ACCSCIENCE PUBLISHING

Journal of Seismic Exploration

ARTICLE

Convolutional autoencoder with local wavefield characteristic constraint for suppressing near-surface seismic scattered waves

Daling Hou, Jixiang Xu*, and Meng Li

Institute of Oil and Gas Geophysics, Research Institute of Petroleum Exploration and Development, Beijing, China

Abstract

In mountainous seismic exploration, complex near-surface environments cause strong wave scattering, reducing the signal-to-noise ratio and complicating data processing. Therefore, suppressing scattered waves is crucial. To address this issue, this study proposes a convolutional autoencoder constrained by local scattered wave characteristics to suppress near-surface scattered waves (NSWs) in full-wavefield seismic data. This method uses the scattered waves predicted by seismic interferometry as the network input, and the original seismic records containing true scattered waves as the label. Since there are differences in amplitude and phase between the predicted scattered waves and the true scattered waves, the network introduces local wavefield features of the scattered waves for constraint, and adds a smoothness regularization term in the loss function to ensure the continuity of the output waveform. After training, the network maps the energy of predicted scattered waves into the actual seismic records, thereby accurately extracting scattered waves. Finally, by subtracting the network output from the original records, clean data with scattered waves removed can be obtained. This method is a self-supervised learning strategy and does not require additional clean signal samples. During training, the weights of each item in the loss function can be dynamically adjusted to guide the network to focus on local scattered-wave features, avoid learning effective wave information, and ensure that only scattered-wave components are retained in the output. Practical application results show that this method can effectively suppress NSWs and improve the signal-to-noise ratio of seismic data.

Keywords: Near-surface scattered waves; Convolutional autoencoder; Interferometry; Wavefield characteristics

***Corresponding author:**
Jixiang Xu
(xjx6611@petrochina.com.cn)

Citation: Hou D, Xu J, Li M. Convolutional autoencoder with local wavefield characteristic constraint for suppressing near-surface seismic scattered waves. *J Seismic Explor*. doi: 10.36922/JSE025470112

Received: November 17, 2025

Revised: December 23, 2025

Accepted: December 26, 2025

Published online: January 27, 2026

1. Introduction

When seismic waves propagate underground, variations caused by three-dimensional spatial heterogeneity are referred to as scattered waves. In complex regions, especially in the near-surface, medium heterogeneity is the greatest. The widespread presence of small-scale heterogeneities, strong velocity and density contrasts, and irregular interfaces leads to significant impedance contrasts along the wave propagation path, thereby generating wave "reflections." Because these heterogeneities are much smaller in scale and more irregular in geometry than subsurface reflecting interfaces, the resulting

wavefield response mainly manifests as scattering. These scattered waves overlap with the effective waves in time, space, and frequency, reducing the signal-to-noise ratio (SNR) of seismic data and affecting the accuracy of subsequent inversion and imaging. Wu *et al.*[1] pointed out that the primary cause of low SNR in seismic data is the influence of near-surface scattered waves (NSWs). Therefore, suppressing NSWs is a critical step in obtaining high-quality seismic data.

Scattered-wave suppression is a complex task that involves not only analyzing the properties of scattered waves but also developing advanced processing techniques.[2] Traditional filtering methods often remove scattered waves at the cost of losing effective waves. Thus, an alternative approach—analyzing the characteristics of scattered waves, predicting the true scattered wavefield, and subtracting it from seismic data—minimizes the impact on effective waves and is widely applied in practice. Campman *et al.*[3] introduced impedance functions to model and suppress NSWs. Liu *et al.*[4] estimated and removed scattered waves from radial and vertical components using filtering and correlation analysis based on transverse component assumptions, thereby improving data SNR. Almuhaidib and Toksöz[5] employed controllable filters to estimate spatial variations in noise and used directional non-linear filters for noise separation. In recent years, seismic interferometry has been effectively applied to suppress scattered waves. Xu *et al.*[6] provided a detailed discussion on NSW characteristics and the principles of scattered-wave interferometry, enhancing the coherence of scattered waves to separate them from true records. Halliday *et al.*[7] proposed a model-driven seismic interferometry approach, using models as inputs to estimate scattered surface waves between sources and receivers. Xu *et al.*[8] applied an improved source–receiver interferometry method to reconstruct NSW fields from pre-stack data and then adaptively subtracted the predicted scattered waves using matched filtering, effectively suppressing NSWs.

Deep learning, a subfield of machine learning, has been the fastest-growing branch in the past decade.[9] Deep neural networks construct multi-layer networks to represent targets at multiple levels, aiming to use hierarchical features to capture abstract semantic information from data. Among them, convolutional neural networks (CNNs) leverage local connectivity and weight sharing to effectively capture local features while reducing the number of parameters. In recent years, deep neural networks have shown promising applications in seismic denoising. Through synthetic and real-data demonstrations, deep learning has exhibited significant potential in attenuating random noise, linear noise, and multiple reflections.[10] Yu *et al.*[11] used a denoising CNN to suppress surface waves in synthetic data and transferred the trained network to real seismic data. Liu *et al.*[12] trained a convolutional autoencoder (CAE) using synthetic noise and standard induced polarization signals to denoise continuous random noise in induced polarization signals. Sui *et al.*[13] developed data-adaptive dictionaries through deep unfolding dictionary learning to construct training networks suitable for different noise levels. Li *et al.*[14] designed an automated training data generation workflow based on the noise characteristics of nodal seismic acquisition systems, achieving superior performance in suppressing environmental noise compared to traditional methods. In practical seismic denoising, obtaining noise-free sample data is challenging. Supervised learning often uses traditional denoised data or synthetic data as labels, whereas self-supervised and unsupervised learning methods eliminate the need for clean labels. As a result, deep learning applications in seismic denoising are shifting from supervised to self-supervised and unsupervised approaches, which better align with seismic processing workflows. Song *et al.*[15] employed a CAE for unsupervised random noise suppression. Shao *et al.*[16] applied the Noisy2Noisy framework to remove random noise, achieving supervised learning-level denoising performance without labeled data. Wu *et al.*[17] proposed a self-supervised deep learning approach based on the Neighbor2Neighbor framework to suppress random noise. For coherent noise suppression, a series of self-supervised denoising networks was introduced to remove surface-related multiples, demonstrating promising results;[18-21] Sun *et al.*[22] proposed a multi-channel U-Net with multi-mode and multi-model capabilities for suppressing multiples; Qi *et al.*[23] introduced an interpretable Fast Iterative Shrinkage-Thresholding Algorithm Network for adaptive multiple suppression. Despite significant progress and breakthroughs in seismic denoising using deep learning, research and applications targeting NSW suppression remain largely unexplored.

In image processing and computer vision, commonly used image quality assessment methods are often based on error visibility, such as mean-squared error (MSE) and Peak SNR. While these methods effectively quantify image differences, they fail to accurately reflect human visual perception of image quality. To address this, Wang and Bovik[24] proposed the structural similarity index (SSIM), which evaluates image similarity by considering luminance, contrast, and structure, thereby aligning more closely with human visual perception. SSIM has been widely applied in image denoising and reconstruction. In recent years, SSIM has also been adopted in seismic denoising, seismic interpolation, and velocity modeling. Almadani *et al.*[25] used SSIM to evaluate the denoising performance

of convolutional sparse coding. Sun and Williamson[26] applied SSIM to assess the advantages of frequency-constrained neural networks in seismic interpolation. Simon *et al.*[27] employed SSIM to evaluate the effectiveness of transfer learning in seismic velocity modeling. In this study, we modify and extend SSIM to better capture the characteristics of seismic wavefields and incorporate it into the loss function to enhance network performance.

To more effectively suppress NSWs, this study proposes a CAE constrained by local scattered wave characteristics (LSC-CAE). This method does not rely on additional clean data as labels but instead constructs a self-supervised denoising process. The core idea is to predict the scattered-wave structures, restore their energy to the real-data level, and suppress them by subtraction from the full-wavefield data. Specifically, NSWs are first predicted using seismic interferometry and then used as the input of the CAE; at the same time, full-wavefield data containing scattered waves are used as labels to guide the network in learning the spatial and temporal characteristics of scattered waves. To enable effective training, we designed a composite loss function composed of MSE, a local scattered-wave feature loss, and a waveform-smoothing regularization term, thereby encouraging the network to focus on the structural and energy features of scattered waves and achieve energy matching with real scattered waves. Finally, by subtracting the network output from the full-wavefield data, the denoised seismic record was obtained. In addition, to enhance the network's ability to model complex wavefield structures, a multi-head self-attention mechanism was introduced in the intermediate layers of the CAE to improve the network's perception of global scattered-wave features. Meanwhile, residual structures were introduced to accelerate training and enhance network stability. Application results on real seismic data show that the proposed method can effectively suppress scattered waves while preserving the integrity of effective waveforms, demonstrating good denoising performance and practical value.

