ACCSCIENCE PUBLISHING

Journal of Seismic Exploration

ARTICLE

U-STDRNet: A unified model integrating swin transformer and residual dense network for seismic image super-resolution and denoising

Mingliao Wu[1], Juan Wu[1]*, Min Bai[1], Haiyu Li[1], Zhixian Gui[1], and Guangtan Huang[2]*

[1]The Key Laboratory of Exploration Technology for Oil and Gas Resources of Ministry of Education, College of Geophysics and Petroleum Resources, Yangtze University, Wuhan, Hubei, China

[2]The State Key Laboratory of Geomechanics and Geotechnical Engineering, Institute of Rock and Soil Mechanics, Chinese Academy of Sciences, Wuhan, Hubei, China

(This article belongs to the *Special Issue: Advanced Artificial Intelligence Theories and Methods for Seismic Exploration*)

Abstract

Enhancing seismic image resolution while effectively suppressing noise remains a critical challenge in accurately characterizing subsurface geological structures for oil and gas exploration. Traditional methods often fail to balance the recovery of fine details with robustness to noise, particularly in complex geological settings or under high-noise conditions. This study proposes a deep learning-based joint model, U-Net Shifted Window (Swin) Transformer-based dense residual network (U-STDRNet). The model integrates the global modeling capability of the Swin Transformer, the hierarchical feature reuse mechanism of the residual dense network, and an attention-guided strategy to jointly perform seismic image super-resolution and denoising. Built upon the U-Net encoder-decoder architecture, the model embeds Swin Transformer-based convolutional residual blocks. These blocks employ both a feature fusion block with the Swin Transformer and a feature fusion block with a convolutional neural network to effectively capture stratigraphic continuity and enhance detailed features such as fault edges. Residual dense blocks further improve weak signal recovery (e.g., thin-layer interfaces) through dense residual connections. Furthermore, the convolutional block attention module is integrated into skip connections, employing a dual-channel spatial weighting mechanism to suppress noise and emphasize key geological regions. Experimental results and field-data experiments demonstrate that U-STDRNet achieves a higher peak signal-to-noise ratio than the traditional U-Net. In addition, the model successfully restores fault and fold continuity details while exhibiting superior noise suppression compared to existing methods.

Keywords: U-Net Swin Transformer-based dense residual network; Seismic image super-resolution; Seismic image denoising; Deep learning models; Swin transformer; Residual dense networks; Convolutional block attention module

***Corresponding authors:**
Juan Wu
(wujuan@yangtzeu.edu.cn)
Guangtan Huang
(gthuang@whrsm.ac.cn)

Citation: Wu M, Wu J, Bai M, Li H, Gui Z, Huang G. U-STDRNet: A unified model integrating swin transformer and residual dense network for seismic image super-resolution and denoising. *J Seismic Explor*. 2026;35(1):14-32.
doi: 10.36922/JSE025400081

Received: September 30, 2025

Revised: November 5, 2025

Accepted: November 5, 2025

Published online: December 1, 2025

1. Introduction

In the field of oil and gas exploration, seismic data processing and interpretation are crucial for constructing models of underground geological structures. Seismic imaging reveals these structures by analyzing seismic wave propagation characteristics; however, real-world data often suffer from low resolution and high noise levels. These problems not only obscure geological features and cause loss of detail but may also mask important geological signals. Consequently, noise suppression, data restoration, and resolution enhancement are essential in seismic data processing. The pursuit of higher resolution and lower noise in seismic imaging has long been a focus in geophysics.

Traditional approaches can be broadly categorized into acquisition-based and processing-based methods. Acquisition-based techniques, such as high-density acquisition,[1,2] and broadband seismic acquisition,[3-5] aim to enhance data quality at the source. However, these techniques significantly increase the financial and computational costs of seismic surveys, which limits their widespread application. Processing-based techniques are widely used for denoising and enhancement after data acquisition. However, conventional methods face significant limitations. Filtering techniques often attenuate effective signals whose frequency spectra overlap with that of the noise.[6,7] Wavelet transform can lead to over-smoothing and loss of detailed features.[8,9] Statistical methods, such as those based on principal component analysis, rely on assumptions about data distributions that are often violated by complex real-world seismic data.[10] Although some advanced traditional methods, such as those based on low-rank and sparse constraints, have been developed to improve performance, they often entail high computational complexity or rely on assumptions that may not always hold for complex field data.[11-13]

