ACCSCIENCE PUBLISHING

Journal of Seismic Exploration

ARTICLE

TSU-former seismic data denoising method based on spatiotemporal attention mechanism

Zhenjing Yao[1,2], Lan Qin[1,2], Meng Lu[3]*, Jingyi Zhu[4], and Chenyu Zhang[1,2]

[1]School of Emergency Communication and Control Engineering, University of Emergency Management, Sanhe, Hebei, China

[2]Hebei Key Laboratory of Seismic Disaster Instrument and Monitoring Technology, Sanhe, Hebei, China

[3]Key Laboratory of Drilling Technology in Mine Rescue, Ministry of Emergency Management, Beijing, China

[4]International Education College, Beijing University of Chemical Technology, Beijing, China

Abstract

Seismic data are frequently contaminated by severe environmental noise, reducing the signal-to-noise ratio (SNR) and degrading subsequent processing and interpretation. Although U-Net-based deep learning has been widely used for denoising, such networks suffer from local receptive fields and an imperfect balance between noise suppression and signal preservation. This paper proposes TSU-Former, a supervised seismic data denoising network built on a symmetric U-Net encoder–decoder with a Transformer bottleneck and a spatiotemporal attention mechanism. An improved convolutional unit with a residual design is integrated into both the encoder and decoder to stabilize training and extract multi-scale local features. The Transformer bottleneck models non-local spatiotemporal dependencies within each input window, preserving the continuity of coherent phase axes. At the network output, a spatiotemporal self-adaptive noise enhancement module recalibrates features along the trace and time dimensions to improve the accuracy of noise prediction. Quantitative evaluations on synthetic and field datasets demonstrate that TSU-Former achieves higher SNR, peak SNR, and structural similarity than five representative denoising baselines. In field-data experiments, the estimated SNR is improved by 13.61 dB relative to the noisy input, and the denoised waveforms retain the principal signal characteristics.

Keywords: Denoising; Deep learning; U-Net; Attention mechanism; Transformer

***Corresponding author:**
Meng Lu
(ll1063477367@163.com)

Citation: Yao Z, Qin L, Lu M, Zhu J, Zhang C. TSU-former seismic data denoising method based on spatiotemporal attention mechanism. *J Seismic Explor.* doi: 10.36922/JSE026290142

Received: July 17, 2026

Revised: October 4, 2026

Accepted: October 16, 2026

Published online: November 19, 2026

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

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

1. Introduction

Seismic events inflict severe damage on built environments, resulting in substantial human fatalities, financial costs, and cascading hazards like slope failures and epidemic surges,[1,2] necessitating focused efforts to mitigate disaster impacts. Concurrently, these phenomena serve as pivotal probes into planetary interior mechanics; by deciphering seismic waveforms, researchers elucidate subterranean architectures[3-6] and kinematic processes governing crustal deformation.[7] During transmission through geologically intricate media—such as laterally varying sedimentary layers—or amid anthropogenic noise sources like urban infrastructure operations, seismic waveforms commonly

suffer signal degradation. This noise-induced attenuation of signal fidelity[8-10] critically compromises downstream interpretive workflows, rendering robust signal-to-noise ratio (SNR) enhancement and targeted denoising protocols indispensable for reliable subsurface characterization.

Field-acquired seismic records frequently exhibit overlapping spectral signatures from diverse noise sources—including non-stationary ground roll and cultural interference—inducing severe waveform distortion. This spectral congestion critically undermines the efficacy of conventional denoising techniques, particularly within geologically complex settings where signal-noise boundaries remain ambiguous,[11] rendering adaptive filtering protocols essential for reliable subsurface imaging. Conventional spectral-domain filters—including bandpass techniques targeting surface-wave attenuation,[12,13] based on self-attention residual network[14] and time-frequency peak-based approaches[15-17]—exploit spectral separation between target signals and coherent interference. However, their efficacy diminishes sharply under band overlap conditions, where waveform smearing and high-frequency detail erosion (e.g., fault-plane truncations) manifest as critical imaging artifacts. Sparsity-driven time-frequency decompositions—including wavelet-based multi-scale denoising for dispersive surface-wave suppression,[18-20] curvelet transforms optimized for fault-fracture zone imaging,[21-23] and Shearlet-adapted thresholding targeting anisotropic noise attenuation[24]—rely on sparse coefficient representation within overcomplete dictionaries. For non-stationary seismic data, optimal threshold selection (e.g., Stein's unbiased risk estimate for hard thresholding) is critical; suboptimal values induce amplitude distortion in thin-bed reflections or phase smearing across fault boundaries, degrading high-resolution reservoir characterization. Consequently, data-adaptive threshold calibration protocols must account for local SNR variations to preserve subtle stratigraphic features. Adaptive modal decomposition frameworks—empirical mode decomposition with ensemble noise level calibration for channelized reservoir delineation,[25-27] complete ensemble empirical mode decomposition with adaptive noise constrained by intrinsic mode function counts to suppress tidal-noise interference[28] and VMD optimized via bandwidth parameters to extract thin-bed tuning responses[29-31]—decompose non-stationary seismic traces into physically interpretable components. However, over-decomposition in gas-cloud zones induces amplitude leakage across stratigraphic boundaries, while under-decomposition obscures subtle fault displacements below 5 m. Conversely, rank-reduction techniques (e.g., singular value decomposition leveraging singular value gaps[32] and multichannel singular spectrum analysis with data-adaptive window lengths[33]) enhance signal coherence in three-dimensional seismic volumes by isolating low-rank signal subspaces from noise-dominated components. Their efficacy diminishes in subsalt imaging where complex wavefield scattering violates low-rank assumptions, necessitating local rank selection protocols that dynamically adjust to structural complexity gradients to preserve channel-edge resolution.

Deep learning has advanced substantially in recent years,[34] particularly in seismic signal denoising. Si & Yuan[35] applied a denoising convolutional neural network (CNN) to suppress random noise in seismic data, using residual learning to yield clearer denoised signals. Zhao *et al.*[36] adapted a denoising CNN for desert seismic data, targeting low-frequency noise suppression and significantly improving the SNR of low-frequency components. Jiang *et al.*[37] developed a convolutional autoencoder-based deep neural network to suppress random noise and interpolate seismic traces in seismic data. Wang & Nealon[38] applied a U-Net architecture for supervised denoising of marine seismic data, reducing noise while preserving critical signal features. Wang *et al.*[39] enhanced a convolutional blind denoising-residual dense network to eliminate both noise and artifacts. For improved denoising, Li *et al.*[40] employed dilated convolutions, Gao *et al.*[41] utilized residual dense blocks, and multi-scale convolution techniques were adopted by Feng *et al.*,[42] Liu *et al.*,[43] and Tan and Le.[44] Current deep learning approaches demonstrate advantages in seismic data denoising, with CNNs dominating noise suppression but limited in capturing global information. During convolution, filters across layers learn distinct features through training, extracting local features via finite-sized kernels. This structure restricts global information comprehension.[45] Li *et al.*[46] introduced the Swin Transformer, enhancing cross-region information interaction through local and shifted window self-attention, though it was tested on a small dataset. Chen *et al.*[47] developed MCA-SCUNet to capture both global features and local details, improving adaptability to complex seismic data, but its denoising efficacy requires further validation on field data. Qi *et al.*[48] proposed MR-SCU, an iterative multi-round denoising method achieving thorough noise reduction while preserving signals, though training duration is long and efficacy diminishes with additional rounds.

However, the aforementioned methods all operate within a supervised learning paradigm, relying on clean/noisy paired data for training, which is often difficult to satisfy in field settings where clean references are typically unavailable, and the distribution discrepancy between synthetic and field data may limit model generalization. To

address this limitation, self-supervised learning methods have been introduced to seismic denoising, aiming to train networks using only noisy data without requiring clean labels. Saad *et al.*[49] proposed a self-supervised multi-stage deep learning network comprising two parallel sub-networks—a strong denoiser with fewer neurons and a weak denoiser with more neurons—trained synchronously on overlapping noisy patches using a Transformer-inspired multi-head attention architecture, with both sub-networks employing the noisy data as the training target and achieving satisfactory denoising performance on both synthetic and field data. Saad and Alkhalifah[50] further extended this self-supervised paradigm to seismic data denoising based on a conditional diffusion model, where the reverse diffusion process is guided by the maximum-scale coefficient of the continuous wavelet transform, and the network is trained in a self-supervised manner with noisy patches serving as both input and target, generating denoised outputs in a single step during inference and demonstrating promising results in distributed acoustic sensing denoising tasks. Self-supervised approaches thus offer a promising direction for reducing reliance on clean references.

Despite the significant advancements in deep learning, existing U-Net-based seismic data denoising models still suffer from critical limitations. First, the local receptive field of U-Net restricts the ability to capture long-range correlations across traces and time, resulting in compromised continuity of coherent phase axes. Second, achieving a balance between noise suppression and effective signal preservation remains challenging, often leading to the loss or waveform distortion of weak reflection events after denoising.

The recently proposed U-Trans model of Saad *et al.*[51] shows that combining a U-shaped encoder-decoder with a Transformer representation can yield strong results in seismic waveform representation and downstream earthquake tasks. In this context, the TSU-Former seismic data denoising method based on the spatiotemporal attention mechanism proposed in this paper improves performance through the following innovative designs:

(i) Improved convolutional unit (ICU): This module achieves symmetric processing within the encoder and decoder. It mitigates the vanishing gradient problem in deep layers via residual connections while extracting multi-scale local features. This design effectively adapts to the multi-scale characteristics of seismic signals, preventing the loss of weak reflection events during downsampling.

(ii) Transformer bottleneck module: Used to capture long-distance correlations across traces and time in seismic profiles, significantly enhancing the ability to identify the continuity of the same-phase axis, overcoming the problem of global information loss in traditional U-Net caused by a local receptive field;

(iii) Spatiotemporal self-adaptive noise enhancement module (ST-ANEM): Placed at the end of the network, it enhances noise features through a spatiotemporal attention mechanism while suppressing effective signal features, enabling the model to more accurately distinguish noise from signals.

2. Methods

2.1. Network architecture

We propose a TSU-Former seismic data denoising network based on a spatiotemporal attention mechanism. As illustrated in Figure 1, TSU-Former is a U-Net variant consisting of three encoder (downsampling) blocks paired with three corresponding decoder (upsampling) blocks, with a Transformer module integrated at the bottleneck layer. Each encoder block employs the ICU, specifically designed in this study, to extract multi-scale local features; similarly, each decoder block incorporates the ICU as the primary component to maintain symmetry and feature consistency throughout the encoding-decoding process. Skip connections transmit the output features from each encoder layer to the corresponding decoder block for feature integration. An ST-ANEM is introduced at the network terminal to enhance noise features.

The initial input channel count is set to 64, doubling after each downsampling stage. Downsampling is accomplished through max pooling, while upsampling utilizes transposed convolutions, which have been demonstrated to more effectively restore spatial details compared to bilinear interpolation.

2.2. Theoretical basis of denoising

In seismic data processing, field-collected noisy seismic data can be represented by the following linear superposition model (**Equation 1**):

$$Y = X_{\text{clean}} + n \tag{1}$$

where Y denotes the observed noisy seismic data (with two-dimensional [2D] $\mathrm{R}^{H \times W}$ for multichannel data and one-dimensional (1D) dimension R^{L} for single-channel time series, where H is the number of seismic channels, W is the number of time sampling points, and L is the length of the single-channel time series), X_{clean} represents the effective signal, and n denotes the noise component.

The core objective of the proposed network trained in a supervised manner is to learn a nonlinear mapping. Specifically, the input is the noisy data Y (**Equation 1**);

the training target is the noise component n, for which the learned mapping $f_\theta(\cdot)$ predicts the noise residual, such that the output $n_{out} = f_\theta(Y)$ closely approximates n, thereby enabling the reconstruction of the denoised signal through $X_{denoised} = Y\text{-}n_{out}$; and the supervision is imposed through a loss function defined on the predicted noise, $L(\theta) = \left\| f_\theta(\mathrm{Y}) - \mathrm{n} \right\|^2$.

This mapping is accomplished through four sequential stages: the encoder layer extracts multi-scale feature representations; the Transformer module captures long-range spatiotemporal dependencies within seismic profiles; the decoder layer integrates and reconstructs high-resolution feature representations; and the spatio-temporal adaptive noise enhancement module executes the final noise augmentation.