2. Methods

2.1. Near-surface interferometry

Seismic interferometry is a wavefield reconstruction technique based on wave theory that extracts intercorrelation information from seismic data to reconstruct seismic wave propagation paths. The theoretical foundation of seismic interferometry lies in the Green's function, which allows the cross-correlation of wavefields between two points to be reconstructed as the seismic response of one point acting as a virtual source and the other as a receiver. Schuster[28] first introduced the concept of seismic wave interferometry in 2001 and subsequently conducted a series of studies exploring its principles.[29-33] Following this, several studies provided comprehensive analyses of the principles, advancements, and applications of seismic interferometry.[34-36] Today, seismic interferometry has been widely applied in seismic imaging and seismic monitoring, playing a crucial role in enhancing subsurface characterization and detecting temporal changes in the Earth's structure.

Under the high-frequency and far-field approximations, the source imaging and seismic monitoring, playing a crucial **Equation (I)**:[37]

$$G(y_1,x_1;\omega)+G^*(y_1,x_1;\omega)\approx\frac{4}{(\rho c)^2}\times \oint_{S_1}G(y_1,x;\omega)dS_1\oint_{S_2}G^*(y,x;\omega)G(y,x_1;\omega)dS_2 \quad \text{(I)}$$

Where ω is the angular frequency, ρ and c represent density and velocity, * denotes the complex conjugate, and $G(y_1,x_1;\omega)$ represents the wavefield generated by point source x_1 at point y_1, and similarly for $G(y_1,x;\omega)$ and $G(y,x_1;\omega)$.

When intersource interferometry is computed using correlation, it introduces wavelet distortion in the frequency domain. The amplitude spectrum of the reconstructed virtual wavelet becomes the cube of the original wavelet spectrum, leading to amplitude distortion and peak-like features. Deconvolutional interferometry can effectively remove wavelet effects from seismic records. Therefore, this study employed the improved source–receiver interferometry method[8], expressed as **Equation (II)**:

$$G(y_1,x_1;\omega)+G^*(y_1,x_1;\omega)\approx\frac{4}{(\rho c)^2}\times \oint_{S_1}G(y_1,x;\omega)dS_1\oint_{S_2}\frac{G(y,x_1;\omega)}{G(y,x;\omega)}dS_2 \quad \text{(II)}$$

We used deconvolution to replace cross-correlation calculations, making the reconstructed wavelet closer to the original in shape and relative amplitude, and reducing nonphysical artifacts and spurious energy generated by source interference. Then, a convolution-based shot-receiver interferometry method compensated for the common path between the near-surface scatterers and common shot points, restoring the wavefield's propagation characteristics. This procedure effectively suppressed the influence of the source wavelet and improved the quality of the reconstructed wavefield. Applying this method to predict the NSWs in the actual seismic data (Figure 1A), a wavefield dominated by scattered-wave energy was obtained (Figure 1B).

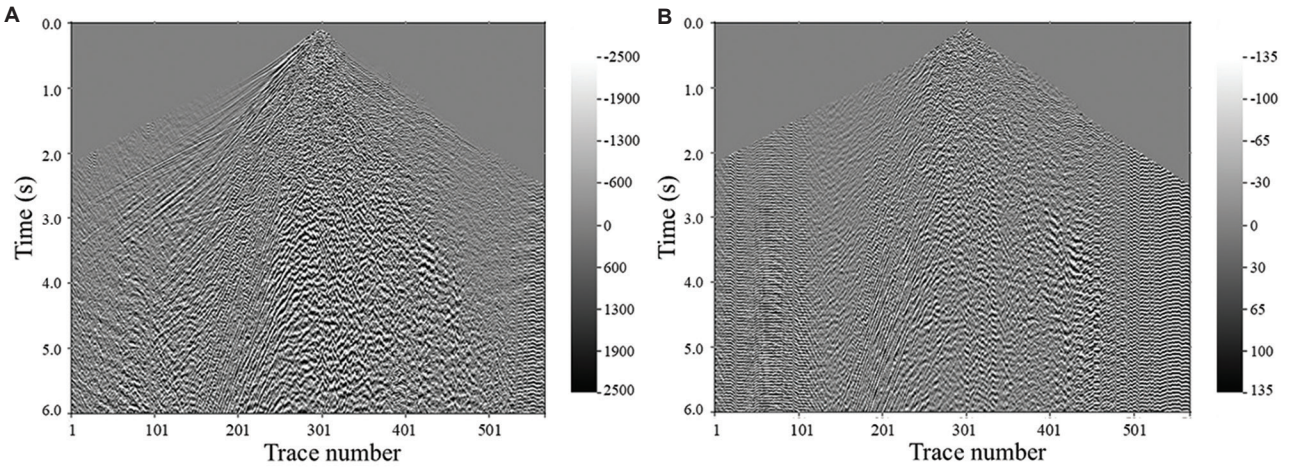

Figure 1. Single-shot record. (A) Full-wavefield data. (B) Interferometric predicted data.

2.2. Improved CAE

A CAE is a deep learning model that combines the concept of an autoencoder with the characteristics of CNNs. It replaces the fully connected layers in traditional autoencoders with convolutional and pooling layers, enabling the CAE to reduce the dimensionality and compress input data with fewer parameters while retaining important spatial features. This makes it well-suited for handling high-dimensional data. The CAE consists of an encoder and a decoder. The encoder, which includes convolutional and pooling layers, extracts invariant features and compresses the input into a latent space representation. The decoder uses deconvolutional layers to reconstruct the input from the latent-space representation. The network architecture is summarized in Figure 2.

During training, the CAE adjusted the hyperparameters of the convolutional and pooling layers to minimize the difference between the reconstructed image (obtained by adjusting the hyperparameters) and the original image. Ultimately, an optimal basic framework was identified, consisting of one convolutional layer, one pooling layer, and their corresponding hyperparameters. Additional convolutional and pooling layers were then stacked to form a large neural network, with a fully connected layer added at the end to create a CNN-based architecture. The learning process of the CAE is represented as **Equation (III)**:

$$y = F(x,\theta) \quad \text{(III)}$$

where x and y represent the network's input and output, respectively. θ refers to the parameters that the network needs to learn, which include the weights w_i and biases b_i of each convolutional layer in the encoder, and the weights w_d and biases b_d of the deconvolutional layer in the decoder.

For seismic data, deep networks are typically required to extract complex waveform features, thereby providing a better and more comprehensive description of the target. However, when the network depth reaches a certain level, increasing it can degrade network performance due to the accumulation of errors during information transfer. To address this, He *et al.*[38] proposed a residual structure with identity mapping to avoid issues, such as gradient vanishing and model degradation, when the network depth increases. The residual structure adds the input to the learning result before passing it to the next learning unit, enabling deep networks to retain the complete information from previous layers. The information from identity mapping in the previous layer can be viewed as pre-learned knowledge, so each layer only needs to learn the difference between the existing knowledge and the ideal model. This can be expressed as **Equation (IV)**:

$$y_{i+1} = h(x_i)+F(x_i,\theta_i) \quad \text{(IV)}$$

where x_i is the input, the output y_{i+1} is the result of adding the identity mapping $h(.)$ and the transformation function $f(.)$, where the transformation function may include multiple operations, such as convolution, normalization, and activation.

In seismic data, signals at different spatiotemporal locations may exhibit complex dependencies. Compared to the local receptive field of CNNs, the multi-head self-attention mechanism can directly capture global-scale features.[39] The core idea of self-attention is to compare each element in a sequence with all others and assign weights that reflect the degree of correlation between them. To further enhance feature extraction, this study employed a multi-head self-attention module. Specifically, eight independently learned linear projections were used to

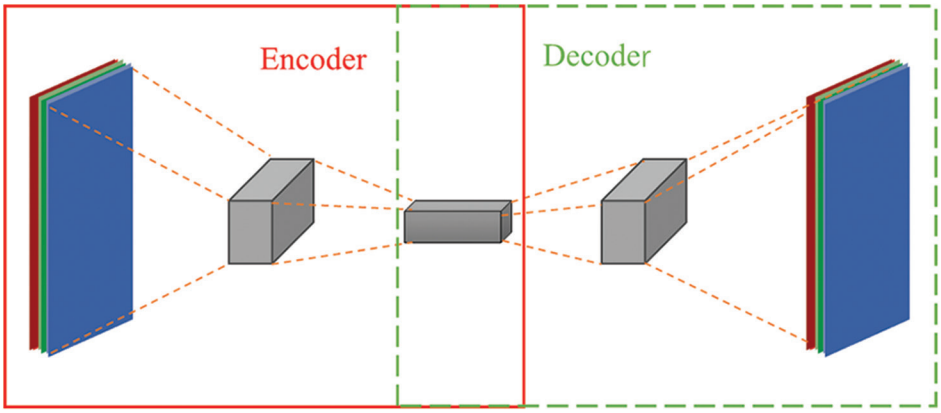

Figure 2. Convolutional autoencoder structure

represent queries, keys, and values in different subspaces. The attention of each head was computed in parallel and then concatenated to generate the final output. The core principle and detailed computational process of the multi-head self-attention mechanism were as follows:

i) Project input feature matrix $\mathbf{X} \in \mathbb{R}^{n \times d_{model}}$ into queries, keys, and values.
Q=XW^{Q}, K=XW^{K}, V=XW^{V}, where $W^{Q}, W^{K}, W^{V} \in R^{d_{model} \times d_k}$, d_k represents the dimension of the Q/K, which is typically set as $d_k = d_{model}/h$, where h is the number of attention heads.