To overcome the limitations of traditional methods, deep learning (DL) has emerged as a powerful alternative, offering data-driven learning capabilities and superior performance in capturing complex, nonlinear features.[14] Early applications primarily treated seismic image enhancement as a computer vision task, where convolutional neural networks (CNNs) were adopted to learn the mapping between low-quality and high-quality seismic data. Models such as the denoising CNN[15] and U-Net[16] became popular baselines. For instance, Wang *et al.*[17] used CNNs with data augmentation for random noise suppression. Numerous other DL methods based on CNN have also been proposed for denoising seismic data.[18-26] These studies demonstrated CNNs' potential but also revealed limitations in generalization and preservation of fine geological structures when dealing with strong noise or features beyond their training distribution. Subsequent research evolved to address specific seismic challenges, such as integrating geophysical priors into networks. For instance, Zhao and Rao[24] proposed a structure-preserving network to protect fault and stratigraphic interfaces during denoising. Others focused on signal compensation; for example, Liu *et al.*[27] used DL to compensate for absorption effects, accelerating high-resolution imaging. Peng *et al.*[28] combined a high-resolution seislet transform with a DL framework to enhance signal recovery in low signal-to-noise ratio data. Furthermore, to mitigate the reliance on vast amounts of labeled data—often scarce in geophysics—self-supervised[29] and unsupervised learning methods have been developed,[30-34] showing promise in learning from unlabeled or weakly labeled field data. Alongside these DL advances, sophisticated traditional methods, such as optimized wavelet transforms and convolutional sparse coding, remain active areas of research.[35,36] Furthermore, specific application-driven denoising approaches like Q-factor compensation highlight the need for tailored solutions for particular seismic processing tasks.[37] A significant advancement came with the introduction of attention mechanisms, which allow networks to focus on semantically critical features. The convolutional block attention module (CBAM), for example, guides networks to emphasize important channels and spatial regions (e.g., fault lines) while suppressing irrelevant noise.[38] More recently, Transformer architectures have been explored for their superior ability to model long-range dependencies. The Shifted Window (Swin) Transformer, with its hierarchical windowing mechanism, is particularly suited for seismic data, as it can efficiently capture the global continuity of seismic events—a key aspect where local CNNs often fail. Ding *et al.*[39] developed a Swin-Conv-UNet architecture, demonstrating the ongoing integration of Transformers and CNNs. Gao *et al.*[40] further advanced this by integrating the Swin Transformer with CNNs for simultaneous seismic data denoising and interpolation, effectively leveraging both global and local features.

Despite these advancements, a critical gap remains. Many existing studies address super-resolution and denoising as separate tasks or propose models that are adaptations of generic computer vision architectures. There is a pressing need for an end-to-end solution explicitly designed for the joint task—seamlessly integrating global contextual modeling (for event continuity), local feature refinement (for detail recovery), and adaptive attention (for geological structure preservation)—within a unified framework tailored to the intricacies of seismic data. Based on the above background and challenges, this paper proposes a novel DL model, U-Net Swin Transformer-

based dense residual network (U-STDRNet), for the joint task of seismic image super-resolution and denoising. The main contributions of this work are summarized as follows:

(i) Novel network architecture for seismic-specific challenges: We propose a dedicated U-Net-based architecture that organically integrates the Swin Transformer (Shifted Window Transformer convolutional residual blocks [STCRB]), residual dense blocks (ResDenB), and the CBAM. This is not a mere combination but a co-design where each module addresses a specific deficiency in seismic image processing. The STCRB captures long-range stratigraphic continuity through its window-based self-attention mechanism, effectively mitigating the issue of event discontinuity caused by noise or missing traces. The ResDenB enhances the recovery of weak yet critical signals, such as thin-layer interfaces, through dense feature reuse, while the CBAM embedded in skip connections acts as a feature filter, dynamically suppressing noise while highlighting spatially important geological structures like faults and folds.

