs	Loss
DnCNN	Adam	1e−3	5%	1	40	MSE
U-Net	Adam	1e−3	5%	1	40	MSE
SwinIR	Adam	1e−3	5%	1	40	MSE
SCU	Adam	1e−3	5%	1	40	MSE
MCA-SCU	Adam	1e−3	5%	1	40	MSE
Our proposed model	Adam	1e−3	5%	1	40	MSE

Abbreviations: DnCNN: Denoising convolutional neural network; MCA: Multi-scale channel attention; MSE: Mean squared error; SCU: Swin-Conv U-Net.

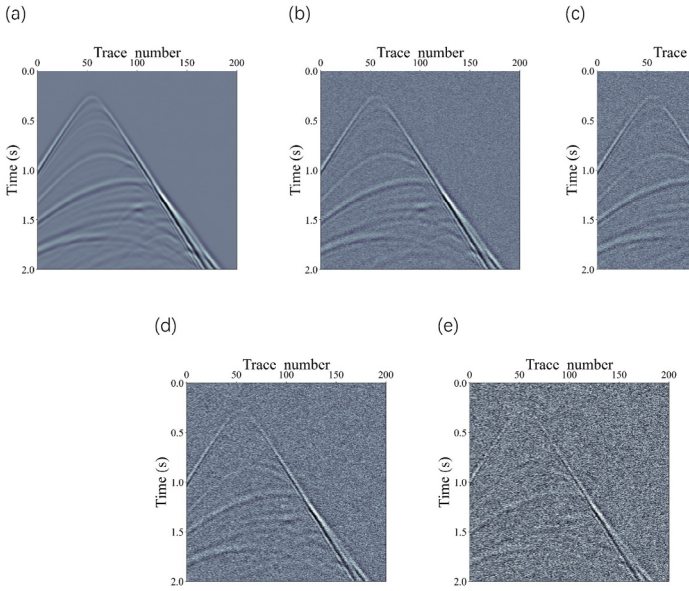


Figure 5. Example of training data. (a) Clean data. (b–e) Noisy data with peak signal-to-noise ratios of (b) 20, (c) 15, (d) 10, and (e) 5 dB.

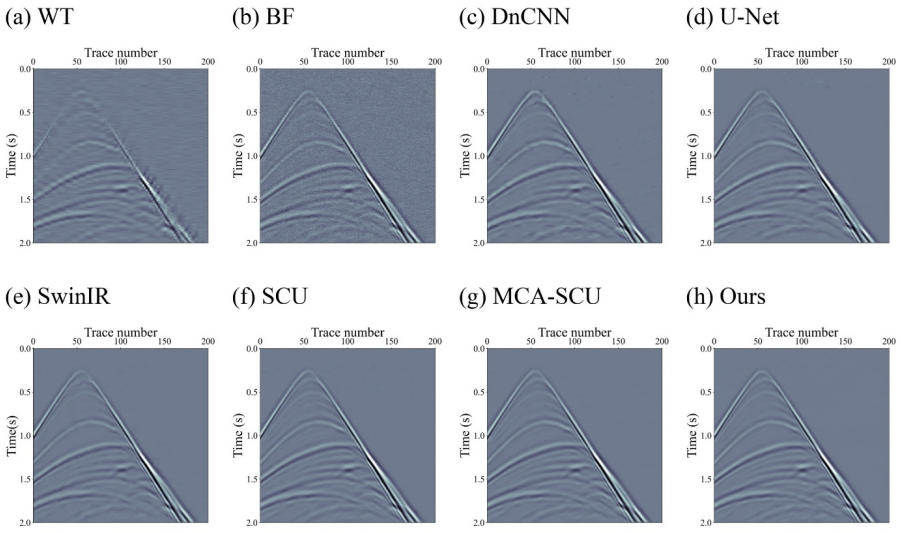


Figure 6. Denoising results of the models at a peak signal-to-noise ratio of 20 dB: (a) wavelet transform (WT), (b) bilateral filtering (BF), (c) denoising convolutional neural network (DnCNN), (d) U-Net, (e) SwinIR, (f) Swin-Conv U-Net (SCU), (g) multi-scale channel attention SCU (MCA-SCU), and (h) our proposed model

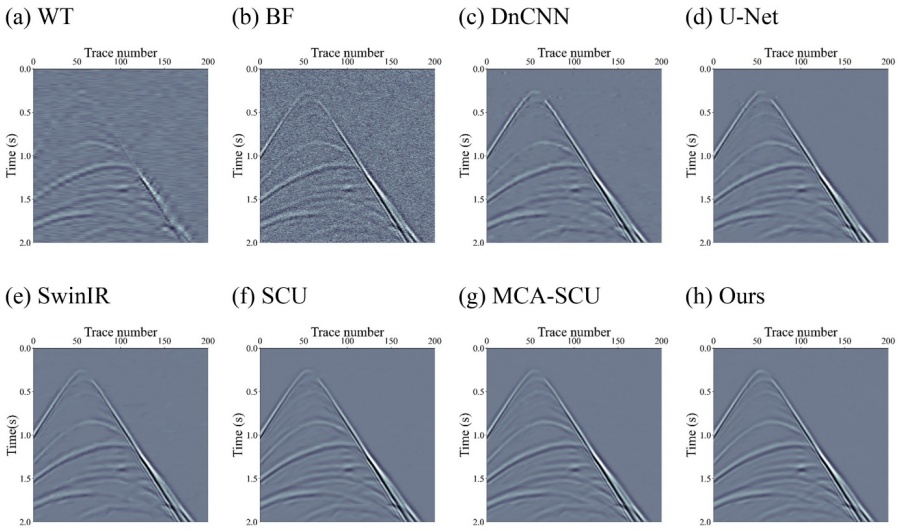


Figure 7. Denoising results of the models at a peak signal-to-noise ratio of 15 dB: (a) wavelet transform (WT), (b) bilateral filtering (BF), (c) denoising convolutional neural network (DnCNN), (d) U-Net, (e) SwinIR, (f) Swin-Conv U-Net (SCU), (g) multi-scale channel attention SCU (MCA-SCU), and (h) our proposed model

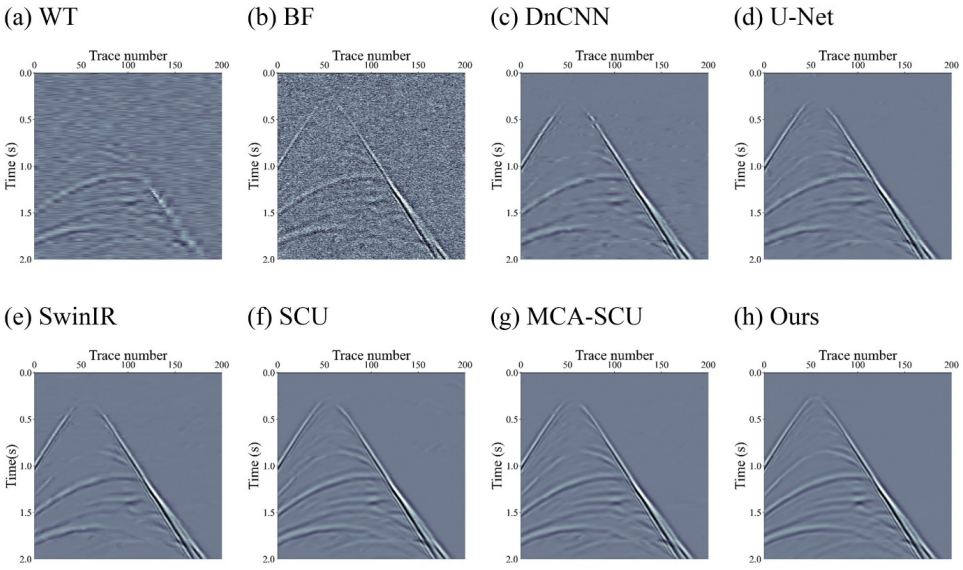


Figure 8. Denoising results of the models at a peak signal-to-noise ratio of 10 dB: (a) wavelet transform (WT), (b) bilateral filtering (BF), (c) denoising convolutional neural network (DnCNN), (d) U-Net, (e) SwinIR, (f) Swin-Conv U-Net (SCU), (g) multi-scale channel attention SCU (MCA-SCU), and (h) our proposed model

show the denoising results of each network model (WT, BF, DnCNN, U-Net, SwinIR, SCU, MCA-SCU, and the proposed network model) at noise levels of 20 dB, 15 dB, 10 dB, and 5 dB, respectively. The comparisons show that WT and BF failed to achieve ideal denoising across all noise intensities, and that DnCNN, U-Net, and SwinIR failed to completely remove noise and severely damaged signal structures at high noise intensities. At low noise intensities, although DnCNN, U-Net, and SwinIR could largely remove noise, they still lost some detailed signal structures. This phenomenon stems from the inherent limitations of the network architectures. Meanwhile, SCU and MCA-SCU achieved notable improvements in denoising results. However, these network models failed to suppress all noise and disrupted the structural continuity of effective signals at high noise intensities. In contrast, the proposed method stably removed noise at all noise intensities and well reconstructed the continuity of seismic events and signal details.