ii) Dimension splitting: The matrices Q, K, and V were split along the feature dimension d_k into h independent submatrices, each with a dimension of d_k:

$$Q=[Q_1,Q_2,\ldots Q_h]$$
$$K=[K_1,K_2,\ldots K_h]$$
$$V=[V_1,V_2,\ldots V_h]$$

iii) Multi-head parallel computation: For each submatrix triplet (Q_i,K_i,V_i), an independent single-head self-attention operation was performed to obtain h output results:

$$Output_i = Soft\max\left(\frac{Q_i K_i^T}{\sqrt{d_k}}\right)V_i, i = 1,2,\ldots\ldots h$$

iv) Concatenation: The h output results were concatenated along the feature dimension to obtain the output:

$$Output_{concat} = [Output_1, Output_{2,\ldots} Output_h]$$

v) Then, a linear layer $W_O \in \mathbb{R}^{d_{model} \times d_{model}}$ was applied for feature fusion, resulting in the final multi-head self-attention output:

$$Output = Output_{contact} W_o$$

As shown in Figure 3, the proposed network with residual connections and a multi-head self-attention mechanism is referred to as the residual self-attention convolutional autoencoder (RS-CAE). The model adopted a U-Net style encoder-decoder framework, where both the encoder and decoder consisted of multiple stacked residual blocks. A multi-head self-attention module was inserted at the bottleneck layer to enable a more comprehensive representation of the global wavefield structure. This design enabled the network to capture both short- and long-range dependencies, establish long-distance relationships in higher-order features, and restore structural details of the wavefield, thereby enhancing the network's ability to distinguish scattered waves from other wavefield components.

The detailed configuration of the proposed RS-CAE is listed in Table 1.

2.3. Loss function

2.3.1. Local scattered-wave feature loss

The SSIM was first proposed by Wang *et al.*[40] and was used for image denoising and reconstruction. The original image information was almost completely recovered through a simple point-to-point inverse linear brightness transformation, while the structural information of the reference image was well preserved. The SSIM primarily measures the luminance $l(x,x')$, contrast $c(x,x')$, and structure $s(x,x')$ of the input data and the output data, and the final similarity is a function of these three components, expressed as **Equation (V)**:

$$S(x,x') = f(l(x,x'),c(x,x'),s(x,x')) \quad (V)$$

For seismic data, luminance reflects the difference in the mean amplitude between the input and output

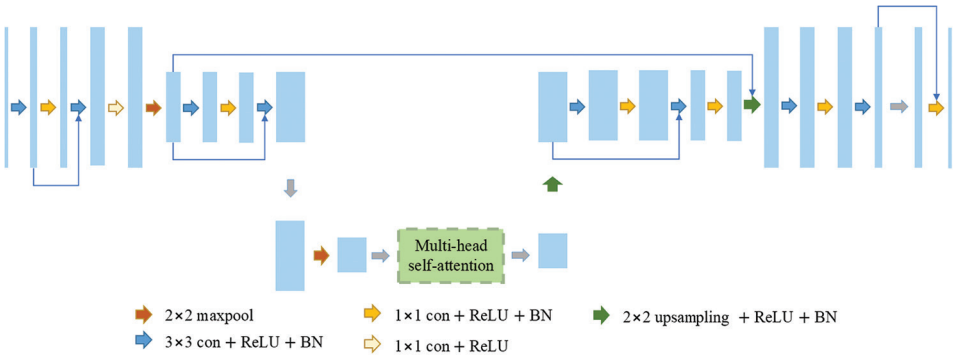

Figure 3. Structure diagram of the improved convolutional encoder
Abbreviations: BN: Batch normalization; ReLU: Rectified linear unit.

Table 1. Parameters of residual-self-attention convolutional autoencoder (RS-CAE)

Module	Layer type	Kernel size	Channels	Output size	Activation
Encoder block 1-1	Conv2D+BatchNorm	3×3	1atc	256×256	ReLU
Encoder block 1-2	Conv2D+BatchNorm	1×1	64→64	256×256	ReLU
Encoder block 2	Conv2D+MaxPool+BatchNorm	3×3/1×1/2×2	64 chN	128×128	ReLU
Encoder block 3	Conv2D+BatchNorm	3×3/1×1	128chNo	128×128	ReLU
Encoder block 4	Conv2D+MaxPool+BatchNorm	3×3/1×1/2×2	128ch56	64×64	ReLU
Bottleneck (attention)	Multi-head self-attention (eight heads)	-	256ads)	64×64	-
Decoder block 1-1	ConvTranspose2D+BatchNorm	3×3	256chNo	128×128	ReLU
Decoder block 1-2	Conv2D+BatchNorm	3×3/1×1	256chNo	128×128	ReLU
Decoder block 2	Conv2D+BatchNorm	3×3/1×1	256chNo	128×128	ReLU
Decoder block 3	ConvTranspose2D+Conv2D+BatchNorm	3×3/1×1	128chN	256×256	ReLU
Decoder block 4	Conv2D+BatchNorm	1×1	64 ch	256×256	ReLU
Output layer	Conv2D	1×1	64 v	256×256	-

Abbreviation: ReLU: Rectified linear unit.

data, indicating the consistency of the average energy distribution between the two datasets. Contrast measures the amplitude differences between the two datasets and evaluates whether the output data can preserve the relative amplitude variations of the input in the output, using the standard deviation as an estimate. Structural similarity reflects the differences in seismic waveform and phase between the input data and the network output, estimated using the value normalized by the standard deviation of the signal itself.

During the interferometric process, coherent signals are enhanced, resulting in a significant overall energy difference between the predicted scattered wave x and the real data y. Therefore, directly applying the luminance term of SSIM is not suitable for scattered-wave analysis, while the relative amplitude and phase differences of the waveform are relatively small. Inspired by the SSIM, we defined a scattered-wave feature loss that constrained the network output's amplitude and waveform based on the contrast and structural differences among the predicted scattered waves, the true scattered waves, and the full-wavefield data, thereby guiding the network to suppress scattered waves without compromising effective waveforms.

The similarity of scattered-wave features is described using standard deviation and covariance, with the calculation formula as shown in **Equation (VI)**:

$$S_{SS} = \frac{2\sigma_{xx'} + 2C_1}{\sigma_x^2 + \sigma_{x'}^2 + C_1} \quad \text{(VI)}$$

Where C_1 is a constant to prevent the denominator from being zero during the calculation, defined as **Equation (VII)**:

$$C_1 = (K_1 L)^2 \quad \text{(VII)}$$

Typically, K_1 << 1and taken as 0.001–0.003, L representing the range of normalized input data. σ_x and $\sigma_{x'}$

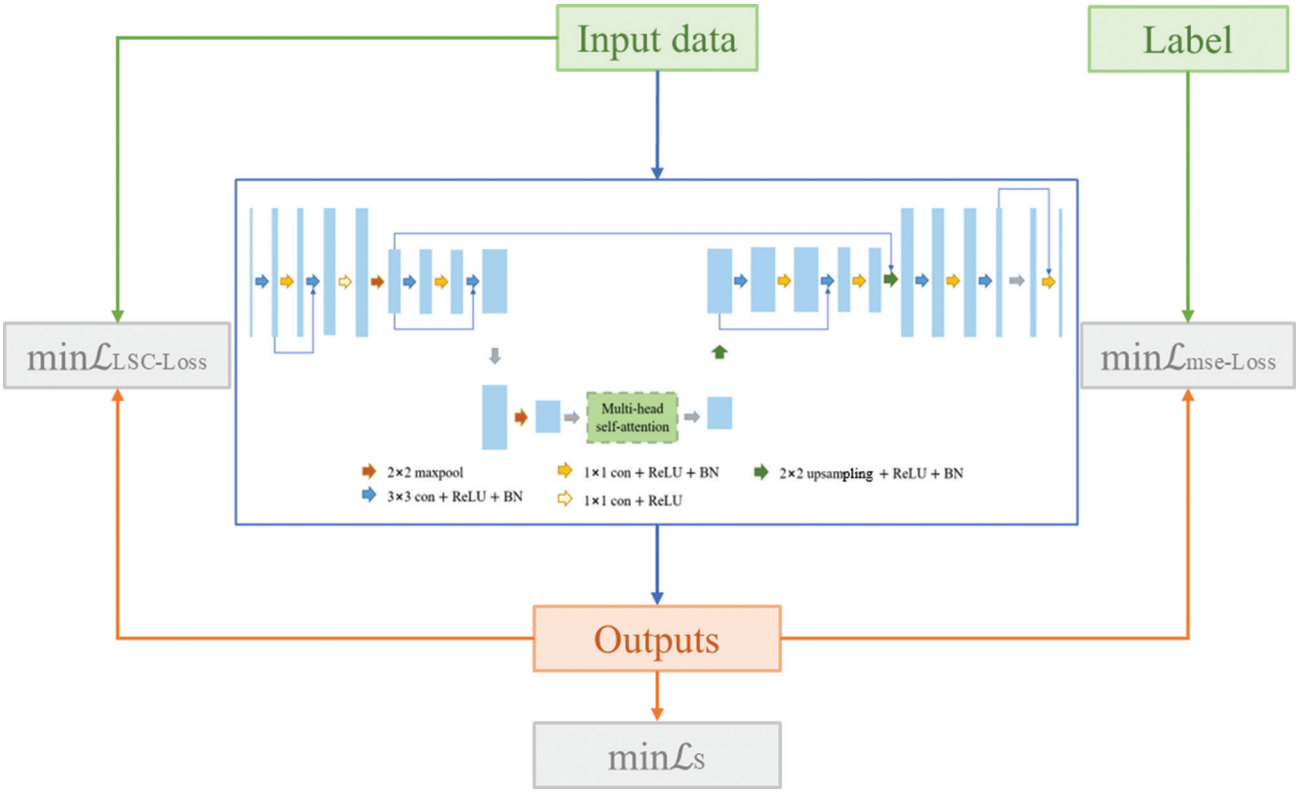

Figure 4. Construction of the loss function

are the standard deviations of the input and output data, defined as **Equations (VIII) and (IX)**:

$$\sigma_x = \left(\frac{1}{N-1} \sum_{i=1}^{N} \left(x_i - \mu_x \right)^2 \right)^{\frac{1}{2}} \quad \text{(VIII)}$$

$$\sigma_{x'} = \left(\frac{1}{N-1} \sum_{i=1}^{N} \left(x_i' - \mu_{x'} \right)^2 \right)^{\frac{1}{2}} \quad \text{(IX)}$$