(ii) Task-specific optimization: We introduce a mixed loss function combining L_1 loss and multiscale structural similarity index (MS-SSIM) loss, specifically tailored for seismic imagery. This combination significantly reduces the over-smoothing of geological edges commonly produced by pixel-wise losses alone, thereby better preserving the structural integrity and perceptual quality essential for geological interpretation.

(iii) Comprehensive validation on synthetic and field data: We extensively validate the proposed model on both synthetic datasets containing realistic geological structures (folds and faults) and two real field datasets. The results demonstrate that U-STDRNet not only achieves superior quantitative metrics (peak signal-to-noise ratio [PSNR], SSIM) but also exhibits significantly enhanced performance in visual quality, particularly in restoring fine details and improving the continuity of geological structures under high-noise conditions, showcasing strong generalization capability for practical applications.

2. Methods

Image super-resolution and denoising are low-level computer vision tasks that can be addressed using similar processing frameworks. First, the basic principles underlying these two problems are analyzed. For seismic super-resolution, the relationship between high-resolution and low-resolution seismic images is expressed as:

$$I_h = D(I_l, \theta) \tag{1}$$

Where I_l and I_h represent low-resolution and high-resolution seismic images, respectively; D represents the down-sampling function, and θ is the parameter of this function. The goal of super-resolution is to obtain I_h from the low-resolution seismic image using a CNN or other methods to approximate the high-resolution image.

For seismic denoising, if I_p and I_n represent the clean and noisy seismic images, respectively, the relationship between them can be expressed as:

$$I_n = I_p + n, \tag{2}$$

Where n is the noise added to I_p. The goal is to estimate the noise distribution n, and then subtract it from the noisy seismic image I_n, yielding an output image approximating the clean seismic image I_p.

In this work, these two problems were solved simultaneously. The noiseless high-resolution seismic image and the noisy low-resolution seismic image were used as the real image I_{output} and the input image I_{input}, respectively, following the formula:

$$I_{output} = N(I_{input}, \sigma) \tag{3}$$

Where N represents the network used, and σ denotes the parameter of the network. The goal is to achieve simultaneous super-resolution and denoising of seismic images using a neural network.

3. Network architecture

3.1. Overall architecture

The overall framework of the proposed U-STDRNet is illustrated in Figure 1. It is built upon a U-Net encoder-decoder backbone, which is renowned for its ability to preserve detailed spatial information through skip connections. To address the specific challenges of seismic image processing, we introduced three key innovations into this backbone:

(i) The STCRB modules replace standard convolutional blocks in the encoder and are designed to capture long-range global dependencies that maintain stratigraphic continuity.

(ii) The ResDenB modules are integrated within the decoder path, leveraging dense connections to maximize feature reuse, enhancing the recovery of fine details and weak signals (e.g., thin layers).

(iii) CBAM modules are embedded in the skip connections, acting as adaptive feature gates, dynamically weighting channel and spatial features to suppress noise in irrelevant areas and highlight crucial geological structures (e.g., faults) before merging encoder and decoder features.