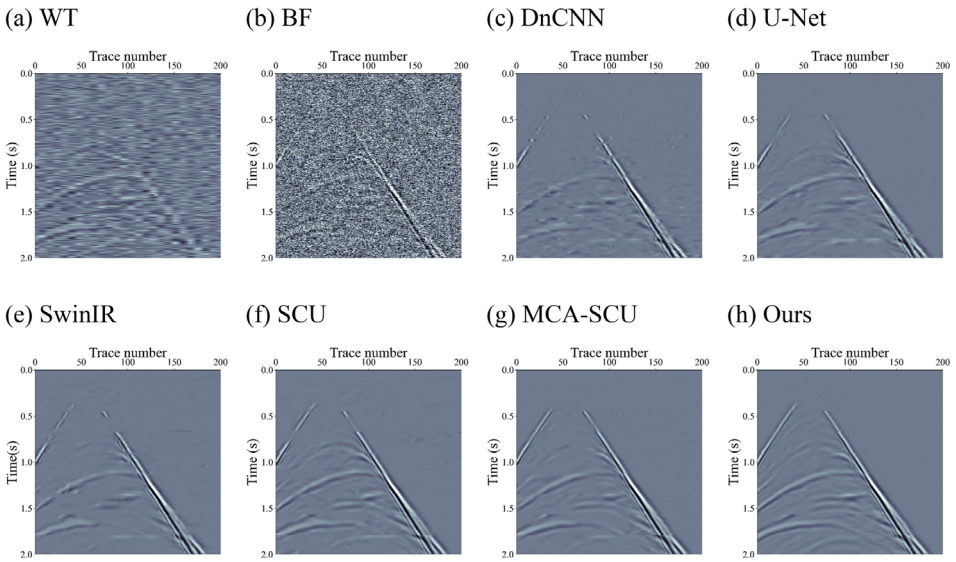


Figure 9. Denoising results of the models at a peak signal-to-noise ratio of 5 dB: (a) wavelet transform (WT), (b) bilateral filtering (BF), (c) denoising convolutional neural network (DnCNN), (d) U-Net, (e) SwinIR, (f) Swin-Conv U-Net (SCU), (g) multi-scale channel attention SCU (MCA-SCU), and (h) our proposed model

To further quantify performance, we compared the PSNR and SSIM of denoising results across these models (Table 2). The proposed model achieved the best PSNR and SSIM at all noise levels. It maintained excellent denoising capability even as noise intensity increased, demonstrating the denoising ability and stability of the proposed architecture.

Aside from the above comparisons, we also computed the residual of the original noisy data at 20 dB with the denoised results (Figure 10). The comparison demonstrates that WT exhibited a relatively blurred boundary between signal and noise, while BF, in order to preserve signal integrity, achieved a comparatively limited extent of noise suppression. DnCNN, U-Net, and SwinIR misidentified a proportion of effective signals as noise and thus damaged these signals. As shown in the red boxes, the residuals of these methods contain clear signal structures, indicating severe signal leakage. SCU and MCA-SCU greatly alleviated this issue; however, they still caused signal leakage by misidentifying parts of weak signals as noise. Conversely, the residual of the proposed method exhibited almost no signal features, demonstrating that it effectively distinguishes weak signals from noise at low noise intensities.

Table 2. Comparison of SSIM and PSNR on random noise across models

PSNR	20 dB		15 dB		10 dB		5 dB	
	PSNR (dB)	SSIM	PSNR (dB)	SSIM	PSNR (dB)	SSIM	PSNR (dB)	SSIM
WT	28.63	0.607	26.00	0.406	23.94	0.237	20.95	0.101
BF	24.19	0.233	18.49	0.084	13.73	0.033	08.13	0.009
DnCNN	38.38	0.895	35.57	0.816	33.60	0.753	30.61	0.667
U-Net	38.49	0.914	36.47	0.866	34.30	0.809	31.76	0.740
SwinIR	39.65	0.919	37.13	0.856	34.57	0.790	31.42	0.698
SCU	42.86	0.945	39.59	0.897	36.96	0.846	33.07	0.753
MCA-SCU	43.06	0.947	39.87	0.902	37.28	0.854	33.26	0.763
Our proposed model	43.26	0.951	40.44	0.915	37.88	0.871	34.35	0.805

Abbreviations: BF: Bilateral filtering; DnCNN: Denoising convolutional neural network; MCA: Multi-scale channel attention; PSNR: Peak signal-to-noise ratio; SCU: Swin-Conv U-Net; SSIM: Structural similarity index; WT: Wavelet transform.

Figure 11 shows the local similarity mapping of denoising results from each model at 15 dB, where the color transition from white to black indicates increasing similarity. The proposed method maintained higher consistency across most regions, indicating that it can more precisely recover the continuity in seismic events and suppress noise.

Lastly, we used single-trace data to compare the recovered signal details of each model. We extracted the 50th (Figure 12a) and 100th (Figure 12b) traces from the denoising results for SCU, MCA-SCU, and the proposed network model at 10 dB. SCU (green) and MCA-SCU (blue) exhibited misidentifications and polarity reversals at weak reflection events. Their amplitude fidelity and phase consistency also showed deviations, as indicated by the black arrows. In contrast, the proposed method (red) achieved improvement, and its overall waveform was closer to the ground truth (black).

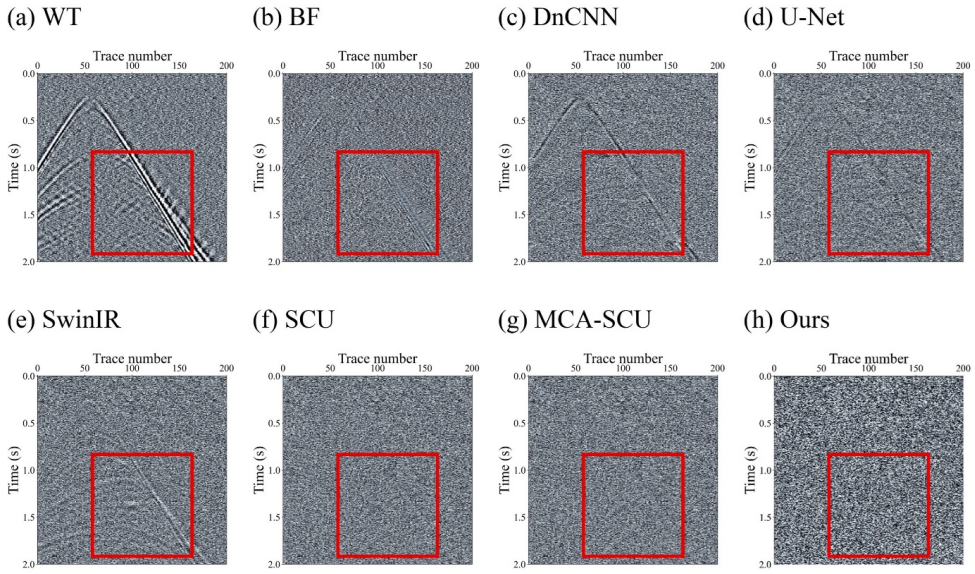


Figure 10. Residuals of the original noisy data at 20 dB with the denoised results across models: (a) wavelet transform (WT), (b) bilateral filtering (BF), (c) denoising convolutional neural network (DnCNN), (d) U-Net, (e) SwinIR, (f) Swin-Conv U-Net (SCU), (g) multi-scale channel attention SCU (MCA-SCU), and (h) our proposed model. The red boxes highlight signal leakage.

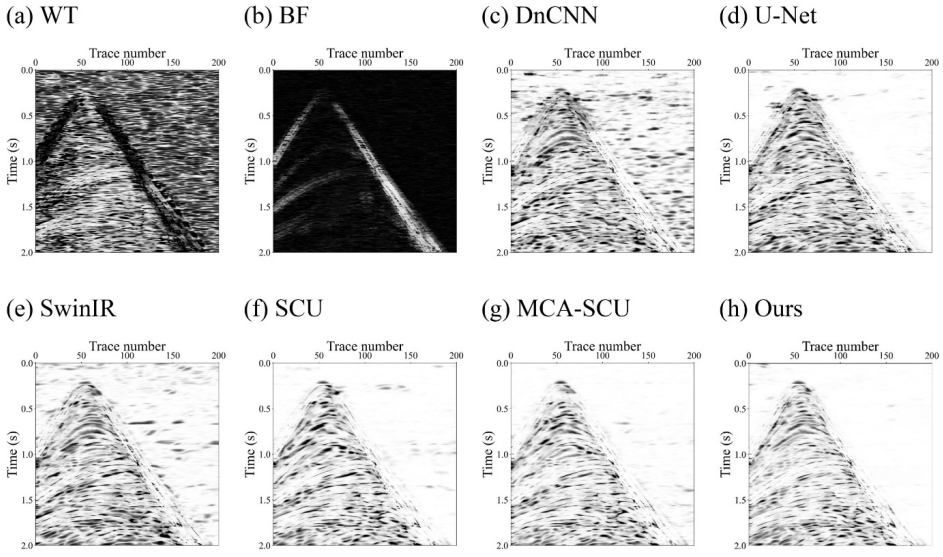


Figure 11. Local similarity mapping between clean shot gather and denoising results across models: (a) wavelet transform (WT), (b) bilateral filtering (BF), (c) denoising convolutional neural network (DnCNN), (d) U-Net, (e) SwinIR, (f) Swin-Conv U-Net (SCU), (g) multi-scale channel attention SCU (MCA-SCU), and (h) our proposed model

Based on the comprehensive experimental analysis, the proposed method accurately removed noise and preserved signal features. The success is due to several reasons. First, the CNN branch uses the local receptive field of the convolution kernels and the BS-Conv property to focus on extracting local features from seismic data. Second, the Transformer branch uses RALA to capture global information and generate diverse output features, ensuring accurate signal recovery under high background noise. Lastly, the introduction of GCB reduces the transmission of noise features, lessening their negative impact on subsequent layers. This design ensures the model removes noise effectively and recovers the signal accurately under high background noise.