2.3. Improved convolutional unit module

We employ the ICU module as the core component of TSU-Former to construct each encoder-decoder pair (Figure 2). It extracts clean-signal features from seismic data, improving model representation. The network includes three downsampling and three upsampling layers. Downsampling is achieved through max pooling operations, while upsampling utilizes transposed convolution. Compared with bilinear interpolation, transposed convolution demonstrates superior capability in restoring spatial details and recovering effective features that may be lost during the downsampling process.

Specifically, input features are processed by the ICU and subsequently downsampled via max pooling; in the decoder pathway, upsampled features are concatenated with the skip connection features from the corresponding encoder layer and then processed by the ICU. This architectural design enables the network to effectively suppress fused noise while simultaneously restoring spatial details.

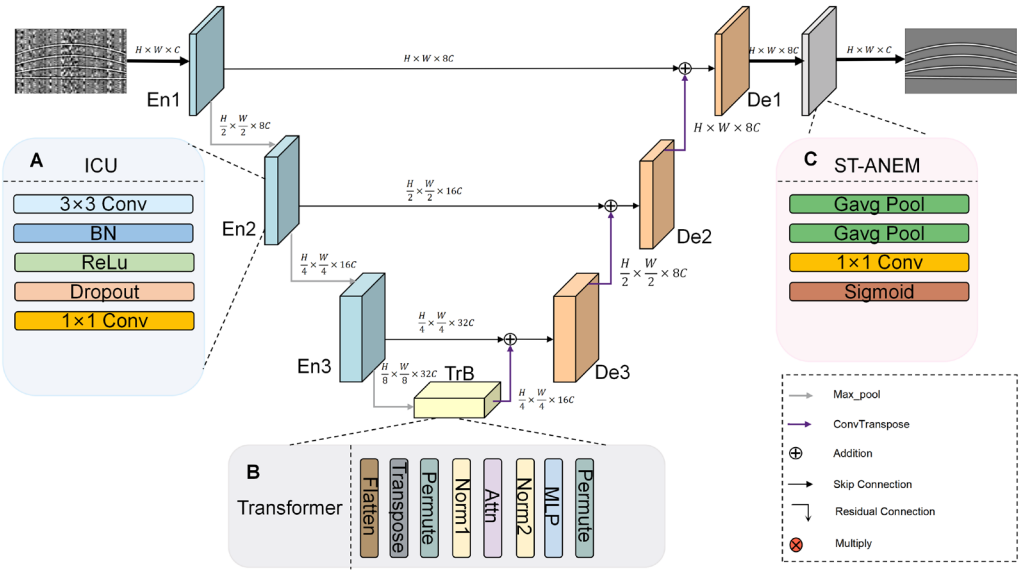


Figure 1. TSU-Former seismic data denoising model based on spatiotemporal attention mechanism. (A) ICU. (B) Transformer. (C) ST-ANEM.
Abbreviations: BN: Batch normalization; ICU: Improved convolutional unit; ReLU: Rectified linear unit; ST-ANEM: Spatiotemporal Self-Adaptive Noise Enhancement Module.

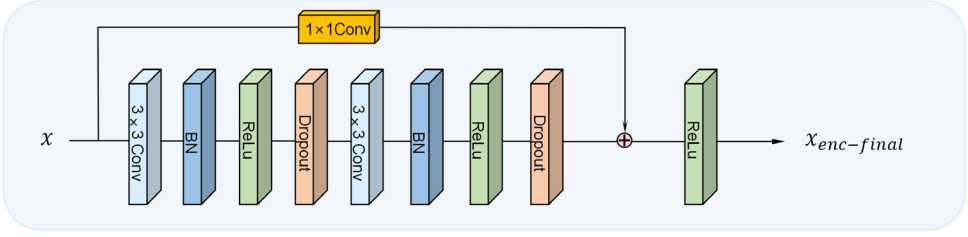


Figure 2. Improved convolutional unit module
Abbreviations: BN: Batch normalization; ReLU: Rectified linear unit.

The central mechanism depicted in Figure 2 elucidates the fundamental principle of the ICU: a residual structure, where the input feature x is processed in parallel through the main path F_{main} and the residual branch F_{res}. The main path undergoes two sequential stages of "convolution, batch normalization, activation function, and dropout," which collectively enhance feature extraction. The residual branch dynamically selects either an identity mapping or a 1 × 1 convolution for channel adaptation, depending on whether the input and output channel dimensions match. The output features generated by the dual pathways are first aggregated through element-wise summation, after which they undergo nonlinear transformation via the rectified linear unit function to obtain the resultant output (**Equation 2**):

$$ICU(x) = ReLu(F_{main}(x) + F_{res}(x)) \tag{2}$$

This innovative design not only mitigates the vanishing gradient problem but also ensures consistency between encoder and decoder features through a unified processing methodology.

2.4. Transformer module

The bottleneck layer incorporates a global Transformer module (Figure 3) to model long-range spatiotemporal dependencies in seismic profiles. The input feature $x_{enc\text{-}final}$ is flattened into a sequence representation Z, which is then processed by normalization, multi-head self-attention, and a multilayer perceptron, with residual connections integrated after each primary operation. Multi-head self-attention captures long-range spatiotemporal dependencies through similarity computation of query (Q), key (K), and value (V) vectors (**Equations 3–5**):

$$Z_{attn} = softmax(\frac{QK^T}{\sqrt{d_k}})V, \quad d_k = \frac{C}{\mathrm{h}} \tag{3}$$

$$Q = Z \cdot W_Q \tag{4}$$

$$K = Z \cdot W_K \tag{5}$$

where W_Q、W_K、W_V are the projection matrices for queries, keys, and values, respectively; d_k is the dimensionality per attention head, h is the number of attention heads (set to h = 4 in this study), and C corresponds to the number of encoder output channels. Finally, the sequence Z_{mlp} is reconstructed into a 2D feature map $x_{Tr\text{-}out}$, which serves as the initial input to the decoder.

2.5. Spatiotemporal self-adaptive noise enhancement module

The ST-ANEM (Figure 4) is positioned at the network terminus to amplify feature representations along both the spatial (seismic trace) and temporal (time sample) dimensions, thereby enhancing the precision of noise prediction.

Given the input feature $x_{dec\text{-}final}$, the module computes a spatial attention map a_s along the seismic trace dimension (height H) and the temporal attention map a_t along the time sample dimension (width W) (**Equations 6** and **7**):

$$a_s = \frac{1}{H}\sum_{h=1}^{H} x_{h,:} \tag{6}$$

$$a_t = \frac{1}{W}\sum_{w=1}^{W} x_{:,w} \tag{7}$$

where $x_{h,:}$ denotes the feature sequence across all time samples for the h-th seismic trace, while $x_{:,w}$ represents the features of all seismic traces at the w-th time sample.

These features are concatenated and subsequently fused via a 1 × 1 convolution to generate the weight map a, yielding the final output $x_{final} = a \odot x_{dec-final}$. This mechanism adaptively amplifies noise signatures while attenuating effective signal components, thereby enabling the model to distinguish between noise and signal with greater precision. This facilitates a more accurate noise estimation for $n_{out} = f_\theta(Y)$, ultimately allowing for the reconstruction of the effective signal via $X_{denoised} = Y - n_{out}$.

3. Experiments

3.1. Experimental setup

This study benchmarks the efficacy of our approach against five state-of-the-art seismic denoising frameworks—U-Net, Separable-Residual-Dense-Inverted U (SRDU)-Net, Transformer, Noise2Noise, and Noise2Void—by conducting a comparative analysis. Hyperparameter optimization is performed via grid search techniques across models to achieve peak values for metrics including SNR, peak SNR (PSNR), and the structural similarity index (SSIM). This study's experiments were performed utilizing a specialized workstation featuring an Intel Core i7-14700KF central processing unit (operating at 4.88 GHz), paired with 32 gigabytes of high-speed random access memory and an NVIDIA GeForce RTX 4080 SUPER graphical processing unit. Development was facilitated through the PyTorch deep learning ecosystem integrated within the Python programming language environment (3.12.10, Python Software Foundation, Netherlands).

This study conducts an empirical assessment using two types of seismic datasets: synthetically generated noisy data and field-recorded seismic data contaminated by environmental noise. The former comprises four independent subsets, denoted datasets (1)–(4) for reference. Synthetic dataset (1) is a 1D dataset. Its clean

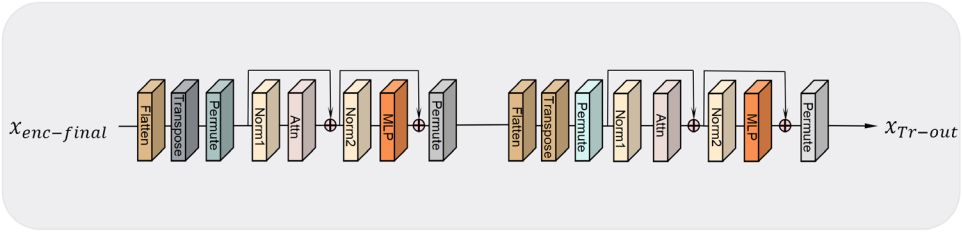


Figure 3. Transformer module
Abbreviation: MLP: Multilayer perceptron.

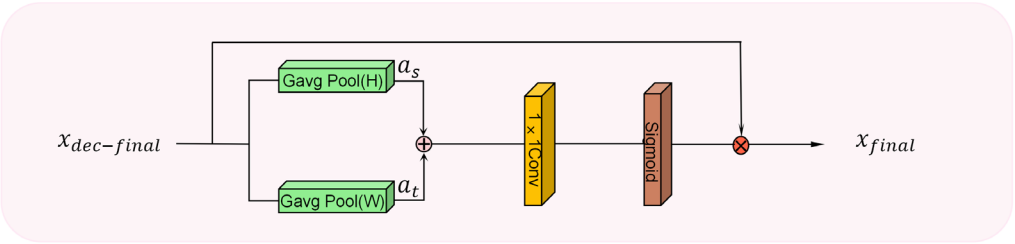


Figure 4. Spatiotemporal self-adaptive noise enhancement module

signals are generated by zero-phase Ricker wavelets with center frequencies randomly drawn from 10 to 40 Hz, sampled at 1 ms, giving 6,000 samples per trace. The noise is taken from real ambient background noise in the publicly available Stanford Earthquake Dataset (STEAD). The two are superimposed to obtain the noisy signals. This dataset is split into training, validation, and test sets at a ratio of 7:2:1, i.e., 7,000 traces for training, 2,000 for validation, and 1,000 for testing, with noise realizations drawn independently for each subset. Synthetic dataset (2) is constructed by adding noise to clean signals that are strictly disjoint from the above 10,000 traces; the noise consists of Gaussian white noise and bandpass noise with standard deviations σ of 0.1, 0.2, and 0.3, giving 3,600 traces in total. This dataset is not used for training, validation, or hyperparameter selection, and serves only as an independent test set. Synthetic dataset (3) contains 5,000 patches of size 16 × 128, generated by superimposing Ricker wavelets with main frequencies of 23–27 Hz onto real noise from the STEAD dataset, and is split into training, validation, and test sets at a ratio of 7:2:1 (3,500/1,000/500 patches). Synthetic dataset (4) contains 4,920 noisy patches of size 16 × 128, generated from clean patches disjoint from those of dataset (3) and contaminated by Gaussian noise and bandpass noise with standard deviations σ of 0.1, 0.2, and 0.3, respectively, serving as an independent test set under multi-noise scenarios. The field dataset comprises 100 real 1D seismic records and one noisy seismic section of 184 traces (each with 500 time samples), both taken from the DiTing330km project released by the National Earthquake Science Data Center.

The decoding blocks employ a dropout rate of 0.1 during training only (Xu *et al.*[52]). For the 1D architecture, the learning rate (LR) is set to 0.0001 with 300 epochs of training; for the 2D network, the LR is set to 0.001 while the number of training epochs remains the same. The above configurations are applied to the corresponding datasets.

To evaluate the seismic signal denoising performance of the model quantitatively, this study adopts three widely used objective evaluation metrics: gauge signal clarity (SNR), measure PSNR, and analyze structural coherence through similarity indexing (SSIM).