$\sigma_{xx'}$ is the covariance of x and x' that reflect the covariance of signals in structure, and is used to measure the similarity of data in local structure, that is, shape and phase. The calculation formula is shown in **Equation (X)**:

$$\sigma_{xx'} = \frac{1}{N-1} \sum_{i=1}^{N} \left(x_i - \mu_x \right) \left(y_i - \mu_y \right) \quad \text{(X)}$$

$\mu_{x'}$ and $\mu_{y'}$ represent the mean values of the output data and the label data, respectively. S_{SS} evaluates the similarity between the network output and the input. The training objective is to minimize the difference, so it needs to be transformed into a difference value to obtain the final local scattered-wave feature loss, which is given by **Equation (XI)**:

$$L_{LSC} = 1 - S_{SS} \quad \text{(XI)}$$

2.3.2. Smoothing regularization

In 2017, Goodfellow *et al.*[41] stated in the book *Deep Learning* that any modifications that can improve the model's accuracy during testing and enhance the model's generalization ability can be considered regularization. The application of regularization terms is achieved by adding a penalty term to the model's loss function to enhance the model's generalization ability, such as L_1 and L_2 regularizations. Unlike the loss function, the added regularization term is independent of the label data, imposing certain prior knowledge constraints on the data to improve the model's denoising effectiveness. During the self-supervised denoising process, coherent scattered noise data exhibits good local continuity. Therefore, a smoothing regularization term is added to the loss function to constrain waveform continuity, using the first derivative differences of adjacent points in the network output to measure waveform continuity, as shown in **Equations (XII) and (XIII)**:

$$Loss_t = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M}\left(u(i+1,j)-u(i,j)\right)^2 \quad \text{(XII)}$$

$$Loss_x = \frac{1}{M}\sum_{i=1}^{N}\sum_{j=1}^{M}\left(u(i,j+1)-u(i,j)\right)^2 \quad \text{(XIII)}$$

where N and M represent the time dimension and spatial dimension, respectively. **Equations (XII) and (XIII)** represent the squared differences of adjacent points in the two dimensions. Combining these two terms yields the final smoothing regularization term, as in **Equation (XIV)**:

$$L_S = Loss_t + Loss_x \quad \text{(XIV)}$$

2.3.3. Composite loss function

The local scattered-wave characteristics loss is combined with the MSE loss, and a smoothing regularization term is added to guide the network in learning the scattered waves of the label data while avoiding excessive attention to effective waves, as shown in Figure 4. Specifically, MSE Loss serves as the reconstruction loss, minimizing the difference between the network output and the label data, ensuring the predicted scattered-wave energy closely matches the true values. In addition, to ensure that the network focuses on learning scattered-wave characteristics and ignores other wavefields, including reflected waves, a term L_{LSC} is introduced to minimize the structural differences between the network output and the input data wavefield. This helps the network output retain the waveform and location information of the scattered waves. Seismic coherent noise and reflected waves differ from random noise, typically exhibiting continuity and specific shapes. Deep learning methods are data-driven and often lack interpretability in complex workflows, such as seismic signal denoising. Therefore, adding a regularization term to the loss function ensures the continuity of coherent noise wavefields, calculates the network output, and ensures the data conforms to wavefield characteristics, standardizing the learning process and enhancing the physical rationality of the network.

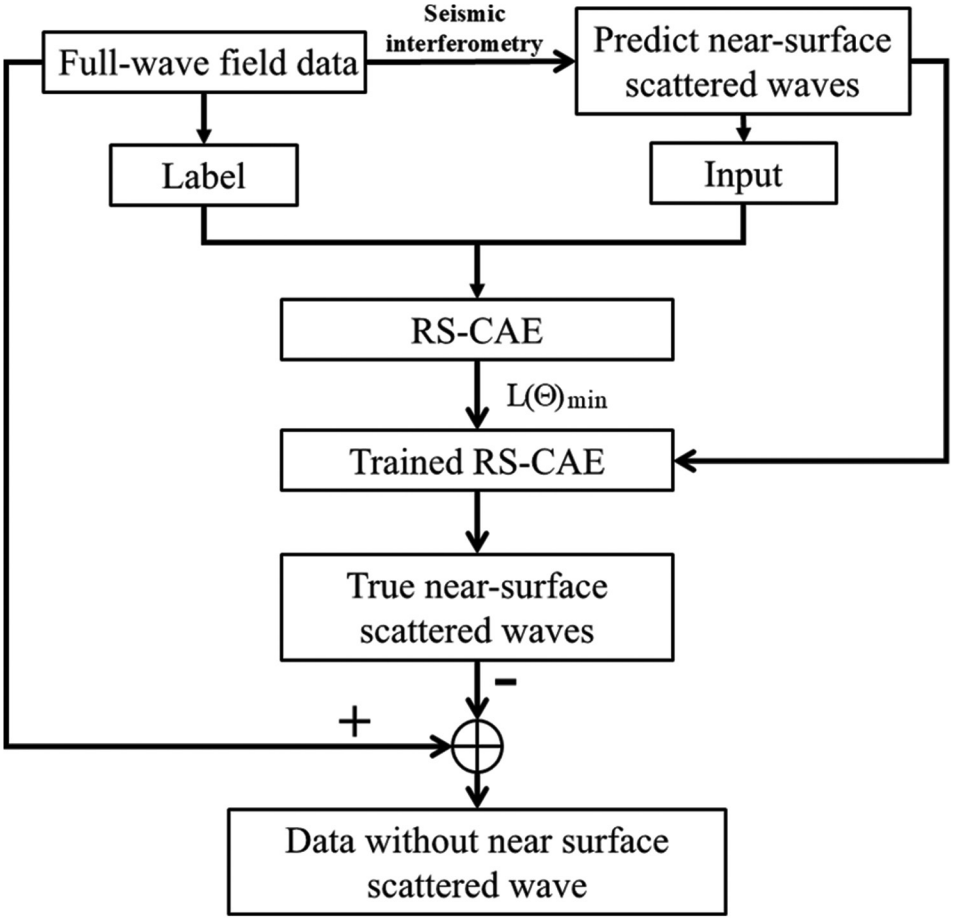

Figure 5. Flow chart of denoising
Abbreviation: RS-CAE: Residual-self-attention convolutional autoencoder.

Finally, the expression for the total loss function is as in **Equation (XV)**:

$$L = \lambda_1 L_{mse} + \lambda_2 L_{LSC} + \lambda_3 L_s \quad \text{(XV)}$$

where λ_1, λ_2, and λ_3 are the weight coefficients for the three types of losses, satisfying the condition: $\lambda_1+\lambda_2+\lambda_3=1$, L_{mse} is the MSE loss, expressed as **Equation (XVI)**:

$$L_{mse} = \frac{1}{N}\sum_{i=1}^{N}\left(x_i' - y_i\right)^2 \quad \text{(XVI)}$$

2.4. Near-surface scattered-wave suppression

In the process of NSW suppression, seismic interferometry reconstructs the predicted scattered-wave data X, whereas the original wavefield data Y serves as the network label. Using the learning process of CAEs, the loss function L is minimized to obtain the true scattered-wave data, which is then subtracted from the original data to achieve denoising. The specific denoising process is shown in Figure 5. Before training, both synthetic and field data were divided into training and test sets at an 8:2 ratio, with the latter used for final model performance evaluation. Meanwhile, to make the model easier to train and converge, it is necessary to normalize both the predicted scattered waves and the original wavefield data so that their ranges are similar. The normalization was applied to single-shot data, and due to the large amplitude differences, mean-variance normalization was used to process the data, resulting in the predicted scattered wave x and label y for the network input (**Equations [XVII] and [XVIII]**):

$$x = \frac{X - \mu_1}{\sigma_1} \quad \text{(XVII)}$$

$$y = \frac{Y - \mu_2}{\sigma_2} \quad \text{(XVIII)}$$

where μ and σ represent the mean and standard deviation of the data, respectively.