3.1.3. Analysis of robustness to random noise intensity

We also conducted tests to assess the robustness of the proposed method to noise intensity. Specifically, the model was trained on noisy samples with PSNRs of 5, 10, 15, and 20 dB, and then it was tested at higher PSNR levels (21, 22, and 23 dB), corresponding to lower noise intensities, and lower PSNR levels (2, 3, and 4 dB), corresponding to higher noise intensities.

Figures 13 and 14 compare the denoising results and residual for a noise level of 23 dB. When the noise intensity deviated from the training noise level, signal leakage became more pronounced (as highlighted by the red boxes in Figure 14). By contrast, the proposed network achieved the cleanest separation between noise and effective signals and demonstrated its robustness to variations in noise intensity.

Figure 15 compares denoising results at an extreme noise level with a PSNR of 2 dB, higher than the training noise intensity. WT, BF, DnCNN, U-Net, SwinIR, SCU, and MCA-SCU failed to remove all noise and severely damaged signal details, such as the continuity of seismic events at high noise intensity. These robustness experiments confirm that the proposed network can reduce noise and reconstruct the signal even when the noise intensity is outside the training distribution.

Table 3 compares the PSNR and SSIM for each model at both higher and lower noise levels. The PSNR and SSIM for all models decreased when the test noise intensity exceeded the training noise intensity. This is because these models face stronger noise interference, and thus their denoising ability becomes limited. Conversely, when the noise intensity was below the training level, PSNR and SSIM increased because weaker noise allows the models to recover signals more easily. The proposed method achieved the highest PSNR and SSIM at all noise levels. This success stems from the joint effect of BS-Conv, RALA, and GCB, which enhances the capability for the model to distinguish signals from noise.

3.1.4. Ablation experiments

Next, we conducted ablation experiments to test the effect of every module and the necessity of the dual-branch parallel architecture. We designed eight ablation configurations (Figure 16a–h; Table 4): Figure 16a with U-Net++; Figure

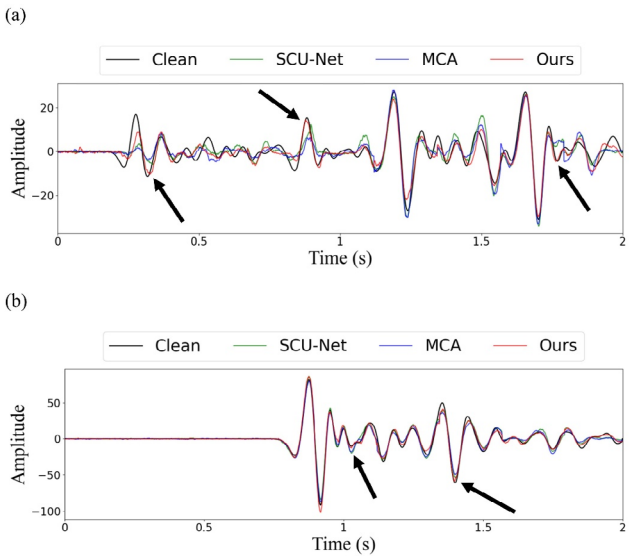


Figure 12. Comparison of single traces using SCU, MCA-SCU, and the proposed method at (a) the 50th and (b) the 100th traces. The black arrows indicate waveform deviations of other methods from the ground truth, while the proposed method is more consistent.
Abbreviations: MCA: Multi-scale channel attention; SCU: Swin-Conv U-Net.

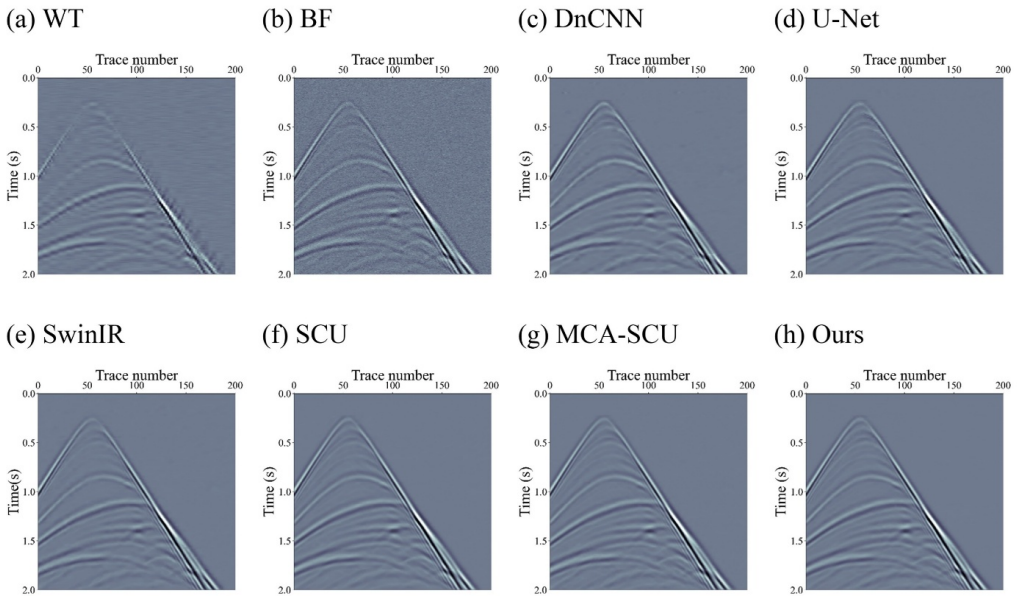


Figure 13. Denoising results of the models at a peak signal-to-noise ratio of 23 dB: (a) wavelet transform (WT), (b) bilateral filtering (BF), (c) denoising convolutional neural network (DnCNN), (d) U-Net, (e) SwinIR, (f) Swin-Conv U-Net (SCU), (g) multi-scale channel attention SCU (MCA-SCU), and (h) our proposed model

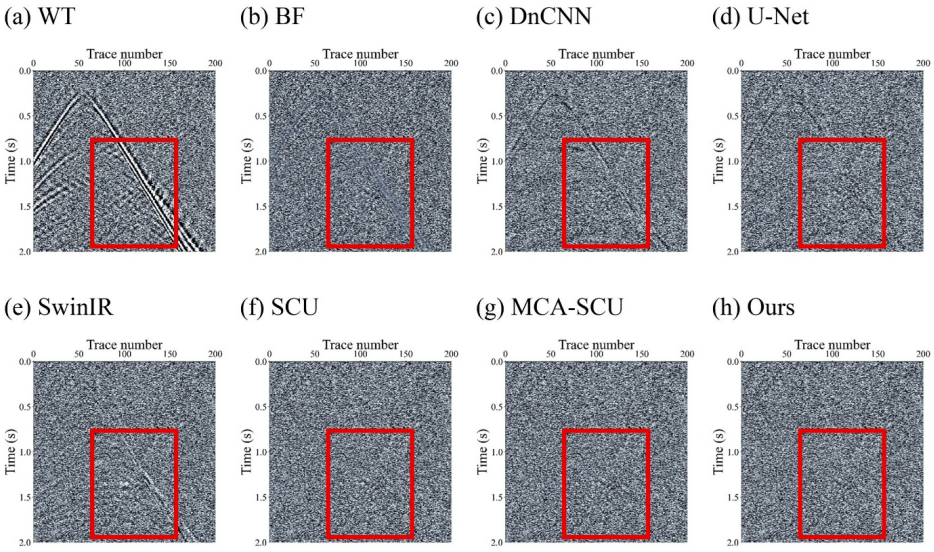


Figure 14. Residual of the original noisy data at 23 dB with the denoised results across models: (a) wavelet transform (WT), (b) bilateral filtering (BF), (c) denoising convolutional neural network (DnCNN), (d) U-Net, (e) SwinIR, (f) Swin-Conv U-Net (SCU), (g) multi-scale channel attention SCU (MCA-SCU), and (h) our proposed model. The red boxes highlight signal leakage.

16b and c incrementally introduce BS-Conv and GCB into U-Net++ to evaluate their individual effects; Figure 16d incorporates RALA and GCB into U-Net++ to assess the role of RALA; Figure 16e stacks BS-Conv, RALA, and GCB together within a U-Net++ to verify the necessity of the dual-branch design; and Figure 16f–h represent the complete dual-branch parallel architecture proposed in this paper, where Figure 16f adopts the SSIM loss, Figure 16g adopts a hybrid MSE+SSIM loss, and Figure 16h adopts the MSE loss.