3.2. Experimental results

3.2.1. Comparative experiments

A rigorous examination of the developed technique's efficacy in mitigating seismic noise is systematically conducted across heterogeneous data collections integrating simulated scenarios and field-recorded observations. Analysis outcomes are presented through statistical tabulations and visual depictions to elucidate both quantitative metrics and qualitative enhancements, thereby validating the approach's robustness. Quantitative and qualitative results are presented by tables and images, respectively.

Firstly, a comprehensive appraisal of the noise abatement capabilities exhibited by disparate techniques

is executed across the synthetic seismic dataset (1). The analytical outcomes, designated as SignalSet_1D and tabulated in Table 1, reveal the pristine signal's heavy noise contamination, as evidenced by pre-denoising metrics: 4.370 dB for SNR, 18.590 dB for PSNR, and 0.250 for the SSIM. After denoising, the proposed method (TSU-Former) significantly outperforms all competing models, including the self-supervised baselines Noise2Noise (N2N) and Noise2Void (N2V). Specifically, the SNR is increased to 41.710 dB (an absolute improvement of 37.340 dB over the noisy input), the PSNR is improved to 55.920 dB, and the SSIM is enhanced to 0.999. These results indicate that the denoised signals show high structural consistency with the original clean signal and are significantly superior to the other five comparative models.

To compare the denoising performance on 2D data, we conducted experiments on the synthetic seismic dataset (3), with results shown as SignalSet_2D in Table 1. The average SNR, PSNR, and SSIM of the noisy section before denoising are −4.250 dB, 10.250 dB, and 0.122, respectively. After denoising, the proposed method outperforms the baseline models on all three quantitative metrics. In particular, the average SNR is improved from −4.250 dB to 20.510 dB (an absolute improvement of 24.760 dB), and the PSNR is enhanced from 10.250 dB to 34.980 dB. Notably, the SSIM is increased from 0.122 to 0.986. This demonstrates that the proposed method achieves excellent performance in both amplitude fidelity and structural consistency, surpassing not only traditional supervised methods like U-Net but also self-supervised approaches like N2N and N2V.

3.2.2. Experiments with different noise types

To further verify the robustness of the proposed method under different noise types and intensities, we conduct statistics on the average performance on multiple 1D synthetic datasets. Table 2 displays the average quantitative results of different denoising methods on synthetic seismic dataset (2). Whether on data with low SNR or high noise level, the proposed methods outperform comparative methods in metrics of SNR, PSNR, and SSIM, which shows good stability.

To verify the adaptability of the proposed method in complex 2D noisy environments, we further examine its performance in various types and intensities of noise on a synthetic seismic dataset (4). The results are shown in Table 3. The experiments indicate that the proposed method achieves the best quantitative performance under different noise scenarios. This advantage is mainly attributed to the proposed collaborative architecture design of U-Net and

Table 1. Average quantitative results of different denoising methods on synthetic seismic datasets (1) and (3)

Noise dataset	Method	Metrics			Time (s)
		SNR	PSNR	SSIM	
SignalSet_1D	Observed	4.37 ± 0.00	18.59 ± 0.00	0.250 ± 0.00	
	U-Net	34.98 ± 0.10	49.20 ± 0.09	0.997 ± 0.002	499.18
	SRDU-Net	36.53 ± 0.08	50.74 ± 0.08	0.998 ± 0.002	975.60
	N2N	28.75 ± 0.12	43.14 ± 0.10	0.996 ± 0.003	523.10
	N2V	27.90 ± 0.13	42.97 ± 0.13	0.996 ± 0.003	513.70
	Transformer	26.44 ± 0.15	40.65 ± 0.14	0.995 ± 0.004	576.39
	TSU-Former	41.71 ± 0.03	55.92 ± 0.05	0.999 ± 0.001	487.46
SignalSet_2D	Observed	-4.25 ± 0.00	10.25 ± 0.00	0.122 ± 0.00	
	U-Net	10.13 ± 0.10	24.59 ± 0.09	0.872 ± 0.003	178.63
	SRDU-Net	11.15 ± 0.09	25.64 ± 0.09	0.886 ± 0.002	203.70
	N2N	6.69 ± 0.11	19.36 ± 0.12	0.810 ± 0.003	179.46
	N2V	5.81 ± 0.11	18.47 ± 0.13	0.790 ± 0.003	179.69
	Transformer	3.17 ± 0.13	17.93 ± 0.14	0.750 ± 0.003	180.36
	TSU-Former	20.51 ± 0.04	34.98 ± 0.04	0.986 ± 0.001	172.08

Abbreviations: N2N: Noise2Noise; N2V: Noise2Void; PSNR: Peak signal-to-noise ratio; SNR: Signal-to-noise ratio; SRDU-Net: Separable-Residual-Dense-Inverted U-Net; SSIM: Structural similarity index.

Table 2. The average quantitative results of different denoising methods on synthetic seismic dataset (2)

Noise dataset	Method	Gaussian (σ = 0.1)			Gaussian (σ = 0.2)			Gaussian (σ = 0.3)			Bandpass (σ = 0.1)			Bandpass (σ = 0.2)			Bandpass (σ = 0.3)			Time (s)
		SNR	PSNR	SSIM	SNR	PSNR	SSIM	SNR	PSNR	SSIM	SNR	PSNR	SSIM	SNR	PSNR	SSIM	SNR	PSNR	SSIM	
SignalSet_1D (1)	Observed	−3.62	20.75	0.316	−9.84	15.54	0.113	−13.24	13.09	0.055	4.78	28.68	0.907	−1.38	22.81	0.747	−4.88	19.38	0.605	
	U-Net	25.88	49.12	0.999	23.12	46.97	0.999	21.28	45.10	0.999	26.30	49.66	0.999	24.25	47.80	0.999	22.04	45.74	0.992	0.024
	SRDU-Net	25.74	48.85	0.999	22.96	46.48	0.999	20.31	43.90	0.998	25.74	48.97	0.999	24.50	48.04	0.999	22.26	45.87	0.999	0.003
	N2N	18.63	42.47	0.997	16.84	41.16	0.997	13.68	39.59	0.996	18.71	42.48	0.997	18.57	41.97	0.997	17.82	41.31	0.996	0.016
	N2V	16.94	40.85	0..997	14.63	39.69	0.997	11.47	37.61	0.996	16.97	40.86	0.997	16.86	40.65	0.997	15.92	40.16	0.996	0.019
	Transformer	14.26	37.70	0.996	11.49	35.31	0.993	9.20	32.98	0.991	15.68	39.00	0.998	15.45	38.97	0.997	14.07	37.73	0.997	0.004
	TSU-Former	34.08	56.40	0.999	28.82	52.01	0.999	25.07	48.47	0.999	33.85	56.44	0.999	29.62	52.56	0.999	27.88	50.93	0.999	0.002
SignalSet_1D (2)	Observed	−3.96	20.64	0.316	−9.79	15.69	0.112	−13.48	12.92	0.054	4.72	28.74	0.907	−1.34	22.73	0.747	−4.77	19.42	0.605	
	U-Net	25.52	49.14	0.999	23.19	46.95	0.999	21.45	45.53	0.999	26.59	49.53	0.999	24.06	47.55	0.999	22.21	45.83	0.999	0.002
	SRDU-Net	25.73	49.18	0.999	23.46	46.95	0.999	20.67	44.45	0.998	26.28	49.58	0.999	24.30	47.85	0.999	22.47	46.02	0.999	0.003
	N2N	18.59	42.48	0.997	16.86	41.15	0.997	13.69	39.61	0.996	18.70	42.46	0.997	18.53	41.95	0.997	17.83	41.33	0.996	0.003
	N2V	16.91	40.86	0.997	14.64	39.69	0.997	11.49	37.62	0.996	16.96	40.85	0.997	16.85	40.64	0.997	15.93	40.17	0.996	0.003
	Transformer	14.34	38.06	0.996	11.48	35.28	0.993	9.31	33.32	0.991	15.88	39.32	0.998	14.97	38.54	0.997	14.47	37.99	0.997	0.003
	TSU-Former	33.71	56.33	0.999	29.52	52.60	0.999	25.23	48.83	0.999	33.43	55.78	0.999	30.35	53.31	0.999	27.66	50.85	0.999	0.001
SignalSet_1D (3)	Observed	−3.91	20.65	0.316	−9.92	15.55	0.113	−13.43	13.00	0.055	4.71	28.70	0.907	−1.22	22.78	0.747	−4.85	19.38	0.606	
	U-Net	25.84	49.44	0.999	23.64	47.54	0.999	21.48	45.44	0.991	26.71	49.73	0.999	24.05	47.45	0.999	22.28	45.99	0.999	0.002
	SRDU-Net	25.97	49.30	0.999	23.48	47.08	0.999	20.59	44.36	0.998	26.30	49.35	0.999	24.17	47.51	0.999	22.49	46.11	0.999	0.003
	N2N	18.60	42.50	0.997	16.90	41.20	0.997	13.70	39.60	0.991	18.73	42.46	0.998	18.53	41.93	0.998	17.84	41.35	0.998	0.003
	N2V	16.92	40.89	0.997	14.71	39.74	0.997	11.51	37.62	0.991	16.97	40.84	0.998	16.84	40.61	0.998	15.95	40.18	0.998	0.003
	Transformer	14.46	38.12	0.996	11.58	35.49	0.993	9.23	33.21	0.991	15.96	39.36	0.998	15.15	38.54	0.997	14.05	37.71	0.997	0.003
	TSU-Former	34.37	56.52	0.999	29.36	52.58	0.999	25.66	49.10	0.999	34.05	56.27	0.999	30.43	53.07	0.999	28.05	51.00	0.999	0.001

Abbreviations: N2N: Noise2Noise; N2V: Noise2Void; PSNR: Peak signal-to-noise ratio; SNR: Signal-to-noise ratio; SRDU-Net: Separable-Residual-Dense-Inverted U-Net; SSIM: Structural similarity index.

doi: 10.36922/JSE026290142

Table 3. Average quantitative results of different denoising methods on synthetic seismic dataset (4)

Noise dataset	Method	Gaussian (σ = 0.1)			Gaussian (σ = 0.2)			Gaussian (σ = 0.3)			Bandpass (σ = 0.1)			Bandpass (σ = 0.2)			Bandpass (σ = 0.3)			Time (s)
		SNR	PSNR	SSIM	SNR	PSNR	SSIM	SNR	PSNR	SSIM	SNR	PSNR	SSIM	SNR	PSNR	SSIM	SNR	PSNR	SSIM	
SignalSet_2D (1)	Observed	−6.69	19.99	0.38	0.69	13.98	0.23	−2.83	10.46	0.17	14.59	27.88	0.64	6.41	21.86	0.37	0.05	18.37	0.20	
	U-Net	11.50	24.81	0.86	11.66	24.94	0.85	11.32	24.61	0.83	12.51	25.81	0.84	10.87	26.30	0.82	8.63	26.89	0.78	3.71
	SRDU-Net	14.44	27.73	0.95	14.04	27.30	0.92	13.31	26.57	0.88	14.90	28.19	0.95	12.75	28.20	0.93	10.02	28.33	0.90	4.50
	N2N	8.85	20.62	0.81	8.56	20.67	0.81	8.84	20.60	0.79	8.87	20.63	0.80	8.84	20.66	0.80	6.41	20.66	0.79	4.72
	N2V	6.79	19.35	0.79	6.81	19.36	0.79	6.77	19.34	0.79	6.81	19.37	0.80	6.78	19.38	0.80	5.82	19.40	0.79	4.67
	Transformer	4.27	17.24	0.69	4.05	17.03	0.68	3.96	16.95	0.68	4.20	17.17	0.69	3.59	18.81	0.76	2.37	20.62	0.85	4.76
	TSU-Former	20.82	34.09	0.98	19.97	33.23	0.98	19.02	32.28	0.98	20.19	33.44	0.999	20.43	35.82	0.99	22.94	41.18	0.99	3.61
SignalSet_2D (2)	Observed	6.68	19.97	0.38	0.70	13.98	0.23	−2.83	10.47	0.17	14.63	27.91	0.67	6.46	21.89	0.39	0.39	18.35	0.21	
	U-Net	12.15	25.44	0.86	11.77	25.04	0.85	11.42	24.72	0.82	11.21	24.50	0.84	10.21	25.62	0.82	8.98	26.88	0.77	2.75
	SRDU-Net	14.94	28.22	0.95	14.25	27.50	0.92	13.59	26.86	0.88	14.31	27.59	0.94	12.09	27.52	0.92	10.11	28.06	0.90	3.44
	N2N	8.86	20.64	0.80	8.56	20.69	0.81	8.84	20.61	0.79	8.86	20.58	0.79	8.81	20.70	0.81	6.44	20.66	0.79	3.39
	N2V	6.81	19.36	0.79	6.82	19.37	0.80	6.78	19.37	0.78	6.75	19.31	0.79	6.73	19.40	0.79	5.83	19.40	0.79	3.40
	Transformer	4.26	17.22	0.69	3.99	16.96	0.67	4.06	17.04	0.68	4.12	17.08	0.70	3.52	18.75	0.77	2.69	20.55	0.84	3.41
	TSU-Former	21.04	34.29	0.98	20.01	33.26	0.98	19.01	32.27	0.98	20.35	35.59	0.98	19.70	35.08	0.99	20.87	38.76	0.99	2.59
SignalSet_2D (3)	Observed	6.72	20.01	0.38	0.67	13.96	0.24	−2.84	10.46	0.17	14.62	27.92	0.68	20.58	21.84	0.35	−0.13	18.30	0.21	
	U-Net	11.60	24.88	0.86	11.70	24.99	0.86	10.99	24.29	0.82	11.50	24.79	0.84	10.17	26.39	0.82	8.46	26.83	0.78	2.60
	SRDU-Net	14.33	27.60	0.94	13.96	27.21	0.92	13.22	26.49	0.88	14.36	27.66	0.94	12.01	28.25	0.92	9.72	28.15	0.90	3.34
	N2N	8.87	20.63	0.81	8.87	20.65	0.82	8.79	20.07	0.77	8.86	20.61	0.81	7.83	18.41	0.71	7.14	21.32	0.80	3.39
	N2V	6.77	19.35	0.79	6.78	19.36	0.79	6.14	19.24	0.76	6.75	19.35	0.79	5.27	16.38	0.69	4.32	19.60	0.79	3.41
	Transformer	4.03	17.01	0.69	4.13	17.10	0.69	4.06	17.04	0.68	4.14	17.11	0.69	3.22	19.28	0.79	2.48	20.85	0.86	3.44
	TSU-Former	20.50	33.75	0.98	20.40	33.66	0.98	19.02	32.28	0.98	20.34	33.59	0.98	20.53	36.72	0.99	20.75	39.12	0.99	2.55