To preserve more local waveform characteristics, the normalized seismic data were divided into equally-sized blocks to create a training dataset for the network. During the learning process, a dynamic learning rate was used, which was reduced by half every 200 iterations. As the learning rate decreased, L continuously dropped, and when it reached its lowest value and stabilized, it indicated that the network learned the feature parameters of the input and label data. Finally, the test data were used as input to obtain the true NSW, which can be directly subtracted from the full-wavefield data. All experiments were implemented using PyTorch (Version 2.5.1, Meta Platforms, Inc., United States of America) and conducted on a workstation equipped with an NVIDIA RTX 4090 GPU with 16 GB memory.

Table 2. Parameter design of the loss function

Iterations	0–200	201–400	401–500
Parameter (λ_1,λ_2,λ_3)	(0.7, 0.25, 0.05)	(0.5, 0.4, 0.1)	(0.3, 0.5, 0.2)
L_{min}	0.1512	0.1704	0.1811

3. Results and discussion

3.1. Training process

For the model dataset, the training process consisted of 500 iterations (Table 2). In the early stage, restoring the energy of the predicted scattered waves was the primary objective. Because there were significant discrepancies between the predicted and true scattered waves, especially for field data, the loss term associated with energy matching was assigned a relatively higher weight to ensure that the network prioritized alignment with the energy distribution of the original data. After 200 iterations, the loss decreased and stabilized at 0.1512, indicating that the network had successfully learned the energy discrepancy between data pairs and had compensated for it accordingly. Subsequently, the weight of L_{mse} was gradually reduced, shifting the network's focus from energy recovery to learning the structural characteristics of waveforms. However, increasing λ_2 may lead to a slight reduction in energy-matching accuracy, highlighting the need for a trade-off between these two aspects.

The training loss curve of the model dataset over 500 iterations is shown in Figure 6. The loss decreased rapidly in the early stage, indicating that the network quickly learned the energy features. Around iteration 200, the curve began to stabilize, eventually reaching approximately 0.15, indicating that the network successfully converged. During

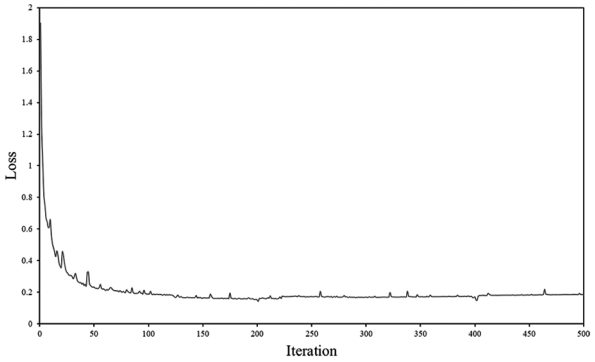

Figure 6. Training loss curve on the model dataset

training, weight adjustments were applied to balance the contributions of different loss components, causing minor fluctuations. Overall, the curve demonstrated that the network effectively captured both the energy and waveform characteristics of the training data.

Figure 7 illustrates the performance of synthetic data at different training stages. For comparison, the network was trained with λ_1 = 1 for 500 iterations, yielding the result as shown in Figure 7A. In complex wavefields, the network primarily focused on simple energy recovery. When adopting

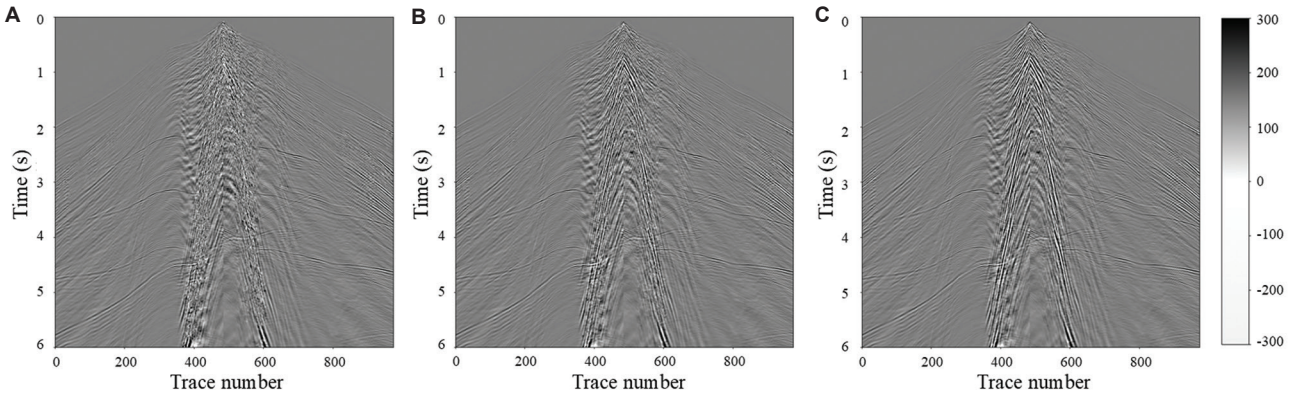

Figure 7. Results of synthetic data processed with different parameters. (A) Results with λ_1=1 after 500 iterations. (B) Result with dynamically adjusted weights after 300 iterations. (C) Result with dynamically adjusted weights after 500 iterations.

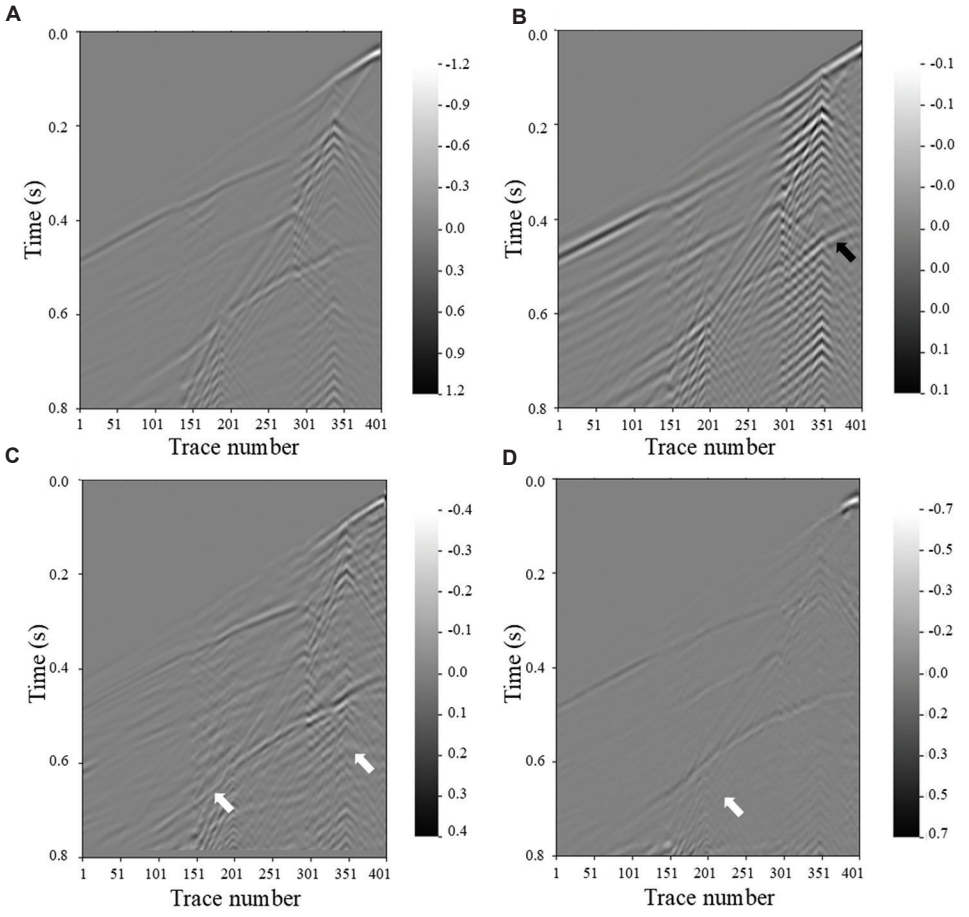

Figure 8. Denoising results of synthetic data 1. (A) Original data. (B) Interferometric data. (C) Results after matched filtering. (D) Denoising results of a convolutional autoencoder constrained by local scattered wave characteristics (LSC-CAE).

a dynamic parameter adjustment strategy to gradually increase λ_2 and λ_3, as shown in Figure 7B and 7C, the waveform characteristics of the test data were better preserved on the basis of energy recovery, demonstrating the effectiveness of LSC-CAE in suppressing noise in seismic data.

3.2. Application of synthetic data

To validate the effectiveness of the proposed method, LSC-CAE was applied to synthetic seismic records to suppress noise. The simple model consisted of two horizontal layers, with a velocity anomaly introduced near the surface to generate scattered waves. A total of 70 shot records were simulated. The predicted scattered-wave data were obtained through interferometric processing. Figure 8A and B shows the original data and the predicted scattered-wave data that served as the network's labels and inputs, respectively. However, due to the limited aperture effect in the simulation, the interferometric processing could not entirely eliminate body waves, leaving residual reflection signals in the predicted data (black arrow). Figure 8C and D shows the denoised results obtained using matched filtering and LSC-CAE, respectively. Compared to matched filtering, the network-based denoising not only better preserved the reflection signals but also more effectively suppressed scattered waves (white arrows). Moreover, the computational time for network-based denoising was only a few seconds, demonstrating significantly higher efficiency than matched filtering.