Figure 16 shows the denoising results for each model at a noise level of 5 dB, and Table 5 compares their SSIM and PSNR. As shown in Figure 16b and c, the single CNN branches (one with only BS-Conv, the other with both BS-Conv and GCB) preserved more weak-signal information in their denoising results compared to

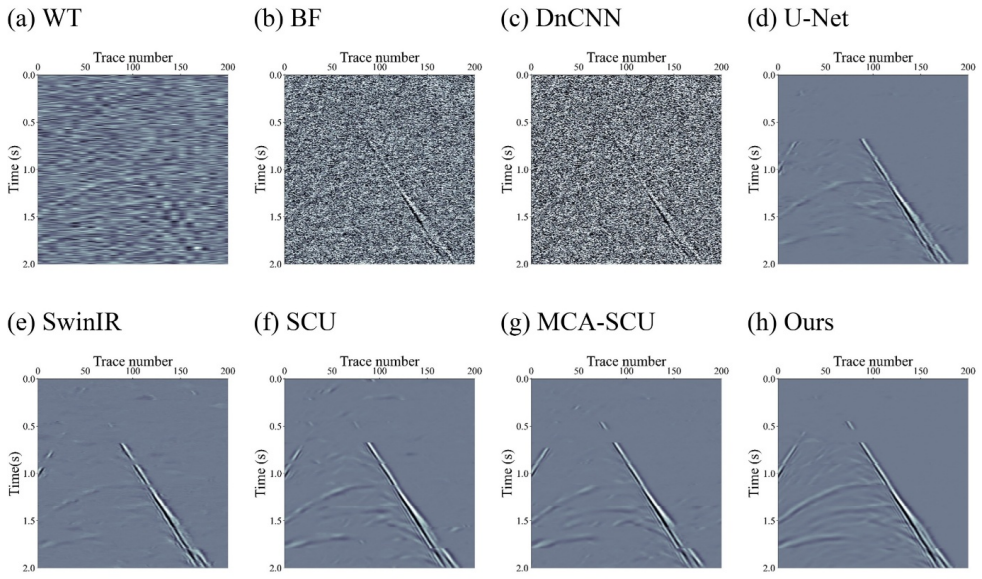


Figure 15. Denoising results of the models at a peak signal-to-noise ratio of 2 dB: (a) wavelet transform (WT), (b) bilateral filtering (BF), (c) denoising convolutional neural network (DnCNN), (d) U-Net, (e) SwinIR, (f) Swin-Conv U-Net (SCU), (g) multi-scale channel attention SCU (MCA-SCU), and (h) our proposed model

Table 3. Robustness comparison: SSIM and PSNR of denoising results across models on random noise

PSNR	23 dB		22 dB		21 dB		4 dB		3 dB		2 dB	
	PSNR (dB)	SSIM	PSNR (dB)	SSIM	PSNR (dB)	SSIM	PSNR (dB)	SSIM	PSNR (dB)	SSIM	PSNR (dB)	SSIM
WT	29.88	0.686	29.37	0.654	28.88	0.624	20.37	0.086	19.55	0.064	18.75	0.050
BF	26.83	0.352	25.67	0.295	24.67	0.252	07.33	0.007	06.23	0.005	05.22	0.004
DnCNN	39.61	0.920	39.38	0.910	39.51	0.905	29.21	0.635	29.21	0.635	28.31	0.614
U-Net	39.00	0.929	38.86	0.928	38.47	0.918	31.12	0.722	30.20	0.692	29.22	0.671
SwinIR	40.30	0.933	40.23	0.934	39.95	0.922	30.74	0.680	29.90	0.628	29.20	0.602
SCU	43.94	0.957	43.69	0.956	43.24	0.951	32.44	0.734	31.38	0.704	30.38	0.665
MCA-SCU	43.90	0.959	43.96	0.958	43.51	0.951	32.41	0.737	31.34	0.700	30.26	0.671
Our proposed model	44.54	0.961	44.28	0.960	43.86	0.954	33.64	0.784	32.53	0.750	31.58	0.728

Abbreviations: BF: Bilateral filtering; DnCNN: Denoising convolutional neural network; MCA: Multi-scale channel attention; PSNR: Peak signal-to-noise ratio; SCU: Swin-Conv U-Net; SSIM: Structural similarity index; WT: Wavelet transform.

U-Net++. Table 5 quantifies these improvements with PSNR gains of 0.25 dB and 0.58 dB over U-Net++. This is because the BS-Conv and GCB modules help the network effectively distinguish weak signals from noise. Meanwhile, Figure 16d shows notable improvements in both weak-signal preservation and noise suppression, indicating that the global receptive field and the output feature diversity of RALA are well suited to the task of seismic data denoising. Notably, the hybrid single-branch model in Figure 16e underperformed the single Transformer branch in Figure 16d, with a PSNR drop of 0.20 dB, as shown in Table 5, as serial stacking of CNN and Transformer modules induces mutual interference between the two mechanisms. Figure 16f–h compare three loss functions (SSIM, hybrid MSE-SSIM, and MSE) on our full architecture. Among them, MSE achieved the best overall performance, followed by the hybrid loss, while SSIM alone performed worst. We therefore retained MSE as the default loss function in our method.

Table 4. Ablation configurations for module effectiveness and dual-branch necessity

Figure	Abbreviation	BS-Conv	RALA	GCB	Loss	Architecture
16a	U-Net++	–	–	–	MSE	Single branch
16b	With BS-Conv	✓	–	–	MSE	Single branch
16c	With BS-Conv and GCB	✓	–	✓	MSE	Single branch
16d	With RALA and GCB	–	✓	✓	MSE	Single branch
16e	Single hybrid	✓	✓	✓	MSE	Single branch
16f	Our proposed model (full architecture)	✓	✓	✓	SSIM	Dual branch
16g	Our proposed model (full architecture)	✓	✓	✓	SSIM+MSE	Dual branch
16h	Our proposed model (full architecture)	✓	✓	✓	MSE	Dual branch

Abbreviations: BS-Conv: Blind-spot convolution; GCB: Gated convolutional neural network block; MSE: Mean squared error; RALA: Rank-augmented linear attention; SSIM: Structural similarity index.

Table 5. PSNR and SSIM of denoising results for each configuration in Figure 16 at a noise level of 5 dB

PSNR	5 dB	
	PSNR (dB)	SSIM
U-Net++	32.27	0.735
Single CNN branch with BS-Conv	32.52	0.755
Single CNN branch with BS-Conv and GCB	32.85	0.760
Single Transformer branch with RALA and GCB	34.17	0.797
Single branch hybrid CNN and Transformer	33.97	0.784
Full architecture with SSIM loss	32.91	0.780
Full architecture with hybrid loss	34.07	0.797
Full architecture with MSE loss	34.35	0.805

Abbreviations: BS-Conv: Blind-spot convolution; CNN: Convolutional neural network; GCB: Gated convolutional neural network block; MSE: Mean squared error; PSNR: Peak signal-to-noise ratio; RALA: Rank-augmented linear attention; SSIM: Structural similarity index.

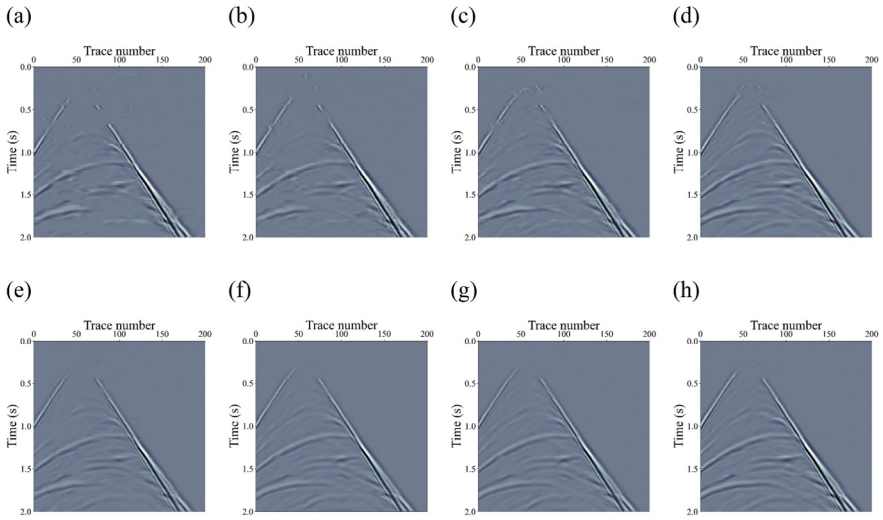


Figure 16. Ablation experiment at a noise level of 5 dB. (a) U-Net++, (b) with BS-Conv, (c) with BS-Conv and GCB, (d) with RALA and GCB, (e) single hybrid, and (f–h) the complete dual-branch parallel architecture proposed in this paper, where (f) adopts the SSIM loss, (g) adopts a hybrid MSE+SSIM loss, and (h) adopts the MSE loss.
Abbreviations: BS-Conv: Blind-spot convolution; GCB: Gated convolutional neural network block; MSE: Mean squared error; RALA: Rank-augmented linear attention; SSIM: Structural similarity index.

To verify the adaptive selection capability of GCB, we extracted the feature maps before and after GCB at the middle skip connection and visualized them for comparison. Figure 17a and b show the feature maps before and after GCB gating, respectively. We observed that after GCB gating, feature activations in strong reflection regions became significantly enhanced (brighter), while the activations in background noise regions were effectively suppressed (darker). This differentiated response between signal regions and noise regions directly verifies that GCB can adaptively enhance effective signal features and suppress noise features.

3.1.5. Receptive field analysis and feature visualization

As shown in Table 6, the hybrid convolutional blocks in the CNN branch adopted 3×3 kernels, resulting in an initial receptive field of only 1.3% when data entered this branch. Even at the bottleneck layer, where the receptive field expanded to 21.4%, the feature maps were already highly abstract and lost original spatial details. In contrast, the Transformer branch maintained a global receptive field at every stage through RALA. This mathematical distinction fundamentally demonstrates that the two branches operate at different spatial scales.