Abbreviations: N2N: Noise2Noise; N2V: Noise2Void; PSNR: Peak signal-to-noise ratio; SNR: Signal-to-noise ratio; SRDU-Net: Separable-Residual-Dense-Inverted U-Net; SSIM: Structural similarity index.

Transformer: U-Net can capture the local detailed features of seismic signals effectively, while Transformer enhances the model's capability to capture long-range spatiotemporal dependencies. By introducing a spatiotemporal attention mechanism, the model can adaptively focus on critical noisy regions, better preserving the integrity and continuity of geological structures, thereby demonstrating excellent denoising performance and robustness under multiple noise conditions.

Based on the quantitative analysis, to further evaluate the denoising performance intuitively, we compare the denoising visualization results of various methods, shown in Figures 5 and 6.

Among them, Figure 5 shows the denoising performance of each model on 1D single-channel seismic signals in the synthetic seismic dataset (2), which clearly shows the differences in detailed processing of each method: The self-supervised methods, N2V and N2N, preserve the main signal trend but show limited capability in suppressing background noise fluctuations. The denoising results of U-Net and SRDU-Net still show obvious noise fluctuations and fail to achieve complete noise suppression; Transformer causes significant signal distortion, and key details of the original signal are lost to varying degrees. In contrast, the proposed method demonstrates superior comprehensive performance—while efficiently filtering out Gaussian noise, it reproduces the waveform features of the original signal accurately, achieving the best performance in preserving signal details.

Figure 6 presents an evaluation of denoising performance on the 2D seismic section from synthetic dataset (4), integrating seismic profiles, f–k spectra, estimated noise maps, difference sections, and local similarity distributions. The denoising results indicate that while U-Net and SRDU-Net produce smoother sections, they tend to over-smooth the data and attenuate

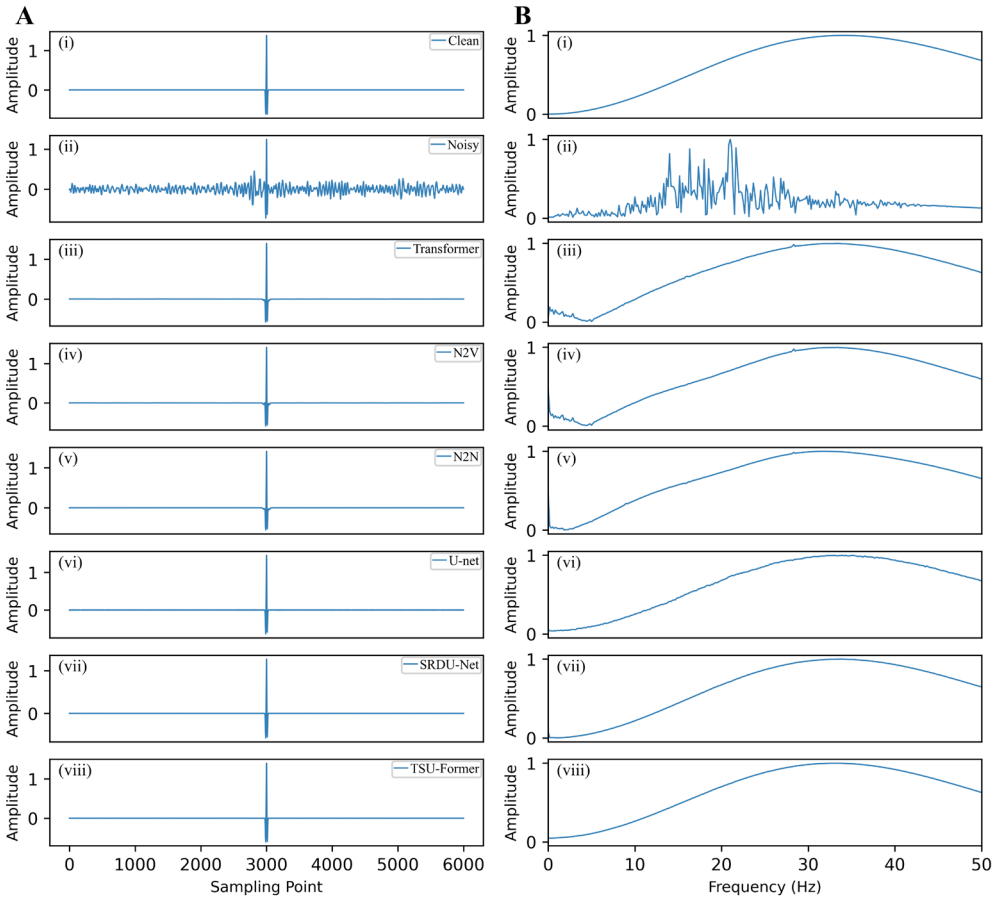


Figure 5. Comparison of denoising performance of different methods on the synthetic dataset (2). (A) The seismic signals in the time domain and (B) their corresponding frequency spectra. The subplots represent: (i) the clean signal and its spectrum; (ii) the noisy signal with Gaussian noise $\sigma = 0.1$ and its spectrum; (iii) the result of the Transformer method; (iv) the result of N2V; (v) the result of N2N; (vi) the result of U-Net; (vii) the result of SRDU-Net; and (viii) the result of the proposed TSU-Former.
Abbreviations: N2N: Noise2Noise; N2V: Noise2Void; SRDU-Net: Separable-Residual-Dense-Inverted U-Net.

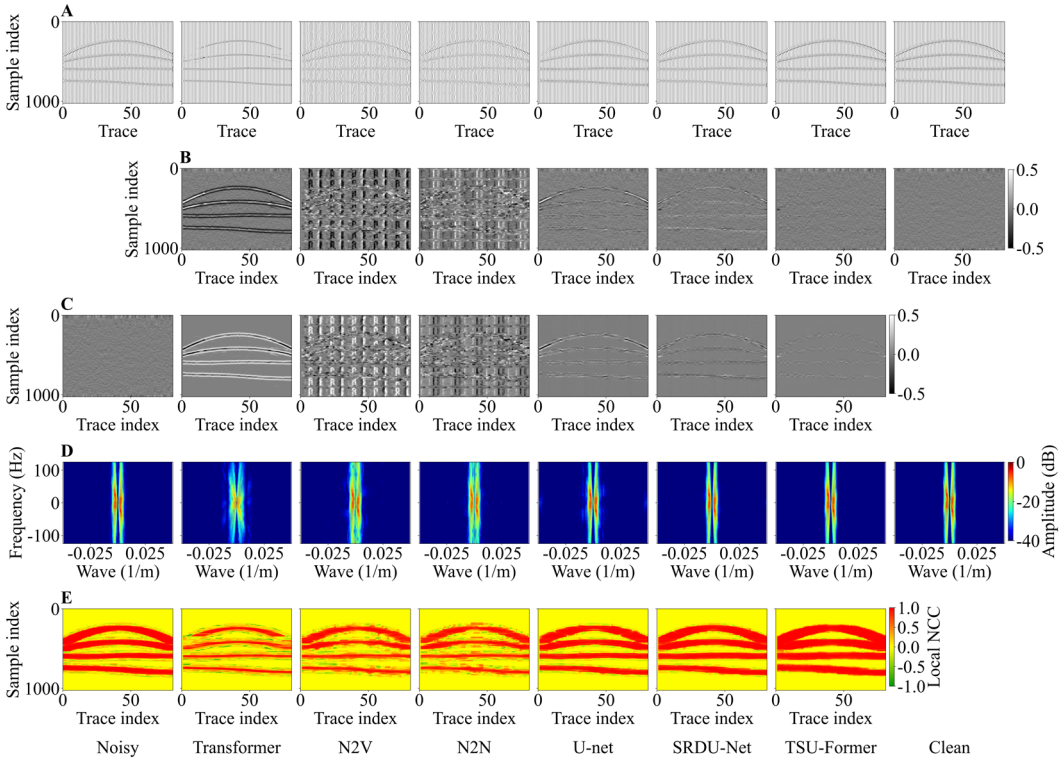


Figure 6. Results of different methods under bandpass noise (σ = 0.1). (A) Noise attenuation results of different methods under bandpass noise (σ = 0.1). (B) Estimated noise (noisy input minus denoised output). (C) Difference section between the denoised data and the clean data. (D) f–k spectra. (E) Local similarity maps (denoised vs. clean).
Abbreviations: N2N: Noise2Noise; N2V: Noise2Void; SRDU-Net: Separable-Residual-Dense-Inverted U-Net.

high-frequency details. Similarly, the Transformer model introduces local artifacts and does not fully suppress background noise, whereas the self-supervised methods N2V and N2N leave residual textures that may obscure weak signals. These observations are supported by the estimated noise maps and difference sections: the noise maps for Transformer, N2V, and N2N contain noticeable coherent structural events rather than purely random noise, suggesting partial signal leakage or incomplete signal–noise separation. In comparison, the noise map of the proposed TSU-Former contains fewer coherent structures, indicating relatively better preservation of the underlying signal. This trend is further reflected in the local similarity distributions and f–k spectra, where the proposed method shows relatively higher similarity and more concentrated energy distribution, suggesting that it can mitigate bandpass noise interference while maintaining the structural fidelity of key geological features.

To further evaluate the denoising effectiveness, we select the 40th seismic trace for single-trace comparison. Figure 7 presents the analysis results across three rows. In the time-domain plots (Figure 7A), the Transformer and self-supervised methods N2V and N2N exhibit noticeable residual fluctuations or waveform distortions compared to the clean signal. The U-Net and SRDU-Net produce smoother waveforms but show slight deviations in peak amplitudes. In contrast, the proposed TSU-Former yields a waveform that aligns more closely with the original trace, effectively preserving the signal's dynamic range. This observation is supported by the amplitude spectra (Figure 7B), where the TSU-Former's result maintains a spectral distribution consistent with the clean data, avoiding the high-frequency energy loss seen in other methods. Furthermore, the phase spectra (Figure 7C) demonstrate that the proposed method preserves the phase characteristics of the effective signal components more accurately than the comparative models, indicating relatively better fidelity in signal reconstruction.