To further verify the effectiveness of the network-based denoising, a more complex undulating surface model was constructed for forward modeling, generating 70 synthetic shot records. Figure 9A and B shows the original data and the predicted scattered-wave data, respectively. In the original data, the reflection signals were overwhelmed by noise. The interferometric processing results demonstrate that surface and scattered waves were well predicted, with waveform amplitude and phase closely matching the original data, providing high-quality training samples for the network.

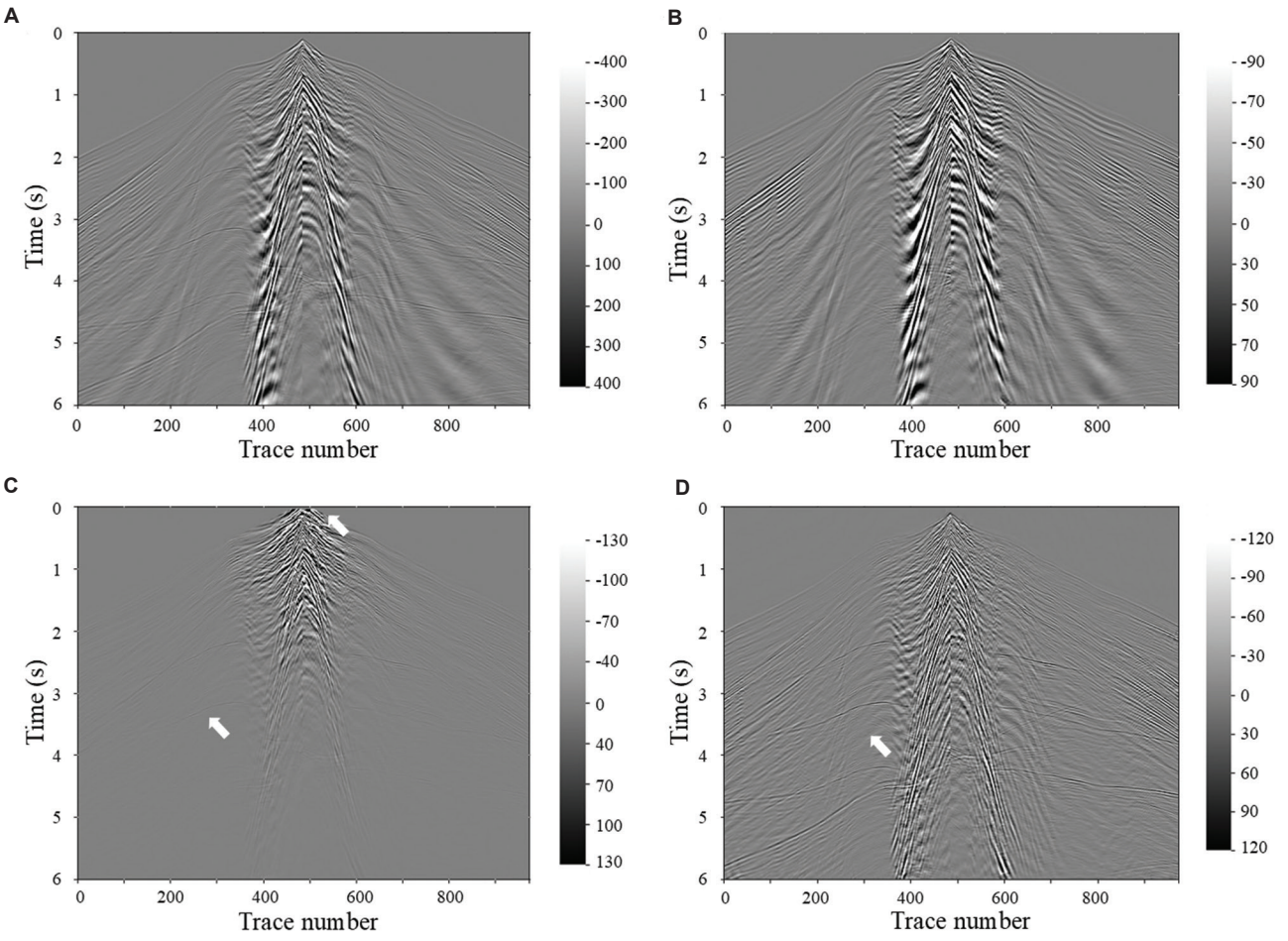

Figure 9. Denoising results of synthetic data 2. (A) Original data. (B) Interferometric data. (C) Results after matching filtering. (D) Denoising results of a convolutional autoencoder constrained by local scattered wave characteristics (LSC-CAE).

After applying the matched filter for noise suppression (Figure 9C), the SNR improved slightly. However, a portion of the body waves was attenuated during the filtering process (white arrows). When processed using LSC-CAE (Figure 9D), the body wave energy was effectively restored, the SNR was further enhanced, and both waveform continuity and phase information were preserved, demonstrating the effectiveness of the proposed method.

3.3. Field data application

The synthetic data experiments demonstrated that the proposed self-supervised learning approach enabled the predicted scattered waves to match the energy of the true wavefield. The LSC-CAE was then applied to field seismic data following the workflow described in Section 2.4. Due to the complexity of real seismic data, a larger dataset was required to capture the data characteristics effectively. Therefore, 60-shot records from mountainous terrain were selected as training samples for the network. First, shot-receiver interferometry was performed to obtain the predicted scattered waves. The 60-shot records were then partitioned into 15,000 data pairs, each of size 256 × 256. These predicted scattered wavefield slices and full-wavefield data slices were used as network inputs and labels, respectively, for training.

Unlike synthetic data, the predicted scattered waves generated during field seismic data processing exhibit significant energy differences compared to the full-wavefield data. Thus, when selecting coefficients in the loss function, it was crucial to consider energy recovery and maintain a proper balance among the different loss components. The denoising results for the first field dataset are shown in Figure 10. Figure 10A and F displays the original data and the predicted scattered waves. The shot data shows typical high-frequency scattering and low-frequency, large-curvature side NSWs (black arrow in Figure 10A). The interferometric results demonstrate that structured external interferences and scattered surface waves were well predicted without introducing additional noise. To comprehensively demonstrate the denoising superiority of the LSC-CAE, we compared its denoised results (Figure 10B) and the extracted noise (Figure 10G) with those obtained from the RS-CAE trained with a conventional MSE loss (Figure 10D and 10I), the matched-filtering method (Figure 10E and 10J), and the CAE using a composite loss function (Figure 10C and 10H). The results demonstrate that the LSC-CAE produced cleaner denoised sections and almost completely suppressed the scattered-wave groups (red arrows). In contrast, the conventional CAE and matched filtering approach failed to recover the predicted noise to its original amplitude,

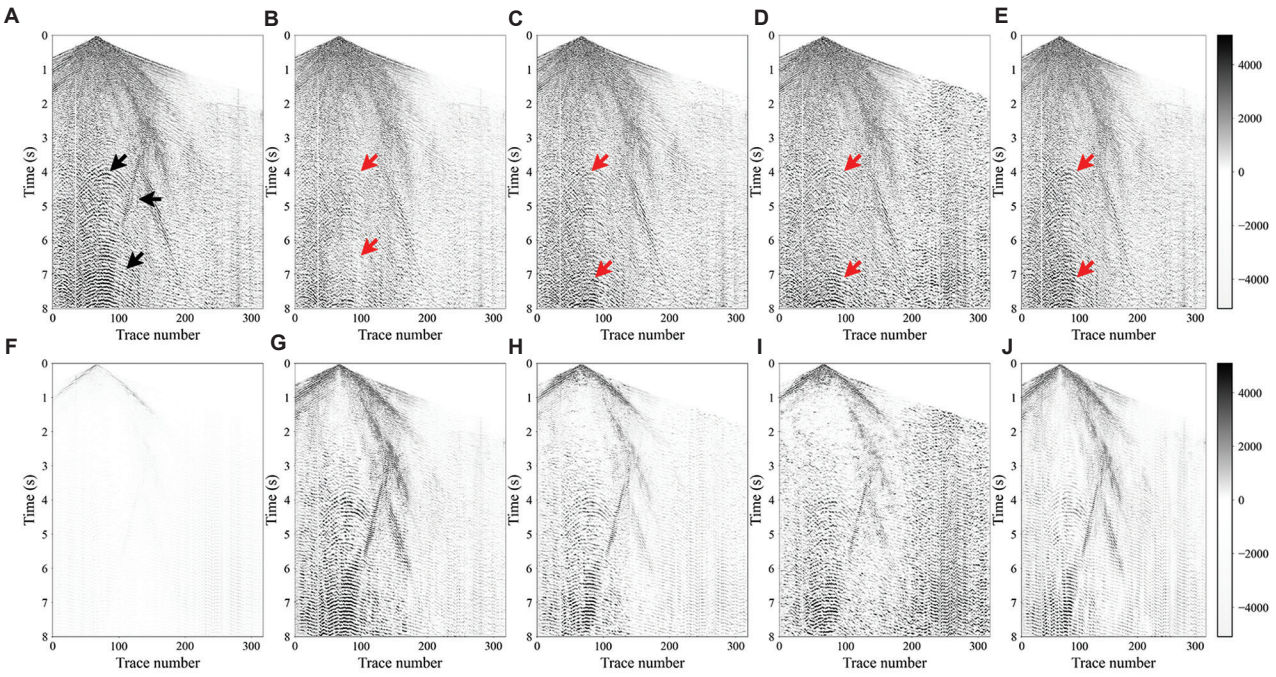

Figure 10. Denoising results of field data 1. (A) Original data. (B) LSC-CAE denoising result (dynamic parameter adjustment strategy). (C) CAE denoising result (dynamic parameter adjustment strategy). (D) LSC-CAE denoising result (λ_1=1). (E) Matched filtering result. (F) Interferometric data. (G) Noise removed by the LSC-CAE. (H) Noise removed by the CAE. (I) Noise removed by the LSC-CAE (λ_1=1). (J) Noise removed by matched filtering
Abbreviations: CAE: Convolutional autoencoder; LSC: Local scattered wave.