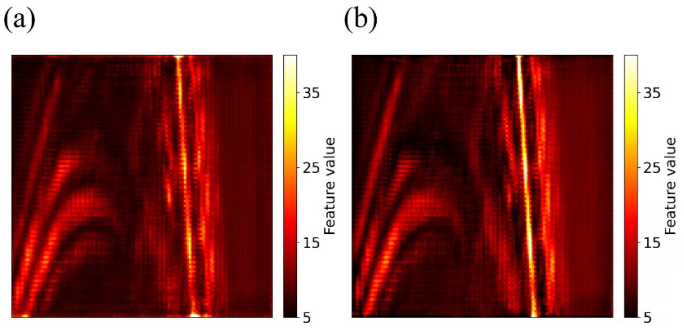


Figure 17. Feature maps (a) before and (b) after gated convolutional neural network block (GCB)

Table 6. Comparison of the receptive fields of the two branches

Branch	Stage	Receptive field	Proportion
CNN	Input	3 × 3	3/224 ≈ 1.3%
CNN	After 1 down-sample	3 × 3	3/112 ≈ 2.7%
CNN	After 2 down-samples	3 × 3	3/56 ≈ 5.4%
CNN	After 3 down-samples	3 × 3	3/28 ≈ 10.7%
CNN	After 4 down-samples	3 × 3	3/14 ≈ 21.4%
Transformer	Every stage	224 × 224	224/224 = 100%

Abbreviation: CNN: Convolutional neural network.

We further visualized the intermediate feature maps of the two branches (Figure 18), where Figure 18a denotes the noisy input, Figure 18b the feature map from the CNN branch, and Figure 18c the feature map from the Transformer branch. The CNN branch activated sparsely around regions with sharp amplitude variations, whereas the Transformer branch produced widespread activations across the entire spatial domain. This visualization, consistent with the theoretical analysis in Table 6, visually confirms the complementary spatial scales of the two branches and validates the necessity of the dual-branch cooperative design.

3.2. Synthetic shot gather test based on real noise from field data

After the random-noise experiments, we validated the practicality of the proposed model using real noise. Therefore, we reconstructed a new training set specifically for real noise.

3.2.1. Construction of synthetic datasets with field noise

The construction of the synthetic dataset with real noise shares similarities with the random-noise dataset, but it also differs in key ways. Unlike the artificial random-noise

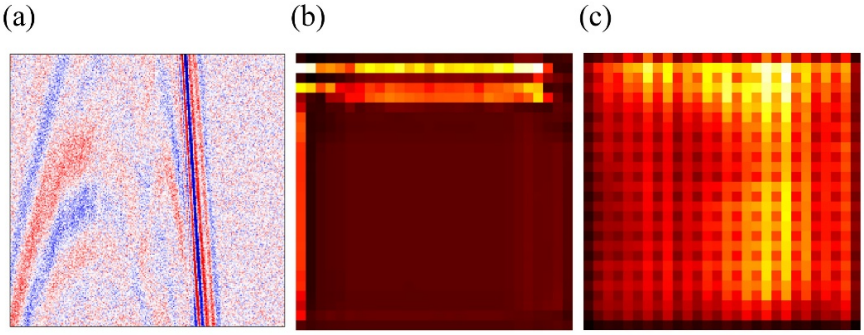


Figure 18. Feature maps of the two branches. (a) Noisy input. (b) Feature map from the convolutional neural network (CNN) branch. (c) Feature map from the Transformer branch.

generator, we extracted pure-noise segments from field data. The clean shot gathers were still generated from the Marmousi model. Then, we added these noise segments to the clean shot gathers to acquire noisy seismic data. Lastly, the same sliding window strategy was applied to increase the number of training samples. Figure 19 illustrates an example: field seismic data with a red box that indicates a pure noise segment (Figure 19a), a clean synthetic shot gather (Figure 19b), and the resulting noisy data at a PSNR of 10 dB (Figure 19c).

To further clarify the statistical distinction between the two noise types, we compared their amplitude spectra (Figure 20), where Figure 20a–c correspond to the amplitude spectra of Gaussian random noise, pure field noise, and raw field seismic data, respectively. The Gaussian noise exhibited a flat and uniform spectrum, consistent with the white-noise assumption. In contrast, the field noise segments showed concentrated frequency content and spatially coherent patterns, revealing clear non-Gaussian characteristics. The raw field data displayed

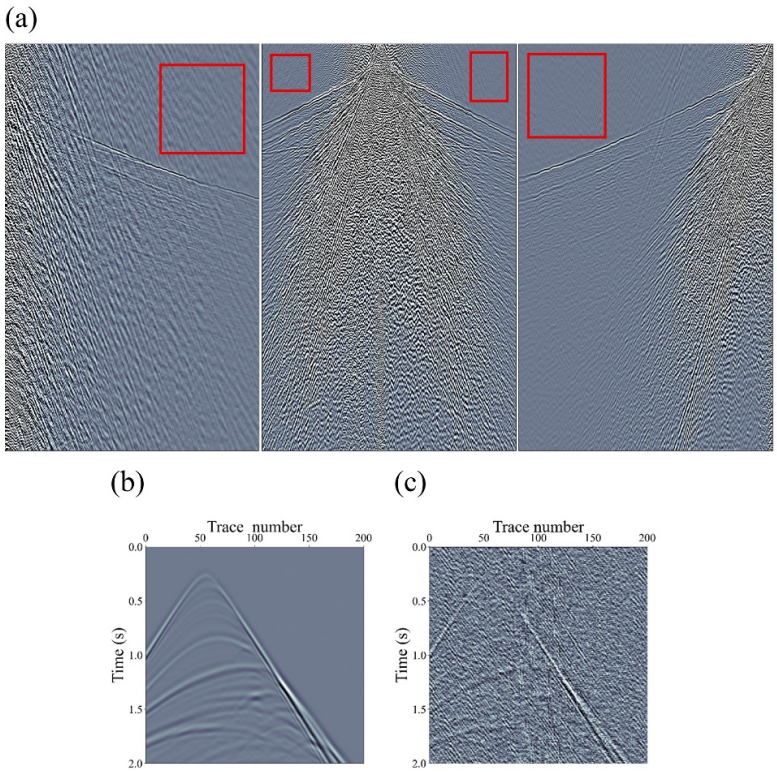


Figure 19. An example of the construction of synthetic datasets with real noise from field data. Field data with a red box indicating a pure noise segment. (a) Field seismic data with a red box that indicates a pure noise segment. (b) A clean synthetic shot gather. (c) The resulting noisy data at a peak signal-to-noise ratio of 10 dB.

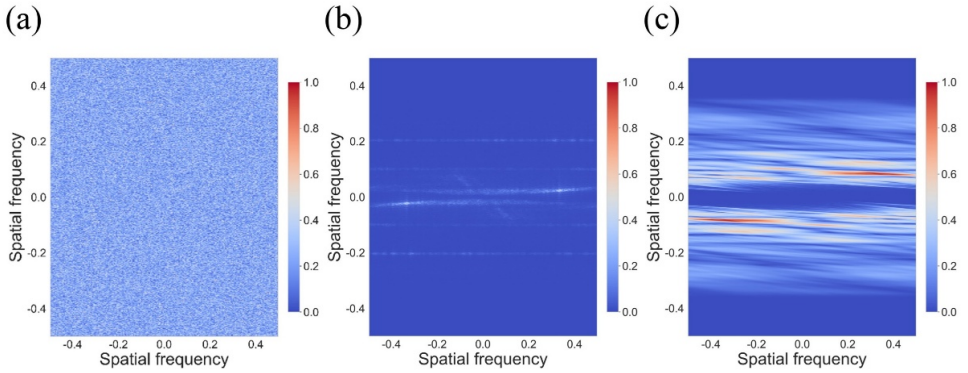


Figure 20. Amplitude spectra of (a) Gaussian random noise, (b) pure-noise segment extracted from field data, and (c) raw field seismic data

even more complex frequency features due to the mixture of signals and noise. These distinct characteristics confirm that the two noise types are statistically different, justifying two independent training sets rather than mixing them.

3.2.2. Comparison of denoising results

Next, we assessed the denoising capability of each model based on real noise at a level of 10 dB. Figure 21 shows the denoising results. Table 7 compares the SSIM and PSNR of these denoising results based on real noise at 10 dB. WT, BF, DnCNN, and U-Net barely removed real noise, and their denoising results achieved extremely low SSIM and PSNR. This is because these methods fail to distinguish between real noise and signal features, indicating that neither traditional methods nor convolution alone is sufficient for real noise. Meanwhile, SwinIR, SCU, MCA-SCU, and the proposed method showed notable improvements, demonstrating the effectiveness of global attention for real seismic data. Among these methods, our improved dual-branch network achieved the highest PSNR and SSIM, showing that it can stably remove real noise and reconstruct the continuity of seismic events and signal details.

Figure 22 shows the local similarity mapping of denoising results from each model at 10 dB. The proposed network showed the highest agreement with the clean data, confirming that it can better recover the continuity for seismic events and suppress noise.