To verify the practicality of the proposed method, we test a field-noisy seismic dataset; the results are shown in Table 4, Figures 8 and 9.

Table 4 presents the average quantitative results on 1D field data. The proposed TSU-Former achieves an average SNR of 27.48 dB, representing a significant improvement over the original noisy signal (13.87 dB) with a relative gain of approximately 98.1%. When compared with other

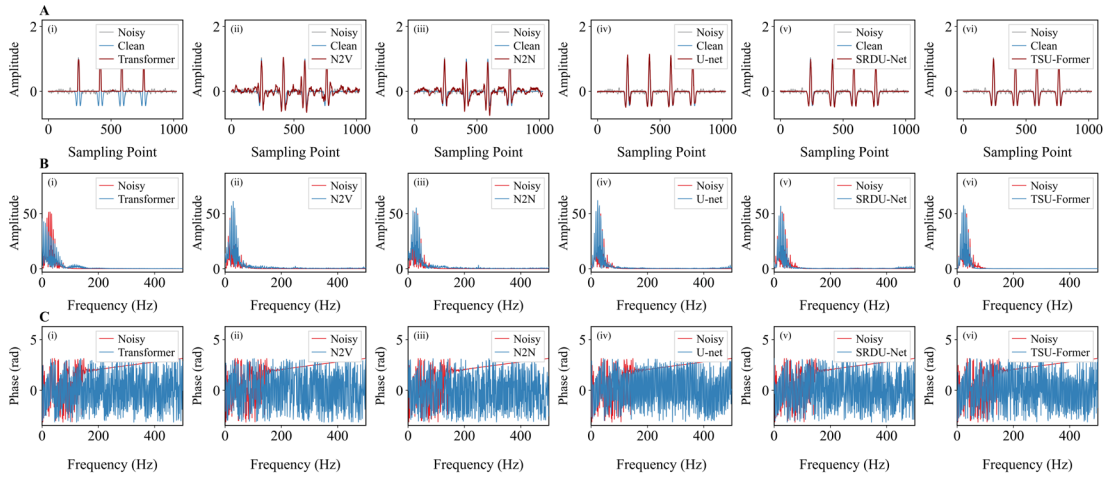


Figure 7. The results of processing the 40th trace line using six different methods. (A) The denoising results of the 40th trace time-domain data after processing with six methods for 2D synthetic seismic records. (B) Frequency analysis. (C) Phase analysis.
Abbreviations: N2N: Noise2Noise; N2V: Noise2Void; SRDU-Net: Separable-Residual-Dense-Inverted U-Net.

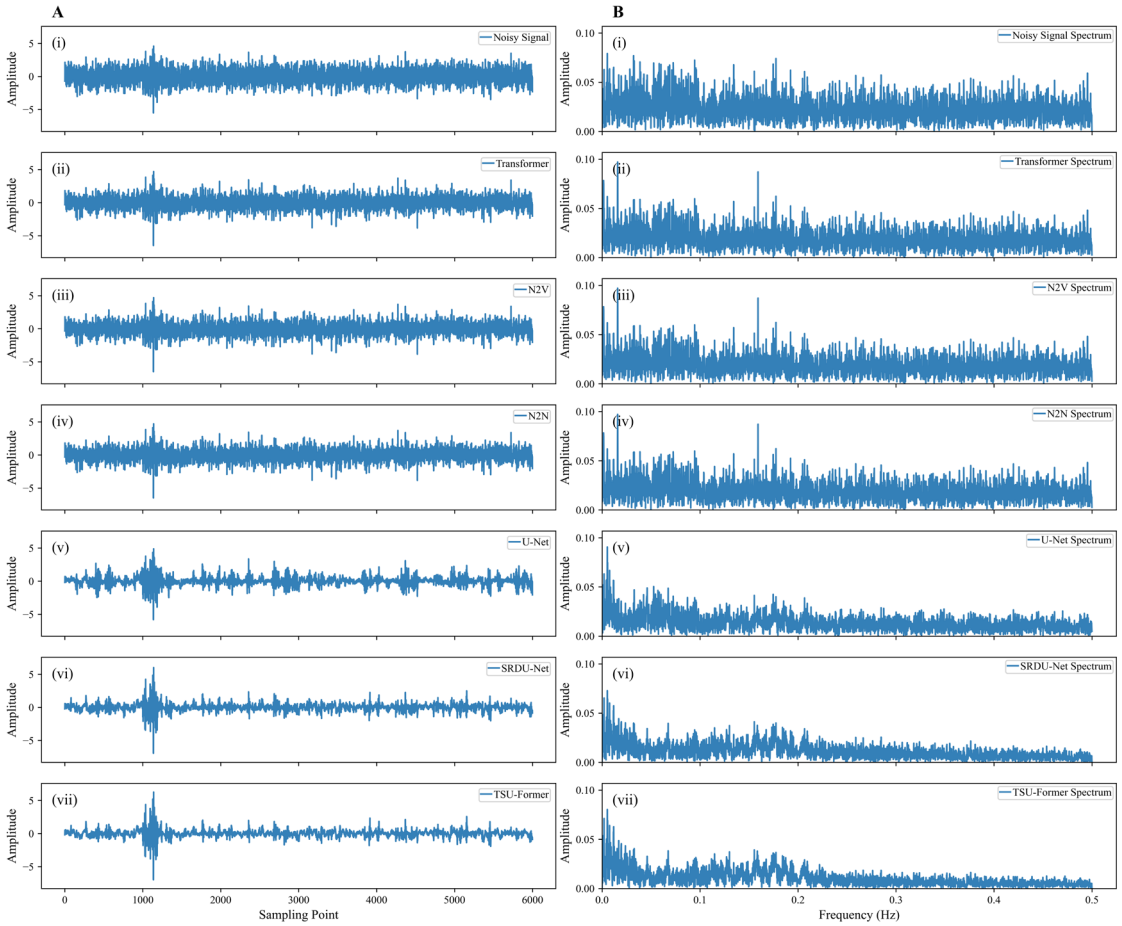


Figure 8. Denoising results of different methods on one-dimensional field seismic data. (A) Time-domain plot. (B) Frequency-domain plot. (i) Field data and corresponding spectrum. (ii) Transformer denoising result and corresponding spectrum. (iii) N2V denoising result and corresponding spectrum. (iv) N2N denoising result and corresponding spectrum. (v) U-Net denoising result and corresponding spectrum. (vi) SRDU-Net denoising result and corresponding spectrum. (vii) TSU-Former denoising result and corresponding spectrum.
Abbreviations: N2N: Noise2Noise; N2V: Noise2Void; SRDU-Net: Separable-Residual-Dense-Inverted U-Net.

Table 4. Average quantitative results of different denoising methods on one-dimensional field data

Original signal	U-Net	SRDU-Net	Transformer	N2N	N2V	TSU-Former
13.87 dB	21.98 dB	21.07 dB	15.67 dB	17.13 dB	16.48 dB	27.48 dB

Abbreviations: N2N: Noise2Noise; N2V: Noise2Void; SRDU-Net: Separable-Residual-Dense-Inverted U-Net.

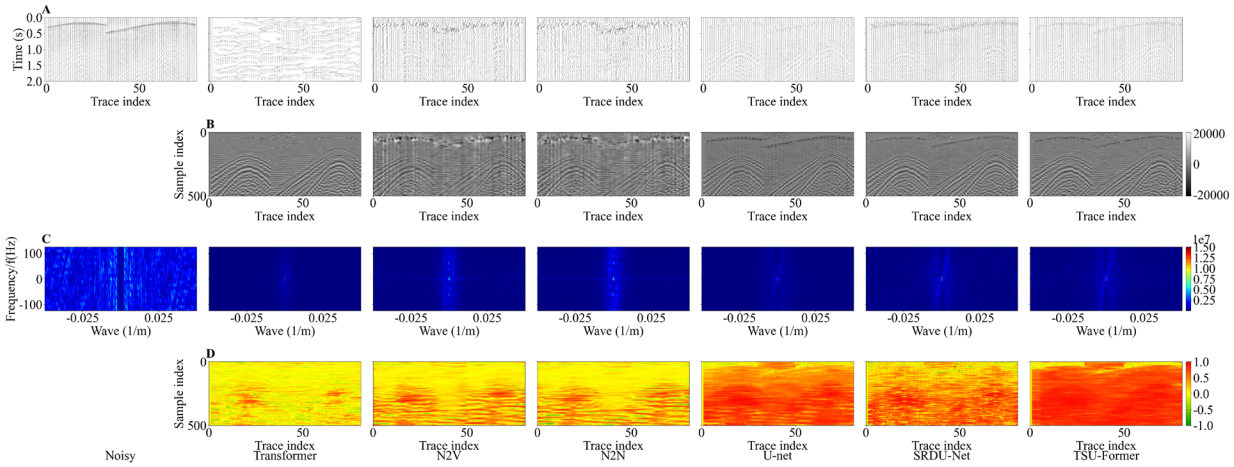


Figure 9. Results of noise attenuation in two-dimensional field seismic data using different methods. (A) Noise attenuation results by different methods for two-dimensional field seismic data. (B) Estimated noise (noisy input minus denoised output). (C) f–k spectra. (D) Local similarity maps (denoised vs. noisy).
Abbreviations: N2N: Noise2Noise; N2V: Noise2Void; SRDU-Net: Separable-Residual-Dense-Inverted U-Net.

state-of-the-art methods, TSU-Former outperforms U-Net and SRDU-Net by 5.50 dB and 6.41 dB, respectively. Furthermore, it demonstrates superior performance against the Transformer and self-supervised methods N2N and N2V, exceeding them by margins of 11.81 dB, 10.35 dB, and 11.00 dB, respectively, which further proves its superior performance under real, complex noise environments.

Figure 8 compares the denoising results of various methods on 1D single-channel field seismic data. The raw noisy signal is characterized by strong random background noise with weak, irregular events. The transformer module effectively removes random noise while retaining the primary signal morphology. In contrast, N2V and N2N exhibit limitations in suppressing background noise, leaving substantial irregular noise components. The U-Net module provides a cleaner background but over-smooths the waveform, leading to a loss of fine structural details. Similarly, the SRDU-Net module generates a smooth signal but fails to preserve subtle signal variations. Ultimately, the proposed TSU-Former module achieves the optimal balance, effectively suppressing the random noise while accurately recovering the effective signal features with minimal waveform distortion.

Figure 9 illustrates the denoising performance of different methods on the 2D field seismic section. The original noisy section is characterized by strong background interference that obscures the continuity of effective signals. U-Net and SRDU-Net tend to over-smooth the data, leading to a loss of effective structural details; the Transformer model leaves residual noise artifacts in certain areas; and the self-supervised methods, N2V and N2N, exhibit noticeable residual textures that may degrade the clarity of weak reflections. In contrast, the proposed TSU-Former yields a cleaner section, effectively preserving reflection events and enhancing signal continuity. These observations are further supported by the estimated noise maps. The noise maps for U-Net, SRDU-Net, Transformer, as well as N2V and N2N, contain varying degrees of coherent structural events rather than pure random noise, suggesting that effective geological features might be removed along with the noise. Conversely, the noise map of the proposed method appears predominantly as random noise with minimal structural information, indicating a more precise separation of signal and noise. This effectiveness is corroborated by the f–k spectra, where the proposed method effectively isolates noise energy without significantly distorting valid frequency-wavenumber components. Finally, the local similarity maps provide quantitative evidence: compared to all other models, the

proposed method achieves relatively higher similarity scores across the section, suggesting its robust capability to maintain structural fidelity while mitigating complex field noise.

Figure 10 presents the denoised profiles and zoom-in views of different methods on the 2D field seismic data. The original noisy profile is characterized by strong background interference that obscures the continuity of reflection events. Analysis of the zoom-in views reveals that the transformer model produces blocky artifacts that distort seismic events after denoising. The U-Net and SRDU-Net methods, while suppressing noise, tend to blur reflection events, potentially leading to a loss of effective structural details. The self-supervised methods, N2V and N2N, exhibit noticeable residual noise textures and may not fully restore weak signals. In contrast, the proposed TSU-Former achieves a balance between noise suppression and signal preservation. As highlighted in the zoom-in views, TSU-Former effectively mitigates complex background noise while preserving the continuity and amplitude of reflection events.