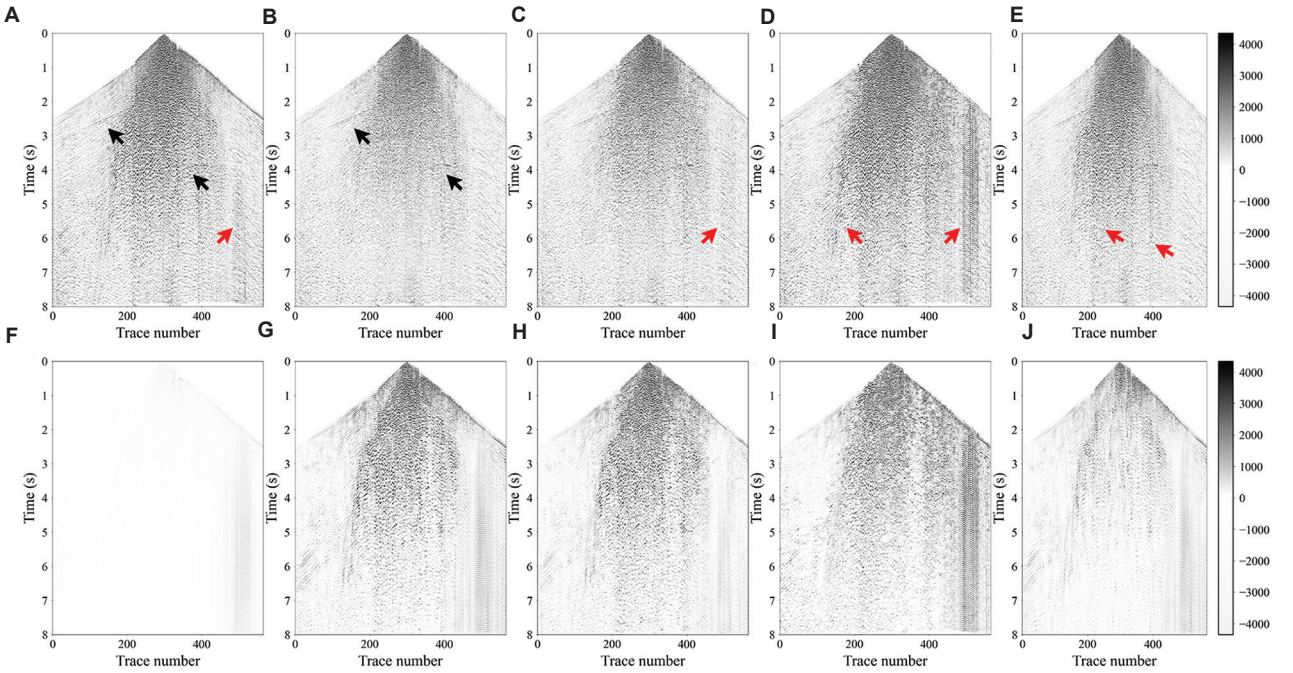

Figure 11. Denoising results of field data 2. (A) Original data. (B) LSC-CAE denoising result (dynamic parameter adjustment strategy). (C) CAE denoising result (dynamic parameter adjustment strategy). (D) LSC-CAE denoising result (λ_1=1). (E) Matched filtering result. (F) Interferometric data. (G) Noise removed by the LSC-CAE. (H) Noise removed by the CAE. (I) Noise removed by the LSC-CAE (λ_1=1). (J) Noise removed by matched filtering
Abbreviations: CAE: Convolutional autoencoder; LSC: Local scattered wave.

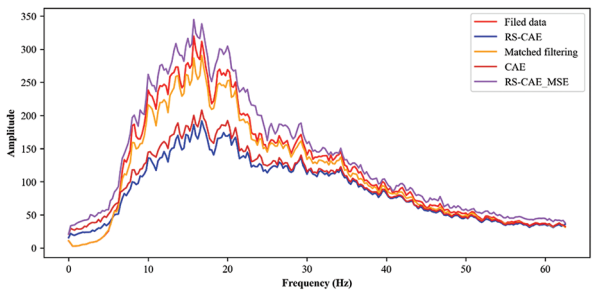

Figure 12. Spectral analysis of the surface wave region after denoising using different methods
Abbreviations: CAE: Convolutional autoencoder; RS-CAE: Residual-self-attention convolutional autoencoder; MSE: Mean squared error.

leaving significant scattered-wave energy in their denoised outputs. Moreover, when only the traditional MSE loss was used (λ_1=1), the network tended to focus primarily on amplitude restoration while neglecting the underlying structural characteristics of the waveforms, ultimately limiting its denoising performance.

For the second field dataset (Figure 11), the data denoised by the LSC-CAE (Figure 11B and 11G) showed a more pronounced attenuation of surface-wave energy compared with the CAE using only a dynamic loss function (Figure 11C and 11H), the LSC-CAE with λ_1=1 (Figure 11D and 11I), and the matched filter (Figure 11E and 11J). Since the scattered-wave energy in field data primarily originates from surface waves, this result indicates that, in addition to effectively suppressing large-curvature low-frequency scattering, the LSC-CAE performed better in complex scenarios where multiple scattered waves were superimposed. Furthermore, the regular interference near trace 500 was also effectively suppressed.

A spectral analysis was performed on the surface wave region to evaluate the capability of different denoising methods in suppressing NSWs. The surface wave energy was mainly concentrated in the 10-30 Hz frequency range, with a peak at 20 Hz. As shown in Figure 12, the RS-CAE method achieved the best performance, reducing the peak amplitude from 320 to 190, corresponding to a denoising rate of approximately 40%. The CAE method with local wavefield feature constraints performed moderately, while the network trained with only MSE (λ_1=1) failed to effectively suppress the surface-wave noise. These results indicate that the RS-CAE network, incorporating residual structures and self-attention mechanisms along with local wavefield characteristic constraints, exhibited a significant advantage in NSW suppression.

Both synthetic and field data examples demonstrated that introducing the local scattered-wave characteristics

loss enhances the network's ability to extract features in complex wavefields. During the training process, the weights of L_{mse}, L_{LSC}, and L_S were dynamically adjusted in phases to meet the learning objectives of each stage. This strategy guided the network to gradually balance energy-related features and waveform structural characteristics while maintaining training stability.

4. Conclusion

We proposed a self-supervised deep learning strategy that eliminated the need for additional clean sample data as labels. During dataset preparation, seismic interferometry was utilized to predict noise from the original noisy data, generating high-quality training samples for the network. Throughout the network training process, a local scattered-wave feature loss term and regularization terms were integrated with the conventional MSE, allowing the network to autonomously focus on waveform characteristics. A dynamic adjustment strategy was employed to control the network's feature extraction priorities at different stages, ultimately enabling the model to learn both energy discrepancies and the waveform characteristics of scattered waves.

Traditional matched filters, relying on fixed parameters, often struggle to achieve satisfactory results in complex survey areas, whereas a trained network can adaptively handle varying noise conditions, providing stable denoising performance. Although network training can be time-consuming, once completed, inference is extremely fast, with a single-shot gather taking less than a second, substantially improving overall efficiency. During training, only a small portion of data from each survey area or model is needed to satisfy the network's sample requirements. Tests on both synthetic and field seismic data show that incorporating local scattered-wave characteristic constraints enables the network to effectively achieve the intended denoising objectives.

Acknowledgment

None.

Funding

This work is supported by the China National Petroleum Corporation Science and Technology Project, "Key Technologies for Seismic Physical Modeling, Acquisition, and Processing in the Tajik Basin" (2024DJ9905).

Conflict of interest

The authors declare that they have no conflict of interest.