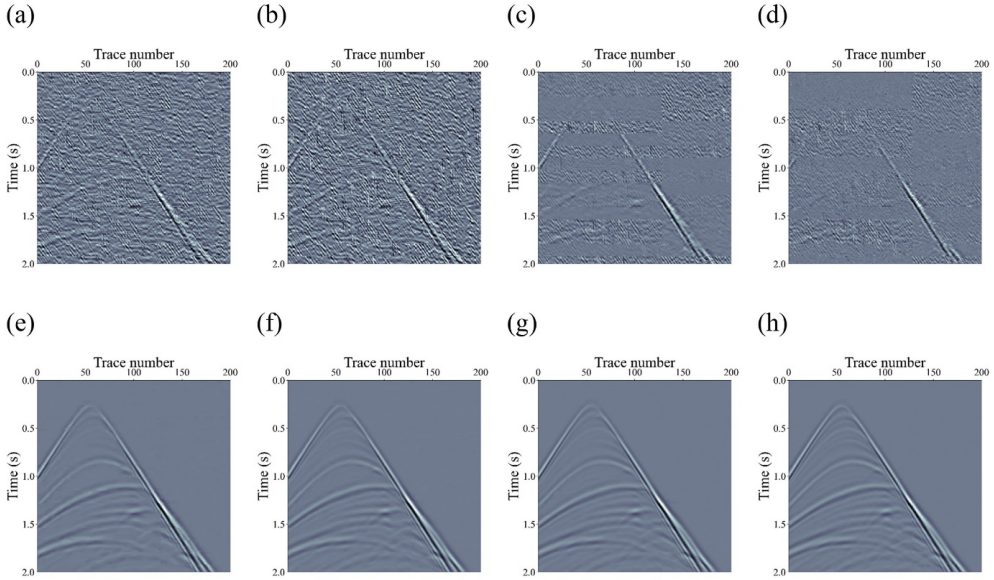


Figure 21. Denoising results of the models with real noise at a peak signal-to-noise ratio of 10 dB: (a) wavelet transform (WT), (b) bilateral filtering (BF), (c) denoising convolutional neural network (DnCNN), (d) U-Net, (e) SwinIR, (f) Swin-Conv U-Net (SCU), (g) multi-scale channel attention SCU (MCA-SCU), and (h) our proposed model

Table 7. Denoising results of different models for real noise

PSNR	10 dB	
	PSNR (dB)	SSIM
WT	14.05	0.021
BF	12.13	0.017
DnCNN	19.07	0.162
U-Net	19.07	0.093
SwinIR	36.45	0.902
SCU	40.69	0.954
MCA-SCU	41.60	0.964
Our proposed model	42.63	0.972

Abbreviations: BF: Bilateral filtering; DnCNN: Denoising convolutional neural network; MCA: Multi-scale channel attention; PSNR: Peak signal-to-noise ratio; SCU: Swin-Conv U-Net; SSIM: Structural similarity index; WT: Wavelet transform.

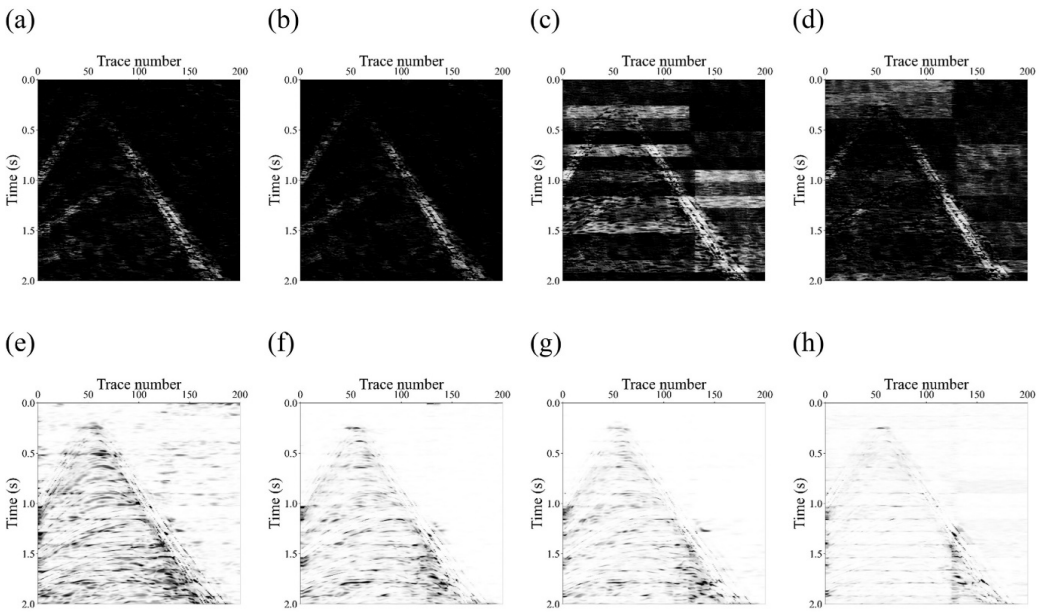


Figure 22. Local similarity mapping between clean shot gather and denoising results based on real noise at 10 dB: (a) wavelet transform (WT), (b) bilateral filtering (BF), (c) denoising convolutional neural network (DnCNN), (d) U-Net, (e) SwinIR, (f) Swin-Conv U-Net (SCU), (g) multi-scale channel attention SCU (MCA-SCU), and (h) our proposed model

3.2.3. Field data test

Next, we verified the model performance with field data. It should be emphasized that the field data test presented in this subsection is not merely a demonstration of practical applicability, but also serves as a rigorous cross-domain generalization evaluation. Figure 23 shows the denoising results of all methods, with the yellow boxes highlighting the extent of noise removal in the noise-only regions. A noteworthy phenomenon occurred: WT, BF, DnCNN, and U-Net exhibited a basic ability to suppress noise in field data despite their poor performance on synthetic tests with real noise. This is because WT and BF are traditional methods with inherent noise-suppressing mechanisms, while DnCNN and U-Net benefit from the translation invariance of CNNs. However, the denoising results of SwinIR, SCU, and MCA-SCU were unsatisfactory. Although the denoised results appeared structurally correct, the amplitude of the effective signal was either too large or too small. This is because these models have learned a mapping between Marmousi synthetic signals and field noise. Consequently, when applied to field data whose signal structures differ significantly from the training set, the learned mapping no

longer holds, rendering the models incapable of effectively distinguishing effective signals from noise. In short, the proposed model achieved the best balance between noise suppression and signal preservation.

We further evaluated the amplitude fidelity of different methods through spectral analysis, as presented in Figure 24. Figure 24a–d show the linear amplitude spectra of the denoised results from SwinIR, SCU, MCA-SCU, and

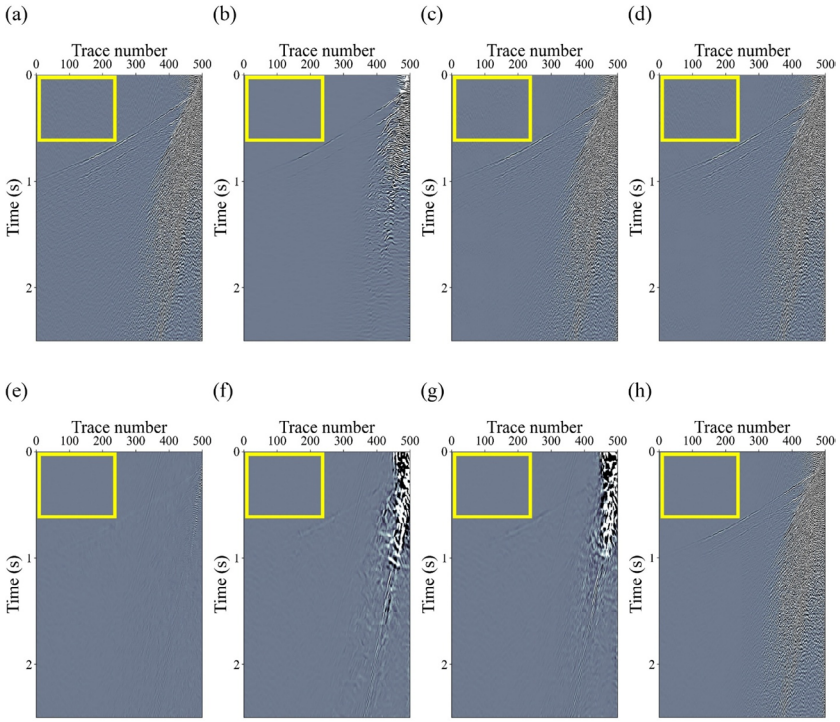


Figure 23. Denoising results of each model on the field data: (a) wavelet transform (WT), (b) bilateral filtering (BF), (c) denoising convolutional neural network (DnCNN), (d) U-Net, (e) SwinIR, (f) Swin-Conv U-Net (SCU), (g) multi-scale channel attention SCU (MCA-SCU), and (h) our proposed model. The yellow boxes indicate denoising performance in pure-noise regions of the field data.