To further evaluate the denoising effectiveness on real-world data, we extract the 40th seismic trace from the 2D field seismic dataset for a detailed single-trace comparison, as shown in Figure 11. In the time-domain plots (Figure 11A), the denoising results of the 40th trace reveal distinct performance differences among the six methods. While the comparative models are able to

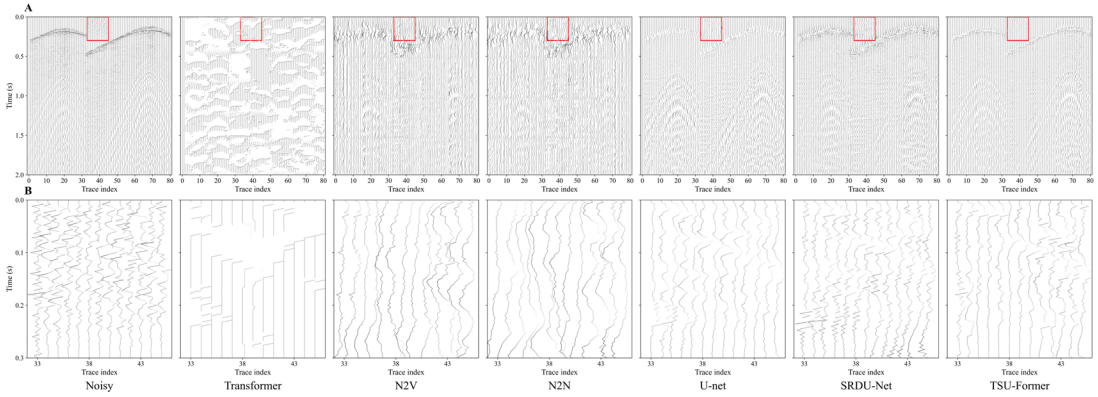


Figure 10. Processing results of two-dimensional field seismic data using different methods. (A) Noise attenuation results by different methods of two-dimensional field seismic data. (B) Zoom-in views corresponding to the red rectangles in the first row.
Abbreviations: N2N: Noise2Noise; N2V: Noise2Void; SRDU-Net: Separable-Residual-Dense-Inverted U-Net.

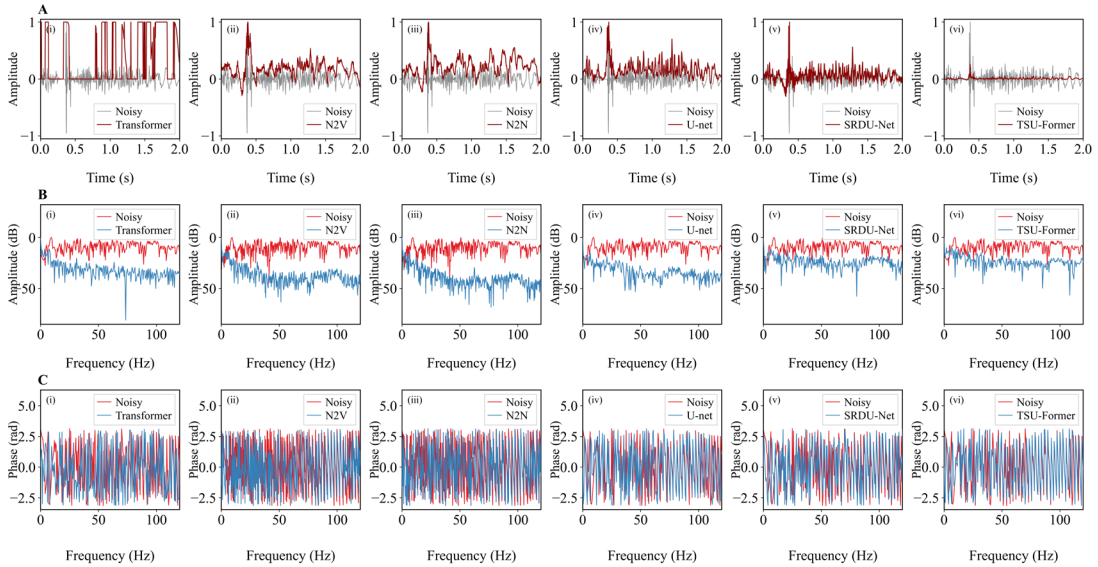


Figure 11. Results of trace 40 after processing two-dimensional field seismic data using six different methods. (A) The denoising results of the 40th trace time-domain data after processing with six methods for two-dimensional field seismic data. (B) Frequency analysis. (C) Phase analysis.
Abbreviations: N2N: Noise2Noise; N2V: Noise2Void; SRDU-Net: Separable-Residual-Dense-Inverted U-Net.

suppress some noise, they tend to either leave noticeable residual fluctuations or over-smooth the signal, leading to waveform distortions. In contrast, the proposed method yields a cleaner waveform that aligns more closely with the expected signal characteristics, effectively preserving the detailed structures and dynamic range of the original trace. Furthermore, the phase analysis (Figure 11C) indicates that the proposed method preserves the phase characteristics of the effective signal components more accurately than the comparative models. This suggests relatively better fidelity in signal reconstruction.

3.2.3. Out-of-distribution experiment

To further evaluate the denoising performance of TSU-Former under field conditions different from those used for model training, an additional independent experiment was conducted on the Texas Earthquake Dataset for AI (TXED) dataset.[53] TXED contains seismic observations collected from 321 stations across Texas between 2017 and 2023, covering different recording locations, instruments, acquisition conditions, and noise environments. In this experiment, 3,600 noisy waveform records were directly selected from TXED as an independent test set, and the trained model was applied without retraining or fine-tuning.

These field records retain complex real-world noise characteristics, including coherent components, impulsive disturbances, harmonic interference, strong non-stationarity, and mixtures of multiple noise components. Therefore, the TXED experiment provides an additional evaluation of TSU-Former under more diverse acquisition conditions and more complex noise environments than those used during training.

The quantitative results are presented in Table 5. The original noisy data have an SNR of 5.580 dB, a PSNR of 17.630 dB, and an SSIM of 0.728. After denoising, TSU-Former increases these values to 21.340 dB, 33.280 dB, and 0.978, respectively, corresponding to improvements of 15.760 dB in SNR, 15.650 dB in PSNR, and 0.250 in SSIM. Among the comparison methods, SRDU-Net achieves an SNR of 17.260 dB, a PSNR of 30.680 dB, and an SSIM of 0.937. Compared with SRDU-Net, TSU-Former further improves the three metrics by 4.080 dB, 2.600 dB, and 0.041, respectively. U-Net, Transformer, N2N, and N2V obtain SNR values of 15.560, 7.420, 12.610, and 12.260 dB, respectively, all lower than those obtained by TSU-Former. The quantitative comparison indicates that TSU-Former maintains effective noise suppression and signal reconstruction performance when tested with real environmental noise from a data source different from that used in the original noise experiments.

For qualitative comparison, two waveform samples were randomly selected from the TXED-based test set, and their time-domain waveforms and corresponding spectra are shown in Figures 12 and 13. In both examples, the noisy signals contain evident fluctuations over a broad temporal range, while their spectral energy is distributed across multiple frequency components. After processing by the comparison methods, different degrees of residual interference remain. Transformer, N2V, N2N, and U-Net reduce part of the noise energy, but noticeable fluctuations can still be observed in the reconstructed waveforms and corresponding spectra. SRDU-Net provides further suppression of several noise components, although residual spectral energy remains visible.

In comparison, TSU-Former produces smoother reconstructed waveforms while retaining the principal temporal variation of the signals. In the frequency domain, the background spectral energy is further reduced, and the dominant frequency components are more clearly

Table 5. Comparison of performance metrics across different methods on the Texas Earthquake Dataset for AI data

Method	SNR	PSNR	SSIM
Noisy	5.580	17.630	0.728
U-Net	15.560	28.740	0.924
SRDU-Net	17.260	30.680	0.937
Transformer	7.420	20.520	0.885
N2N	12.610	26.880	0.918
N2V	12.260	26.190	0.912
TSU-Former	21.340	33.280	0.978

Abbreviations: N2N: Noise2Noise; N2V: Noise2Void; PSNR: Peak signal-to-noise ratio; SNR: Signal-to-noise ratio; SRDU-Net: Separable-Residual-Dense-Inverted U-Net; SSIM: Structural similarity index.

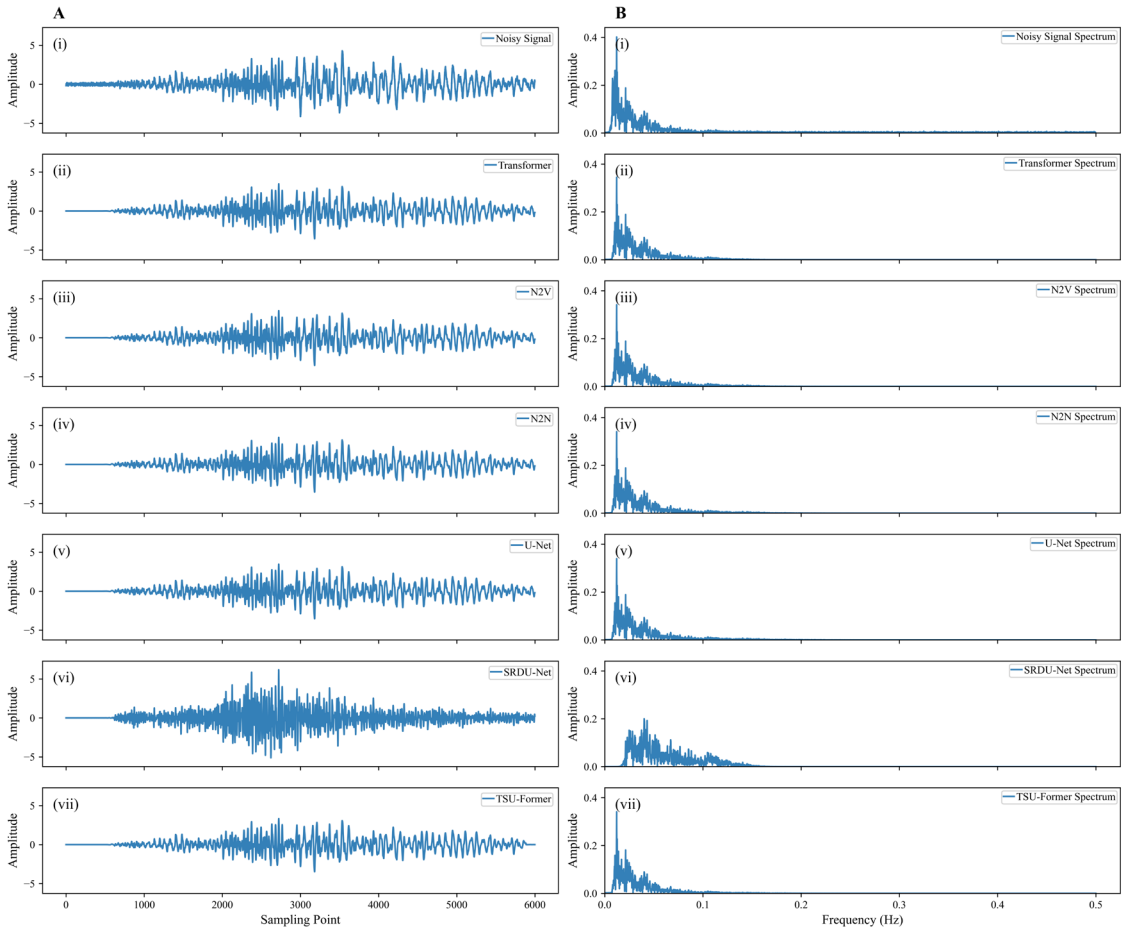


Figure 12. Results comparison of different denoising methods on seismic signals from the Texas Earthquake Dataset for the AI dataset. (A) Time-domain denoised signals. (B) Spectral results of denoised signals. In each subfigure: (i) Noisy signal; (ii) Transformer; (iii) N2V; (iv) N2N; (v) U-Net; (vi) SRDU-Net; (vii) TSU-Former.
Abbreviations: N2N: Noise2Noise; N2V: Noise2Void; SRDU-Net: Separable-Residual-Dense-Inverted U-Net.

preserved. Similar behavior can be observed in both randomly selected examples, which is consistent with the quantitative results in Table 5.