Author contributions

Conceptualization: Daling Hou
Formal analysis: Daling Hou
Funding acquisition: Meng Li
Investigation: Daling Hou
Methodology: Daling Hou
Visualization: Daling Hou
Writing–original draft: Daling Hou
Writing–review & editing: Daling Hou, Jixiang Xu

Availability of data

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

References

1. Wu X, Li Y, Zhang M, He J, Long Z. Causes and countermeasures for low signal-to-noise ratio in seismic data in complex areas (part 1) - surface scattered waves are the root cause of low signal-to-noise ratio in seismic data. *Nat Gas Ind*. 2012;32(1):27-32. [Article in Chinese].

   doi: 10.3787/j.issn.1000-0976.2012.01.005

2. Shao J, Tang J, Sun C. Progress of seismic wave scattering theory and application. *Prog Geophys*. 2016;31(1):334-343.

   doi: 10.6038/pg20160139

3. Campman XH, Wijk Kvan, Scales JA, Herman GC. Imaging and suppressing near-receiver scattered surface waves. *Geophysics*. 2005;70(2):V21-V29.

   doi: 10.1190/1.1884831

4. Liu Y, Hu G, Fan T. Suppression of strong scattered waves using the transverse component. *J Appl Geophys*. 2015;112:226-235.

   doi: 10.1016/j.jappgeo.2014.11.021

5. Almuhaidib AM, Toksöz MN. Suppression of near-surface scattered body-to-surface waves: A steerable and nonlinear filtering approach. *Geophys Prospect.* 2016;64(2):392-405.

   doi: 10.1111/1365-2478.12275

6. Xu J, McLean JX, Song J. A near-surface seismic scattered wave separation method. *Pet Explor Dev*. 2014;41(6):771-777.

   doi: 10.1016/S1876-3804(14)60091-4

7. Halliday D, Bilsby P, West L, Kragh E, Quigley J. Scattered ground-roll attenuation using model-driven interferometry. *Geophys Prospect*. 2015;63(1):116-132.

   doi: 10.1111/1365-2478.12165

8. Xu J, Dong S, Cui H, Zhang Y, Hu Y, Sun X. Near-surface scattered waves enhancement with source-receiver interferometry. *Geophysics*. 2018;83(6):Q49-Q69.

   doi: 10.1190/geo2017-0806.1

9. Zhou P, Li P, Wang Y, Wei Z. *Principle and Practice of Deep Convolution Neural Network*. Beijing, China: Publishing House of Electronics Industry. 2020. p. 17-18.

10. Yu S, Ma J, Wang W. Deep learning for denoising. *Geophysics*. 2019;84:V333-V350.

doi: 10.1190/geo2018-0668.1

11. Yu S, Yang W, Li H, Wang H, Ma H. Scattered ground roll intelligent attenuation based on deep learning. *Chin Sci Bull (Chinese Version)*. 2021;66(18):2343-2354.

doi: 10.1360/TB-2020-0564

12. Liu W, Lü Q, Chen Q, *et al*. A synthetic denoising algorithm for full-waveform induced polarization based on deep learning. *Geophysics*. 2023;88(1):WA267-WA279.

doi: 10.1190/geo2022-0234.1

13. Sui Y, Wang X, Ma J. Deep unfolding dictionary learning for seismic denoising. *Geophysics*. 2023;88(1):WA129-WA147.

doi: 10.1190/geo2022-0198.1

14. Li X, Qi Q, Yang Y, Duan P, Cao Z. Removing abnormal environmental noise in nodal land seismic data using deep learning. *Geophysics*. 2024;89(1):WA143-WA156.

doi: 10.1190/geo2023-0143.1

15. Song H, Gao Y, Chen W, Zhang X. Seismic data denoising based on convolutional autoencoder. *Oil Geophys Prospect.* 2020;55(6):1210-1219+1160-1161. [Article in Chinese].

doi: 10.13810/j.cnki.issn.1000-7210.2020.06.006

16. Shao D, Zhao Y, Li Y, Li Y. Noisy2Noisy: Denoise pre-stack seismic data without paired training data with labels. *IEEE Geosci Remote Sens Lett.* 2022;19:1-5.

doi: 10.1109/lgrs.2022.3145835

17. Wu T, Meng T, Liu H, Li T. A seismic random noise suppression method based on self-supervised deep learning and transfer learning. *Acta Geophys*. 2024;72:655-671.

doi: 10.1007/s11600-023-01105-5

18. Wang K, Hu T, Wang S, Wei J. Seismic multiple suppression based on a deep neural network method for marine data. *Geophysics*. 2022a;87(4):V341-V365.

doi: 10.1190/geo2021-0206.1

19. Wang K, Hu T, Zhao B. An unsupervised deep neural network approach based on ensemble learning to suppress seismic surface-related multiples. *IEEE Trans Geosci Remote Sens*. 2022b;60:1-14.

doi: 10.1109/TGRS.2022.3225809

20. Wang K, Hu T, Zhao B, Wang S. Surface-related multiple attenuation based on a self-supervised deep neural network with local wavefield characteristics. *Geophysics*. 2023a;88(5):V387-V402.

doi: 10.1190/geo2022-0599.1

21. Wang K, Hu T, Zhao B, Wang S. An unsupervised learning method to suppress seismic internal multiples based on adaptive virtual events and joint constraints of multiple deep neural networks. *IEEE Trans Geosci Remote Sens*. 2023b;61:1-18.

doi: 10.1109/tgrs.2023.3243106

22. Sun K, Li Z, Wang Y, Ma J, Dai X, Li B. Adaptive subtraction based on expanded multichannel U-Net with multi-pattern multiple model for surface-related multiple removal. *IEEE Trans Geosci Remote Sens*. 2025;63:5912511.

doi: 10.1109/tgrs.2025.3565358

23. Qi Z, Li Z, Sun N, Li X, Li B, Wang Y. Semi-supervised interpretable FISTA-Net for adaptive subtraction and removal of seismic multiples. *IEEE Trans Geosci Remote Sens*. 2025;63:5912214.

doi: 10.1109/tgrs.2025.3564071

24. Wang Z, Bovik AC. A universal image quality index. *IEEE Signal Proc Let*. 2005;9(3):81-84.

doi: 10.1109/97.995823

25. Almadani M, Waheed U, Masood M, Chen Y. Dictionary learning with convolutional structure for seismic data denoising and interpolation. *Geophysics*. 2021;86(5):V361-V374.

doi: 10.1190/geo2019-0689.1

26. Sun Y, Williamson P. Seismic data interpolation via frequency-constrained 3D inception Unet. In: *Second International Meeting for Applied Geoscience and Energy*; 2022. p. 1679-1683.

doi: 10.1190/image2022-3751145.1

27. Simon J, Fabien-Ouellet G, Gloaguen E, Khurjekar I. Hierarchical transfer learning for deep learning velocity model building. *Geophysics*. 2023;88(1):R79-R93.

doi: 10.1190/geo2021-0470.1

28. Schuster GT. Theory of Daylight/Interferometric Imaging: Tutorial. In: *63rd EAGE Annual International Conference and Exhibition*, Amsterdam, Netherlands; 2001.

doi: 10.3997/2214-4609-pdb.15.A-32

29. Schuster GT. Imaging the Most Bounce out of Multiples. In: *65th EAGE Conference and Exhibition*, Stavanger, Norway; 2003.

doi: 10.3997/2214-4609.201405725

30. Schuster GT, Yu J, Sheng J, Rickett J. Interferometric/daylight seismic imaging. *Geophys J Int*. 2004;157(2):838-852.

doi: 10.1111/j.1365-246X.2004.02251.x

31. Schuster GT. Fermat's interferometric principle for multiple reflection tomography. *Geophys Res Lett*. 2005a;32(12):L12303.

doi: 10.1029/2005gl022351

32. Schuster GT. Fermat's interferometric principle for target-oriented traveltime tomography. *Geophysics*. 2005b;70(4):U47-U50.

doi: 10.1190/1.1997368

33. Schuster GT, Zhou M. A theoretical overview of model-based and correlation-based redatuming methods. *Geophysics*. 2006;71(4):103-110.

doi: 10.1190/1.2208967

34. Wapenaar K, Draganov D, Snieder R, Campman X, Verdel A. Tutorial on seismic interferometry: Part 1-basic principles and applications. *Geophysics*. 2010a;75(5):75A195-75A209.

doi: 10.1190/1.3457445

35. Wapenaar K, Slob E, Snieder R, Curtis A. Tutorial on seismic interferometry: Part 2-Underlying theory and new advances. *Geophysics*. 2010b;75(5):75A211-75A227.

doi: 10.1190/1.3463440

36. Wapenaar K, Ruigrok E, Neut J, Draganov D. Improved surface-wave retrieval from ambient seismic noise by multidimensional deconvolution. *Geophys Res Lett*. 2011;38(1):L01313.

doi: 10.1029/2010GL045523

37. Curtis A, Halliday D. Source-receiver wave field interferometry. *Phys Rev E Stat Nonlin Soft Matter Phys*. 2010;81(4):046601.

doi: 10.1103/physreve.81.046601

38. He K, Zhang K, Ren S. Deep Residual Learning for Image Recognition. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society; 2016. p. 770-778.

doi: 10.1109/CVPR.2016.90

39. Vaswani A, Shazeer N, Parmar N, *et al.* Attention is All You Need. In: *Advances in Neural Information Processing Systems* [Preprint]. Vol. 30; 2017. p. 5998-6008.

doi: 10.48550/arXiv.1706.03762

40. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: From error visibility to structural similarity. *IEEE Trans Image Process*. 2004;13(4):600-612.

doi: 10.1109/TIP.2003.819861

41. Goodfellow I, Bengio Y, Courville A. *Deep Learning*. Cambridge, MA: MIT Press; 2016.