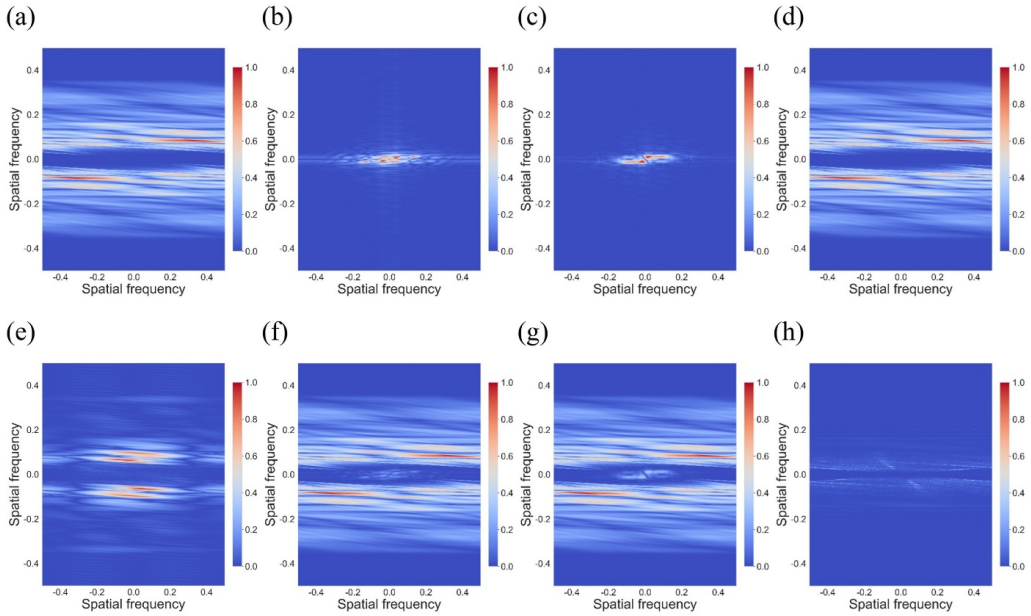


Figure 24. Spectral comparison across methods on field data. Linear amplitude spectra of (a) SwinIR, (b) Swin-Conv U-Net (SCU), (c) multi-scale channel attention SCU (MCA-SCU), and (d) our proposed model. (e–h) The corresponding residual maps.

the proposed method, respectively. Figure 24e–h show the corresponding residual maps, defined as the absolute difference between the amplitude spectrum of the field data and that of each denoised result. SwinIR, SCU, and MCA-SCU exhibited varying degrees of deviation in the frequency domain. In contrast, the amplitude spectrum of the proposed method closely matched that of the raw field data across the entire frequency range, demonstrating higher spectral fidelity.

These experiments demonstrate that our model maintained robust performance across varying signal structures and remained effective on real seismic data. This stability comes from the joint action of BS-Conv, RALA, and GCB. The introduction of BS-Conv enhances the local feature extraction ability of the model. RALA provides global features and increases feature diversity. Thus, RALA ensures that the model can distinguish signal features in seismic data from noise. GCB in the skip connections adaptively selects signal-related features and prevents the propagation of noise features.

4. Discussion

Our proposed dual-branch network model, which integrates the local detail capture of CNN with the global feature extraction capability of Transformers, significantly outperforms existing methods in denoising performance and signal fidelity. Nevertheless, several limitations exist in this study. The proposed method was trained and validated on Gaussian random noise, and its generalizability to other noise types, such as coherent noise, remains to be verified. Furthermore, the reliance of the current supervised learning framework on synthetically generated clean data motivates our future exploration of unsupervised or self-supervised learning frameworks to improve the model's applicability to real-world field data.

5. Conclusion

This paper proposes a dual-branch network for seismic denoising that combines CNN and Transformer branches. The design aims to precisely remove noise while improving signal quality in seismic data. The proposed network uses U-Net++ as the backbone of the dual-branch architecture and incorporates BS-Conv, RALA, and GCB. This network can capture local details and extract global features of seismic data. Consequently, the method improves the continuity of seismic reflection events and suppresses high-level noise. The proposed model was compared with several classical methods and networks, including WT, BF, DnCNN, U-Net, SwinIR, SCU, and MCA-SCU. We first validated its denoising ability on synthetic random noise, then tested robustness, conducted ablation experiments, and finally evaluated applicability on real noise. After comparison, the proposed method achieves higher noise suppression and better continuity of seismic events at different noise levels.

Acknowledgments

None.

Funding

This work is supported by the National Natural Science Foundation of China (grant 42374153), Shaanxi Provincial Key R&D Program (grant 2026CY-YBXM-818), and the Provincial Natural Science Foundation (grant LY23D040001).

Conflict of interest

Naijian Wang is an employee of BGP, Inc., but was not involved in any activities that could constitute a conflict of interest in relation to this study. The authors declare that there are no conflicts of interest regarding the execution, drafting, or dissemination of this study.

Author contributions

Conceptualization: Jingrui Luo, Guoxin Chen
Data curation: Naijian Wang
Formal analysis: Jingrui Luo, Guoxin Chen
Methodology: Zhaoyang Ge
Writing–original draft: Zhaoyang Ge
Writing–review & editing: Jingrui Luo

Availability of data

The data used to support the findings of this study are available from the first author upon request (gzy_xaut@163.com).

References

1. Zhu W, Mousavi SM, Beroza GC. Seismic signal denoising and decomposition using deep neural networks. *IEEE Trans Geosci Remote Sens.* 2019;57(11):9476-9488.

   doi: 10.1109/TGRS.2019.2926772

2. Iqbal N. DeepSeg: Deep segmental denoising neural network for seismic data. *IEEE Trans Neural Netw Learn Syst.* 2023;34(7):3397-3404.

   doi: 10.1109/TNNLS.2022.3205421

3. Zhang S, Zhang L, Qin X. Seismic signal denoising using variational mode decomposition and a denoising convolutional neural network. *J Seism Explor.* 2025;34(2):44-59.

   doi: 10.36922/JSE025260030

4. Cooper HW, Cook RE. Seismic data gathering. *Proc IEEE.*

1984;72(10):1266-1275.

doi: 10.1109/PROC.1984.13016

5. Wang H, Chen W, Huang W, *et al*. Nonstationary predictive filtering for seismic random noise suppression—a tutorial. *Geophysics.* 2021;86(3):W21-W30.

doi: 10.1190/geo2020-0368.1

6. Krohn C, Ronen S, Deere J, Gulunay N. Introduction to this special section—seismic noise. *Leading Edge.* 2008;27(2):163-165.

doi: 10.1190/1.2840362

7. Dong X, Li Y, Yang B. Desert low-frequency noise suppression by using adaptive DnCNNs based on the determination of high-order statistic. *Geophys J Int.* 2019;219(2):1281-1299.

doi: 10.1093/gji/ggz363

8. Canales LL. Random noise reduction. In: *SEG Technical Program Expanded Abstracts.* SEG; 1984:525-527.

doi: 10.1190/1.1894168

9. Abma R, Claerbout J. Lateral prediction for noise attenuation by t-x and f-x techniques. *Geophysics.* 1995;60(6):1887-1896.

doi: 10.1190/1.1443920

10. Gulunay N. Noncausal spatial prediction filtering for random noise reduction on 3-D poststack data. *Geophysics.* 2000;65(5):1641-1653.

doi: 10.1190/1.1444852

11. Trickett SR. F-xy eigenimage noise suppression. *Geophysics.* 2003;68(2):751-759.

doi: 10.1190/1.1567245

12. Liu G, Chen X. Noncausal f-x-y regularized nonstationary prediction filtering for random noise attenuation on 3D seismic data. *J Appl Geophys.* 2013;93:60-66.

doi: 10.1016/j.jappgeo.2013.03.007

13. Liu G, Liu Y, Li C, Chen X. Weighted multisteps adaptive autoregression for seismic image denoising. *IEEE Geosci Remote Sens Lett.* 2018;15(9):1342-1346.

doi: 10.1109/LGRS.2018.2841840

14. Hale D. Structure-oriented bilateral filtering of seismic images. In: *SEG Technical Program Expanded Abstracts.* SEG; 2011:3596-3600.

doi: 10.1190/1.3627947

15. Mousavi SM, Langston CA, Horton SP. Automatic microseismic denoising and onset detection using the synchrosqueezed continuous wavelet transform. *Geophysics.* 2016;81(4):V341-V355.

doi: 10.1190/geo2015-0598.1

16. Liu W, Cao S, Chen Y, Zu S. An effective approach to attenuate random noise based on compressive sensing and curvelet transform. *J Geophys Eng.* 2016;13(2):135-145.

doi: 10.1088/1742-2132/13/2/135

17. Qiao T, Ren J, Wang Z, *et al*. Effective denoising and classification of hyperspectral images using curvelet transform and singular spectrum analysis. *IEEE Trans Geosci Remote Sens.* 2017;55(1):119-133.

doi: 10.1109/TGRS.2016.2598065

18. Li J, Sacchi MD. An lp-space matching pursuit algorithm and its application to robust seismic data denoising via time-domain Radon transforms. *Geophysics.* 2021;86(3):V171-V183.

doi: 10.1190/geo2020-0136.1

19. Anvari R, Siahsar MAN, Gholtashi S, Kahoo AR, Mohammadi M. Seismic random noise attenuation using synchrosqueezed wavelet transform and low-rank signal matrix approximation. *IEEE Trans Geosci Remote Sens.* 2017;55(11):6574-6581.

doi: 10.1109/TGRS.2017.2730228

20. Saad OM, Chen Y, Savvaidis A, *et al*. Unsupervised deep learning for single-channel earthquake data denoising and its applications in event detection and fully automatic location. *IEEE Trans Geosci Remote Sens.* 2022;60:1-10.

doi: 10.1109/TGRS.2022.3209932

21. Anvari R, Kahoo AR, Mohammadi M, Khan NA, Chen Y. Seismic random noise attenuation using sparse low-rank estimation of the signal in the time–frequency domain. *IEEE J Sel Top Appl Earth Obs Remote Sens.* 2019;12(5):1612-1618.

doi: 10.1109/JSTARS.2019.2906360

22. Cadzow JA. Signal enhancement-a composite property mapping algorithm. *IEEE Trans Acoust Speech Signal Process.* 1988;36(1):49-62.

doi: 10.1109/29.1488

23. Trickett S. F-xy Cadzow noise suppression. In: *SEG Technical Program Expanded Abstracts.* SEG; 2008:2586-2590.

doi: 10.1190/1.3063880

24. Oropeza V, Sacchi M. Simultaneous seismic data denoising and reconstruction via multichannel singular spectrum analysis. *Geophysics.* 2011;76(3):V25-V32.