Taken together, the quantitative metrics and the two representative waveform–spectrum examples show that TSU-Former retains effective denoising performance when the noise source is changed from the original synthetic or STEAD-based conditions to real seismic data extracted from TXED. These results provide an additional evaluation of the model under a more complex and previously unseen noise distribution.

4. Discussion

Herein, we delve into the contributions of diverse architectural components and hyperparameter configurations within the TSU-Former framework underpinning spatiotemporal attention mechanisms, thereby deepening insights into the fundamental tenets and empirical efficacy of the advanced approach.

4.1. Ablation experiments

4.1.1. Influence of transformer module

The Transformer module is a key component of the proposed method, since it enhances the in-domain reconstruction performance of denoising by modeling the long-range spatiotemporal dependencies of seismic signals. As shown in Table 6, compared with the TSU-Former model, the ablation model without the Transformer exhibits a decline of 1.79 dB, 1.77 dB, and 0.02 in SNR, PSNR, and SSIM on the test dataset, respectively. This demonstrates the independent contribution of the Transformer module to the overall denoising capability.

4.1.2. Influence of spatiotemporal attention mechanism module

The spatiotemporal attention module is also a key

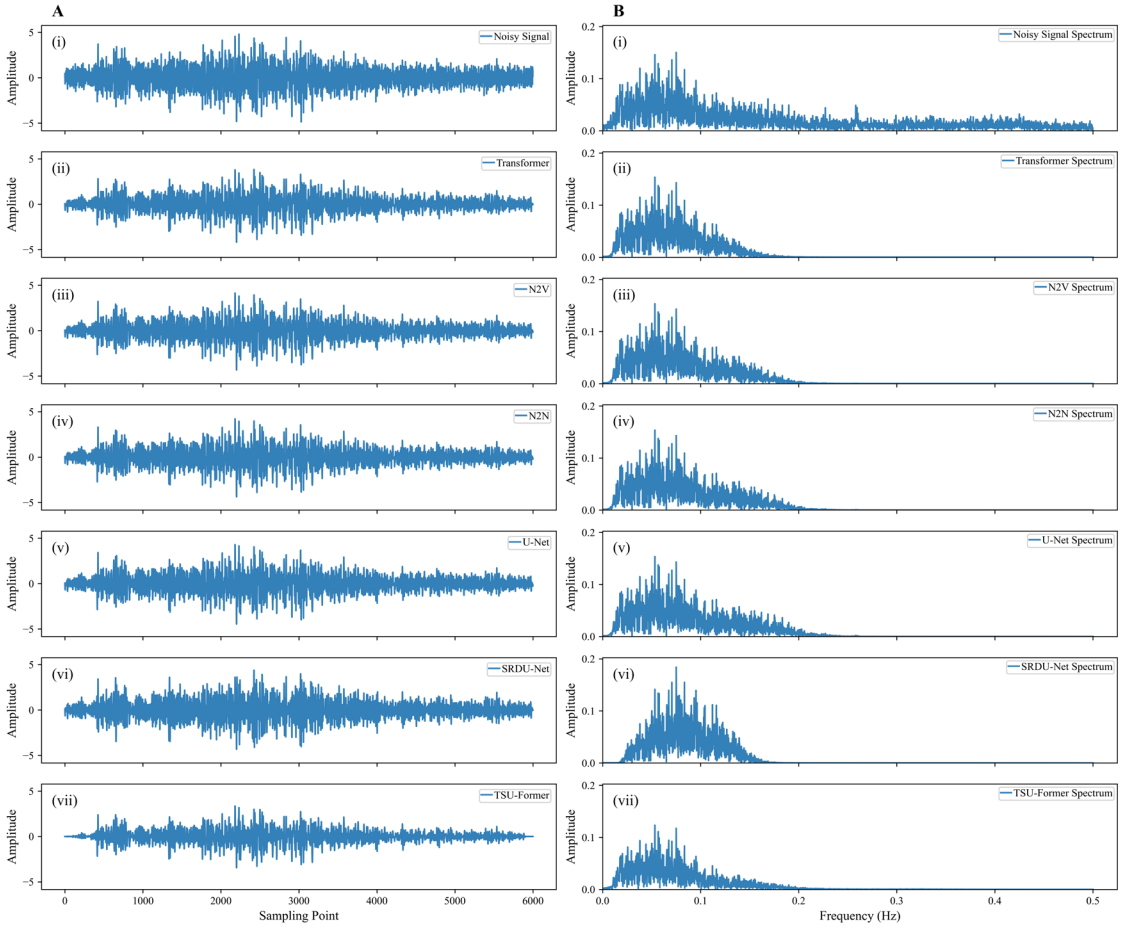


Figure 13. Results comparison of different denoising methods on seismic signals from the Texas Earthquake Dataset for the AI dataset. (A) Time-domain denoised signals. (B) Spectral results of denoised signals. In each subfigure: (i) noisy signal; (ii) Transformer; (iii) N2V; (iv) N2N; (v) U-Net; (vi) SRDU-Net; (vii) TSU-Former.
Abbreviations: N2N: Noise2Noise; N2V: Noise2Void; SRDU-Net: Separable-Residual-Dense-Inverted U-Net.

component of this method, as it enhances the model's perceptual capability regarding seismic signal continuity and structural features by jointly modeling dependencies across space and time. As shown in Table 6, compared with the baseline model (TSU-Former), the ablation model without the spatiotemporal attention mechanism (no ST-ANEM) exhibits a decrease of 2.46 dB, 2.45 dB, and 0.03 in average SNR, PSNR, and SSIM, respectively.

4.1.3. Influence of improved convolutional unit module

The ICU module is also a key component of the proposed method, as it helps extract clean signal features from seismic data, thereby enhancing the model's representation capability. The results shown in Table 6 compare with the baseline model; the ablation model without the ICU module has a decrease of 1.63dB, 1.56dB and 0.02 in average SNR, PSNR, and SSIM.

4.2. Influence of hyperparameters

This subsection systematically assesses the influence of LR and training epochs on the denoising performance of 2D seismic sections and 1D seismic traces. The optimal parameter configurations for dimension-wise denoising were determined via comparative evaluation of metrics assessing signal fidelity, including the power ratios of desired SNR, peak-based assessments (PSNR), and structural coherence measurements (SSIM).

For 2D seismic sections containing complex spatio-temporal structures, experimental results presented in Table 7 indicate that LR = 0.001 combined with 300 epochs yields the optimal performance. Under this configuration, the model achieves peak values across all evaluation metrics (SNR: 15.78 dB, PSNR: 29.82 dB, SSIM: 0.95) while maintaining stable convergence characteristics. Adhering strictly to the parameter optimization methodology

established for 2D data, equivalent tuning was performed for 1D seismic data. As evidenced in Table 7, the configuration of LR = 0.0001 with 300 epochs demonstrates superior performance across all three quantitative metrics.

Figures 14 and 15 respectively illustrate the progressive enhancement of model performance with increasing training epochs under the aforementioned optimal parameter configurations.

Table 6. Ablation experiment results of the TSU-Former model

Model configuration	SNR (dB)	PSNR (dB)	SSIM
TSU-Former	17.52	31.87	0.98
No ST-ANEM	15.06	29.42	0.95
No transformer	15.73	30.10	0.96
No ICU	15.89	30.31	0.96
U-Net	13.58	26.11	0.92

Abbreviations: ICU: Improved convolutional unit; PSNR: Peak signal-to-noise ratio; SNR: Signal-to-noise ratio; SSIM: Structural similarity index; ST-ANEM: Spatiotemporal self-adaptive noise enhancement module.

Table 7. Comparison of denoising performance in two/one-dimensional seismic data under different learning rates and training epochs

LR	Epoch	SNR	PSNR	SSIM
0.01	100	9.41/15.67	23.36/29.88	0.58/0.9906
	200	9.69/20.44	23.67/34.65	0.69/0.9897
	300	9.99/20.20	24.01/34.42	0.82/0.9893
0.001	100	13.61/22.28	27.63/36.49	0.85/0.9928
	200	15.40/22.19	29.42/36.41	0.91/0.9926
	300	15.78/22.10	29.82/36.32	0.95/0.9925
0.0001	100	11.16/22.23	25.17/36.44	0.68/0.9928
	200	13.22/22.13	27.26/36.34	0.81/0.9924
	300	14.45/22.38	28.48/36.60	0.87/0.9928
0.00001	100	3.09/21.10	17.12/35.32	0.51/0.9919
	200	5.19/21.40	19.21/35.61	0.50/0.9924
	300	5.68/21.43	19.71/35.65	0.60/0.9924

Abbreviations: LR: Learning rate; PSNR: Peak signal-to-noise ratio; SNR: Signal-to-noise ratio; SSIM: Structural similarity index.

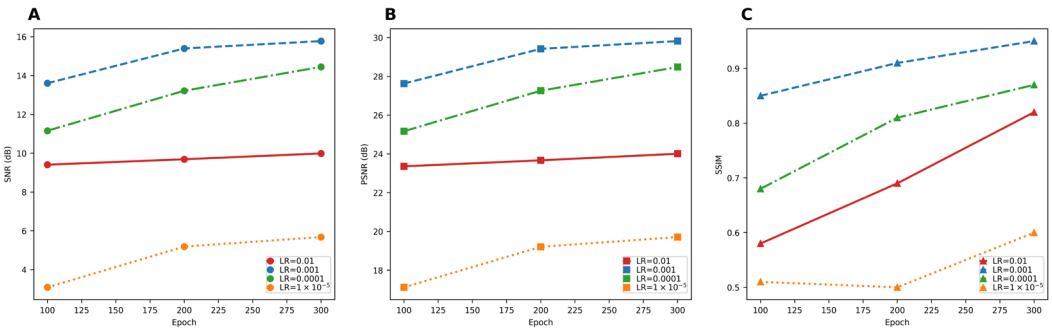


Figure 14. Variation trends of SNR, PSNR, and SSIM of the model during the training process under different learning rates in the denoising process of two-dimensional seismic data. (A) SNR vs. epoch. (B) PSNR vs. epoch. (C) SSIM vs. epoch.
Abbreviations: LR: Learning rate; PSNR: Peak signal-to-noise ratio; SNR: Signal-to-noise ratio; SSIM: Structural similarity index.

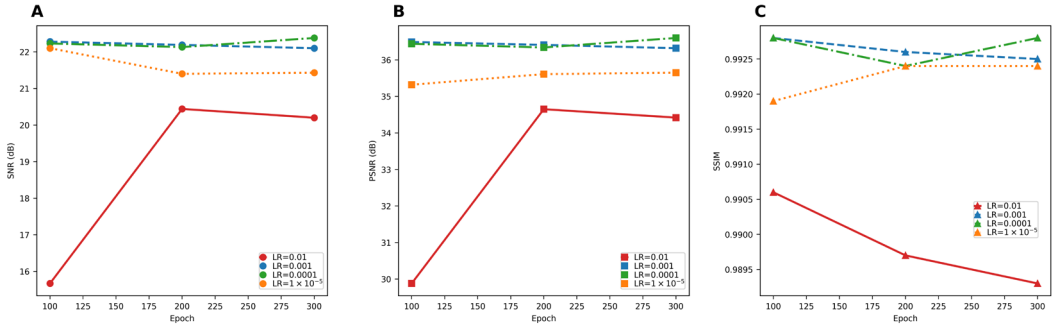


Figure 15. Variation trends of SNR, PSNR, and SSIM of the model during the training procedure under different learning rates in the one-dimensional seismic data denoising process. (A) SNR vs. epoch. (B) PSNR vs. epoch. (C) SSIM vs. epoch.
Abbreviations: LR: Learning rate; PSNR: Peak signal-to-noise ratio; SNR: Signal-to-noise ratio; SSIM: Structural similarity index.

5. Conclusion

This study proposes a TSU-Former seismic data denoising method based on spatiotemporal attention mechanisms. The approach employs an ICU module to extract multi-scale local features from noisy seismic data, while an ST-ANEM module enhances noise and suppresses signals. The Transformer module captures long-range spatiotemporal correlations in seismic profiles, effectively preserving coherence along the same seismic axis. Validation experiments demonstrated the efficacy of each network component. Large-scale experiments using synthetic data and real-world noisy seismic data confirmed that this method demonstrates significant advantages over traditional and deep learning-based base seismic data denoising techniques.

Acknowledgments

None.

Funding

None.

Conflict of interest

The authors declare they have no competing interests.

Author contributions

Conceptualization: Lan Qin, Zhenjing Yap
Investigation: Lan Qin, Jingyi Zhu, Chenyu Zhang
Supervision: Zhenjing Yao, Meng Lu
Validation: Lan Qin
Writing–original draft: Lan Qin
Writing–review & editing: Lan Qin, Zhenjing Yao, Jingyi Zhu, Chenyu Zhang

Availability of data

The data presented in this study are available upon request from the corresponding author.

References

1. Bayram H, Rastgeldi Dogan T, Şahin ÜA, Akdis CA. Environmental and health hazards by massive earthquakes. *Allergy*. 2023;78(8):2081-2084.

   doi: 10.1111/all.15736

2. Mavrouli M, Mavroulis S, Lekkas E, Tsakris A. The impact of earthquakes on public health: a narrative review of infectious diseases in the post-disaster period aiming to disaster risk reduction. *Microorganisms*. 2023;11(2):419.

   doi: 10.3390/microorganisms11020419

3. Jackson J, McKenzie D, Priestley K. Relations between earthquake distributions, geological history, tectonics and rheology on the continents. *Philos Trans R Soc A Math Phys Eng Sci*. 2021;379(2193):20190412.

   doi: 10.1098/rsta.2019.0412

4. Abudeif AM, Abdel Aal GZ, Abdelbaky NF, Abdel Gowad AM, Mohammed MA. Evaluation of engineering site and subsurface structures using seismic refraction tomography: a case study of Abydos site, Sohag governorate, Egypt. *Appl Sci*. 2023;13(4):2745.

   doi: 10.3390/app13042745

5. Mohammed MA, Abudeif AM, Abd El-aal AK. Engineering geotechnical evaluation of soil for foundation purposes using shallow seismic refraction and MASW in 15th Mayo, Egypt. *J Afr Earth Sci*. 2020;162:103721.

   doi: 10.1016/j.jafrearsci.2019.103721

6. Abudeif AM, Raef AE, Abdel Moneim AA, Mohammed MA, Farrag AF. Dynamic geotechnical properties evaluation of a candidate nuclear power plant site (NPP): P- and S-waves seismic refraction technique, North Western Coast, Egypt. *Soil Dyn Earthq Eng*. 2017;99:124-136.

   doi: 10.1016/j.soildyn.2017.05.006

7. Abudeif AM. Seismotectonic analysis of Hamada oil pool,

Libya, using Landsat and seismic data. *Arab J Geosci*. 2015;8(9):6811-6834.

doi: 10.1007/s12517-014-1700-8

8. Dittmann T, Liu Y, Morton Y, Mencin D. Supervised machine learning of high-rate GNSS velocities for earthquake strong motion signals. *J Geophys Res Solid Earth*. 2022;127(11).

doi: 10.1029/2022JB024854

9. Xiong M, Li Y, Wu N. Random-noise attenuation for seismic data by local parallel radial-trace TFPF. *IEEE Trans Geosci Remote Sens*. 2014;52(7):4025-4031.

doi: 10.1109/TGRS.2013.2278981

10. Li F, Sun F, Liu N, Xie R. Denoising seismic signal via resampling local applicability functions. *IEEE Geosci Remote Sens Lett*. 2022;19:1-5.

doi: 10.1109/LGRS.2020.3048110

11. Ding M, Zhou Y, Chi Y. Self-attention generative adversarial network interpolating and denoising seismic signals simultaneously. *Remote Sens*. 2024;16(2):305.

doi: 10.3390/rs16020305

12. Ma H, Qian Z, Li Y, Lin H, Shao D, Yang B. Noise reduction for desert seismic data using spectral kurtosis adaptive bandpass filter. *Acta Geophys*. 2019;67:123-131.

doi: 10.1007/s11600-018-0232-0

13. Li ZX, Sun NN, Zeng TS, Zhang H, Tian JQ. Adaptive subtraction based on the matching filter with three frequency bands of modeled multiples for removing seismic multiples. *IEEE Geosci Remote Sens Lett*. 2023;20:1-5.

doi: 10.1109/LGRS.2023.3261059

14. Liu Q, Jia R, Zhao C, Liu X, Sun H, Zhang X. Face super-resolution reconstruction based on self-attention residual network. *IEEE Access*. 2020;8:4110-4121.

doi: 10.1109/ACCESS.2019.2962790

15. Lin H, Li Y, Yang B. Recovery of seismic events by time-frequency peak filtering. *Proc IEEE Int Conf Image Process*. 2007;V:441-V:444.

doi: 10.1109/ICIP.2007.4379860

16. Tian Y, Li Y, Yang B. Variable-eccentricity hyperbolic-trace TFPF for seismic random noise attenuation. *IEEE Trans Geosci Remote Sens*. 2014;52(10):6449-6458.

doi: 10.1109/TGRS.2013.2296603

17. Wu N, Li Y, Ma H, Xu X. Intermediate-frequency seismic record discrimination by radial trace time-frequency filtering. *IEEE Geosci Remote Sens Lett*. 2014;11(7):1280-1284.

doi: 10.1109/LGRS.2013.2292114

18. Gaci S. The use of wavelet-based denoising techniques to enhance the first-arrival picking on seismic traces. *IEEE Trans Geosci Remote Sens*. 2014;52(8):4558-4563.

doi: 10.1109/TGRS.2013.2282422

19. Vargas RN, Paschoarelli Veiga AC. Seismic trace noise reduction by wavelets and double threshold estimation. *IET Signal Process*. 2017;11(9):1069-1075.

doi: 10.1049/iet-spr.2017.0061

20. 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

21. Hennenfent G, Fenelon L, Herrmann FJ. Nonequispaced curvelet transform for seismic data reconstruction: a sparsity-promoting approach. *Geophysics*. 2010;75(6):WB203-WB210.

doi: 10.1190/1.3494032

22. Liu W, Cao S, Li G, He Y. Reconstruction of seismic data with missing traces based on local random sampling and curvelet transform. *J Appl Geophys*. 2015;115:129-139.

doi: 10.1016/j.jappgeo.2015.02.009

23. Song J, Li Z, Wang G, Lei G, Yang J. Weak seismic signal enhancement using curvelet transform and compressive sampling. *IEEE Trans Geosci Remote Sens*. 2023;61:5907011.

doi: 10.1109/TGRS.2023.3253930

24. Liu J, Gu Y, Chou Y, Gu J. Seismic random noise reduction using adaptive threshold combined scale and directional characteristics of shearlet transform. *IEEE Geosci Remote Sens Lett*. 2020;17(9):1637-1641.

doi: 10.1109/LGRS.2019.2949806

25. Gómez JL, Velis DR. A simple method inspired by empirical mode decomposition for denoising seismic data. *Geophysics*. 2016;81(6):V403-V413.

doi: 10.1190/geo2015-0566.1

26. Han J, van der Baan M. Microseismic and seismic denoising via ensemble empirical mode decomposition and adaptive thresholding. *Geophysics*. 2015;80(6):KS69-KS80.

doi: 10.1190/geo2014-0423.1

27. Li J, Li Y, Li Y, Qian Z. Downhole microseismic signal denoising via empirical wavelet transform and adaptive thresholding. *J Geophys Eng*. 2018;15(6):2469-2480.

doi: 10.1088/1742-2140/aacf63

28. Zhang S, Li Y. Seismic exploration desert noise suppression based on complete ensemble empirical mode decomposition with adaptive noise. *J Appl Geophys*. 2020;180:104055.

doi: 10.1016/j.jappgeo.2020.104055

29. Dragomiretskiy K, Zosso D. Variational mode

decomposition. *IEEE Trans Signal Process*. 2014;62(3):531-544.

doi: 10.1109/TSP.2013.2288675

30. Yu S, Ma J. Complex variational mode decomposition for slope-preserving denoising. *IEEE Trans Geosci Remote Sens*. 2018;56(1):586-597.

doi: 10.1109/TGRS.2017.2751642

31. Yao X, Zhou Q, Wang C, Hu J, Liu P. An adaptive seismic signal denoising method based on variational mode decomposition. *Measurement*. 2021;177:109277.

doi: 10.1016/j.measurement.2021.109277

32. Gan S, Chen Y, Zu S, Qu S, Zhong W. Structure-oriented singular value decomposition for random noise attenuation of seismic data. *J Geophys Eng*. 2015;12(2):262-272.

doi: 10.1088/1742-2132/12/2/262

33. Rekapalli R, Tiwari RK, Sen MK, Vedanti N. 3D seismic data denoising and reconstruction using multichannel time slice singular spectrum analysis. *J Appl Geophys*. 2017;140:145-153.

doi: 10.1016/j.jappgeo.2017.04.001

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

doi: 10.1038/nature14539

35. Si X, Yuan Y. Random noise attenuation based on residual learning of deep convolutional neural network. *SEG Technical Program Expanded Abstracts*. 2018:1986-1990.

doi: 10.1190/segam2018-2985176.1

36. 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

37. Jiang J, Ren H, Zhang M. A convolutional autoencoder method for simultaneous seismic data reconstruction and denoising. *IEEE Geosci Remote Sens Lett*. 2022;19:7503405.

doi: 10.1109/LGRS.2021.3073560

38. Wang E, Nealon J. Applying machine learning to 3D seismic image denoising and enhancement. *Interpretation*. 2019;7(3):SE131-SE139.

doi: 10.1190/INT-2018-0224.1

39. Wang R, Zhang R, Bao C, Qiu L, Yang D. Adapting the residual dense network for seismic data denoising and upscaling. *Geophysics*. 2022;87(4):V321-V340.

doi: 10.1190/geo2021-0294.1

40. Li Y, Wang Y, Wu N. Noise suppression method based on multi-scale dilated convolution network in desert seismic data. *Comput Geosci*. 2021;156:104910.

doi: 10.1016/j.cageo.2021.104910

41. Gao Z, Zhang S, Cai J, Hong L, Zheng J. Research on deep convolutional neural network time-frequency domain seismic signal denoising combined with residual dense blocks. *Front Earth Sci*. 2021;9:681869.

doi: 10.3389/feart.2021.681869

42. Feng J, Li X, Liu X, Chen C. Multigranularity feature fusion convolutional neural network for seismic data denoising. *IEEE Trans Geosci Remote Sens*. 2022;60:5907911.

doi: 10.1109/TGRS.2021.3123509

43. Liu Y, Li CF, Li G. Active-source ocean bottom seismometer data denoising based on multiresolution convolutional neural networks. *Proc SEG Workshop, 4th Int Workshop Math Geophys*. 2022:17-20.

doi: 10.1190/iwmg2021-05.1

44. Tan M, Le QV. MixConv: mixed depthwise convolutional kernels. *arXiv*. arXiv:1907.09595.

doi: 10.48550/arXiv.1907.09595

45. Dosovitskiy A, Beyer L, Kolesnikov A, *et al*. An image is worth 16x16 words: transformers for image recognition at scale. *arXiv*. arXiv:2010.11929.

doi: 10.48550/arXiv.2010.11929

46. 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

47. 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

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

doi: 10.1109/LGRS.2025.3640968

49. Saad, Ravasi, Alkhalifah. Self-supervised multi-stage deep learning network for seismic data denoising. *Artif Intell Geosci*. 2025;6(1):100123.

doi: 10.1016/j.aiig.2025.100123

50. Saad, Alkhalifah. Distributed acoustic sensing denoising using a self-supervised conditional diffusion model. *Geophys Prospect*. 2026;74(3).

doi: 10.1111/1365-2478.70154

51. Saad OM, Chen Y, Alkhalifah T. U-Trans: a foundation model for seismic waveform representation and enhanced downstream earthquake tasks. *Sci Rep*. 2026.

doi: 10.1038/s41598-026-41454-x

52. Xu Z, Luo Y, Wu B, Meng D. S2S-WTV: seismic data noise attenuation using weighted total variation regularized self-supervised learning. *IEEE Trans Geosci Remote Sens*.

2023;61:1-15.

doi: 10.1109/TGRS.2023.3268554

53. Chen Y, Savvaidis A, Saad OM, *et al*. TXED: the Texas earthquake dataset for AI. *Seismol Res Lett*. 2024;95(3):2013-2022.

doi: 10.1785/0220230327