doi: 10.1190/1.3552706

25. Chen K, Sacchi MD. Robust reduced-rank filtering for erratic seismic noise attenuation. *Geophysics.* 2015;80(1):V1-V11.

doi: 10.1190/geo2014-0116.1

26. Dong X, Zhong T, Li Y. New suppression technology for low-frequency noise in desert region: the improved robust principal component analysis based on prediction of neural network. *IEEE Trans Geosci Remote Sens.* 2020;58(7):4680-4690.

doi: 10.1109/TGRS.2020.2966054

27. Liu X, Chen X, Li J, Chen Y. Nonlocal weighted robust principal component analysis for seismic noise attenuation. *IEEE Trans Geosci Remote Sens.* 2021;59(2):1745-1756.

doi: 10.1109/TGRS.2020.2996686

28. Bekara M, van der Baan M. Random and coherent noise attenuation by empirical mode decomposition. *Geophysics.* 2009;74(5):V89-V98.

doi: 10.1190/1.3157244

29. Chen Y, Ma J. Random noise attenuation by f-x empirical-mode decomposition predictive filtering. *Geophysics.* 2014;79(3):V81-V91.

doi: 10.1190/geo2013-0080.1

30. Liu N, Li F, Wang D, Gao J, Xu Z. Ground-roll separation and attenuation using curvelet-based multichannel variational mode decomposition. *IEEE Trans Geosci Remote Sens.* 2022;60:1-14.

doi: 10.1109/TGRS.2021.3054749

31. Lin H, Liu C, Wang S, Ye W. Attention mechanism-based deep denoiser for desert seismic random noise suppression. *Acta Geophys.* 2023;71(4):2781-2793.

doi: 10.1007/s11600-023-01062-z

32. Sun Y, Wei Z, Jia X, Zhu C. A deep learning-based method for enhancing isotropic reverse time migration in complex media. *Acta Geophys.* 2025;73(6):4003-4022.

doi: 10.1007/s11600-025-01563-z

33. Li D, Jiang P, Yang S, Zhang F. Multi-task deep learning for multi-parameter elastic inversion. *Acta Geophys.* 2025;73(4):2443-2460.

doi: 10.1007/s11600-024-01500-6

34. Xi H, Luo J, Liu J, *et al*. MSBE-UNet: A deep learning denoising method for effective seismic noise suppression. *Acta Geophys.* 2026;74(1):65.

doi: 10.1007/s11600-026-01799-3

35. Yang Z, Xue Q, Wang X, Wang T. TFDenoiser-Edge: A hybrid convolutional neural network–Transformer framework for real-time seismic denoising on edge devices in extreme environments. *J Seism Explor*. 2026;35(2):025010138.

doi: 10.36922/JSE025010138

36. Goodfellow IJ, Pouget-Abadie J, Mirza M, *et al*. Generative adversarial nets. *Commun ACM.* 2020;63(11):139-144.

doi: 10.1145/3422622

37. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells W, Frangi A, eds. *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015.* Lecture Notes in Computer Science; Vol 9351. New York, NY: Springer; 2015:234-241.

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

38. Zhang K, Zuo W, Chen Y, Meng D, Zhang L. Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising. *IEEE Trans Image Process.* 2017;26(7):3142-3155.

doi: 10.1109/TIP.2017.2662206

39. Xie P, Boelle JL, Puntous H. Generative adversarial network based fast noise removal on land seismic data. In: *SEG Technical Program Expanded Abstracts*. SEG; 2018:2171-2175.

doi: 10.1190/segam2018-2995310.1

40. Zhao Y, Li Y, Dong X, Yang B. Low-frequency noise suppression method based on improved DnCNN in desert seismic data. *IEEE Geosci Remote Sens Lett.* 2019;16(5):811-815.

doi: 10.1109/LGRS.2018.2882058

41. Luo R, Li Y. Ji Yu RUnet Juan Ji Shen Jing Wang Luo De Di Zhen Zi Liao Sui Ji Zao Sheng Ya Zhi Fang Fa [Random seismic noise attenuation based on RUnet convolutional neural network]. *Geophys Prospect Pet.* 2020;59(1):51-59. [In Chinese]

doi: 10.3969/j.issn.1000-1441.2020.01.006

42. Yuan Y, Si X, Zheng Y. Ground-roll attenuation using generative adversarial networks. *Geophysics.* 2020;85(4):WA255-WA267.

doi: 10.1190/geo2019-0414.1

43. Wang H, Li Y, Dong X. Generative adversarial network for desert seismic data denoising. *IEEE Trans Geosci Remote Sens.* 2021;59(8):7062-7075.

doi: 10.1109/TGRS.2020.3030692

44. Yao H, Ma H, Li Y, Feng Q. DnResNeXt network for desert seismic data denoising. *IEEE Geosci Remote Sens Lett.* 2022;19:1-5.

doi: 10.1109/LGRS.2020.3044036

45. Zhong T, Cheng M, Dong X, Li Y, Wu N. Seismic random noise suppression by using deep residual U-Net. *J Pet Sci Eng.* 2022;209:109901.

doi: 10.1016/j.petrol.2021.109901

46. Dodda VC, Kuruguntla L, Mandpura AK, Elumalai K. Simultaneous seismic data denoising and reconstruction with attention-based wavelet-convolutional neural network. *IEEE Trans Geosci Remote Sens.* 2023;61:1-14.

doi: 10.1109/TGRS.2023.3267037

47. Wang W, Chen HL, Chang DK, Wang XY, Wang SJ, Li D. Dual-branch dense network for seismic background noise elimination. *J Seism Explor.* 2025;34(5):53-65.

doi: 10.36922/JSE025290038

48. Liu Z, Lin Y, Cao Y, *et al*. Swin transformer: hierarchical vision transformer using shifted windows. In: Proceedings

of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV). October 11-17, 2021; Online. IEEE; 2021:9992-10002.

doi: 10.1109/ICCV48922.2021.00986

49. Katharopoulos A, Vyas A, Pappas N, Fleuret F. Transformers are RNNs: fast autoregressive transformers with linear attention. In: III HD, Singh A, eds. P*roceedings of the 37th International Conference on Machine Learning.* PMLR; 2020;119:5156-5165.

50. Wang H, Lin J, Li Y, *et al*. Self-supervised pretraining transformer for seismic data denoising. *IEEE Trans Geosci Remote Sens.* 2024;62:1-25.

doi: 10.1109/TGRS.2024.3368282

51. Li F, Liu H, Wang W, Ma J. Swin transformer for seismic denoising. *IEEE Geosci Remote Sens Lett.* 2024;21:1-5.

doi: 10.1109/LGRS.2024.3358234

52. Chen J, Chen G, Li J, *et al*. Efficient seismic data denoising via deep learning with improved MCA-SCUNet. *IEEE Trans Geosci Remote Sens.* 2024;62:1-14.

doi: 10.1109/TGRS.2024.3355972

53. Li J, Li JY, Li Y, Wu N, Zhang M. Dual-branch DAS denoising strategy based on SNR-aware feature fusion. *Meas Sci Technol.* 2025;36(8):086142.

doi: 10.1088/1361-6501/adfc89

54. Chen HS, Wang J, Liu BD. Dual-domain Mamba for seismic random noise suppression. *J Appl Geophys.* 2025;243:105951.

doi: 10.1016/j.jappgeo.2025.105951

55. Guo Y, He JJ, Yi SX, *et al*. Ji Yu DnMamba De Di Zhen Shu Ju Zao Sheng Qu Chu Fang Fa Yan Jiu Yu Ying Yong [Research on DnMamba-based seismic data denoising method and its application]. *Comput Tech Geophys Geochem Explor.* Published online December 1, 2025. [In Chinese]

doi: 10.12474/wthtjs.20250623-0001

56. Hu Y, Cheng D, Huang Z, Chen B, Lin S, Zhang S. DSCA-former: Dual-stem cross-attentive transformer for image denoising. *Neurocomputing.* 2026;671:132656.

doi: 10.1016/j.neucom.2026.132656

57. Krull A, Buchholz TO, Jug F. Noise2Void-Learning denoising from single noisy images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 16-20, 2019; Long Beach, CA. IEEE; 2019:2124-2132.

doi: 10.1109/CVPR.2019.00223

58. Fan Q, Huang H, He R. Breaking the low-rank dilemma of linear attention. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 11-15, 2025; Nashville, TN. IEEE; 2025:25271-25280.

doi: 10.1109/CVPR52734.2025.02353

59. Yu W, Wang X. MambaOut: do we really need Mamba for vision? In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 11-15, 2025; Nashville, TN. IEEE; 2025:4484-4496.

doi: 10.1109/CVPR52734.2025.00423

60. Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, Liang J. Unet++: A nested U-Net architecture for medical image segmentation. In: Stoyanov D, *et al*., eds. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Lecture Notes in Computer Science; Vol 11045. New York, NY: Springer; 2018:3-11.

doi: 10.1007/978-3-030-00889-5_1

61. Liang J, Cao J, Sun G, *et al*. SwinIR: image restoration using Swin transformer. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops. October 11-17, 2021; Online. 2021;1833-1844.

doi: 10.1109/ICCVW54120.2021.00210

62. Zhang K, Li Y, Liang J, *et al*. Practical blind image denoising via Swin-Conv-UNet and data synthesis. *Mach Intell Res.* 2023;20(6):822-836.

doi: 10.1007/s11633-023-1466-0