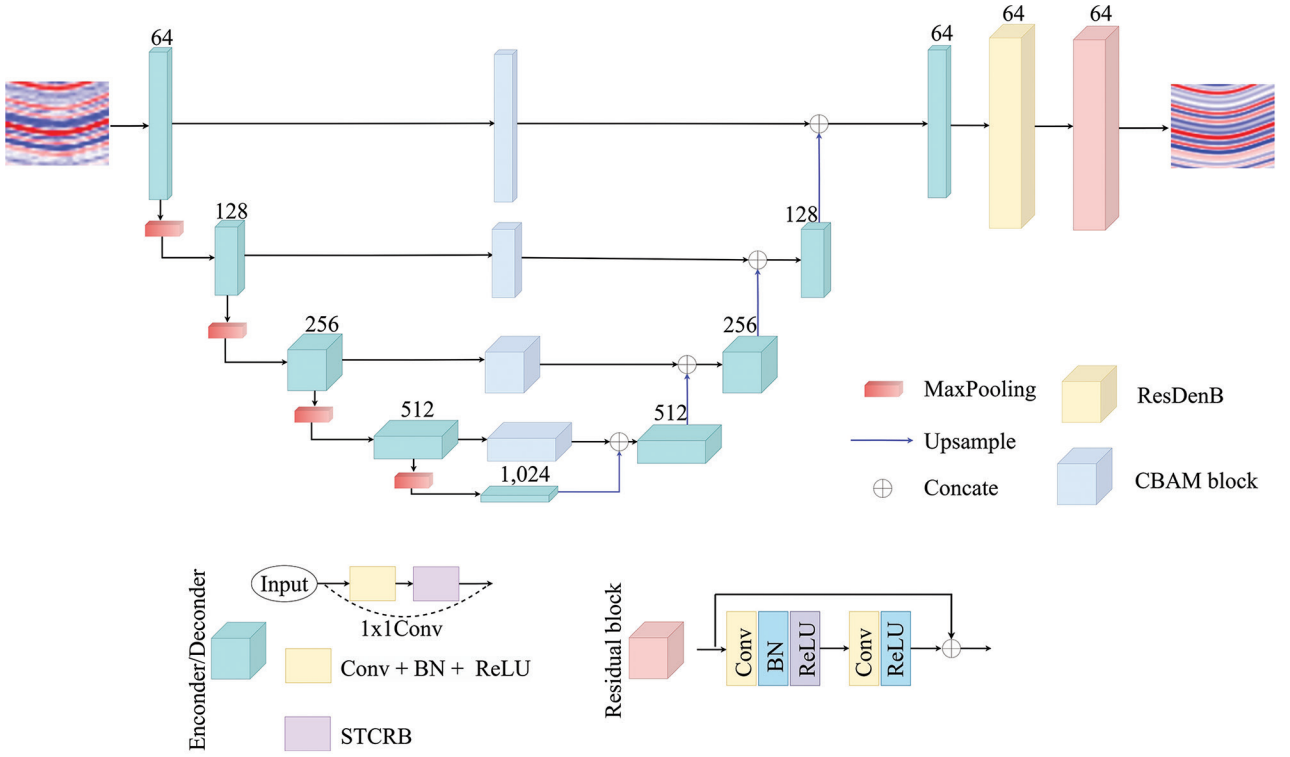

Figure 1. The structure of the U-Net Swin Transformer-based dense residual network
Abbreviations: BN: Batch normalization; CBAM: Convolutional block attention module; Conv: Convolution; ReLU: Rectified linear unit; ResDenB: Residual dense block; STCRB: Shifted Window Transformer convolutional residual blocks.

This synergistic design ensures that global context, local details, and feature importance are jointly optimized throughout the network for the combined task of seismic image super-resolution and denoising.

3.2. STCRB

Traditional CNNs struggle to model the long-range continuity of seismic events, especially in the presence of noise or missing traces. The STCRB module is designed to overcome this limitation by integrating the Swin Transformer's global self-attention mechanism, which effectively captures the contextual relationships across the entire seismic section, thereby maintaining stratigraphic continuity. This network is cascaded with multiple feature fusion blocks (FFB) to enhance denoising performance. The FFB leverages local similarity in seismic data and integrates features with varying amplitudes. Each FFB consists of a Swin Transformer module (FFBST) for global feature extraction and a convolution module (FFBC) for local feature extraction, which are detailed below. Global features capture time-dependent information, maintaining event continuity, while local features focus on highly correlated information, preserving weak signals. The module is shown in Figure 2.

(i) FFBST module: This module aims to address the limitation of traditional CNNs in global feature extraction. Because CNNs extract features through point-to-point local learning, they may lose seismic event continuity when dealing with missing seismic traces or noise and interference. Therefore, the FFBST module was designed to enhance global feature extraction. It integrated several components: layer normalization (LN), window-based multi-head self-attention (WMSA), a residual structure, and a multi-layer perceptron (MLP). This integration significantly enhanced global feature modeling. LN normalized the input features along the channel dimension, stabilizing training and accelerating convergence by ensuring consistent feature distribution across channels. WMSA was the core component of the FFBST module and was used to capture global features. The specific operation is as follows: suppose the input feature dimension is [10, 10, 32]. It is divided into multiple 8×8 window partitions, and the divided feature dimensions are ([8×8], 8×8, 32), where (8×8) represents the number of windows, 8×8 represents the size of each window, and 32 represents the number of channels. Subsequently, through

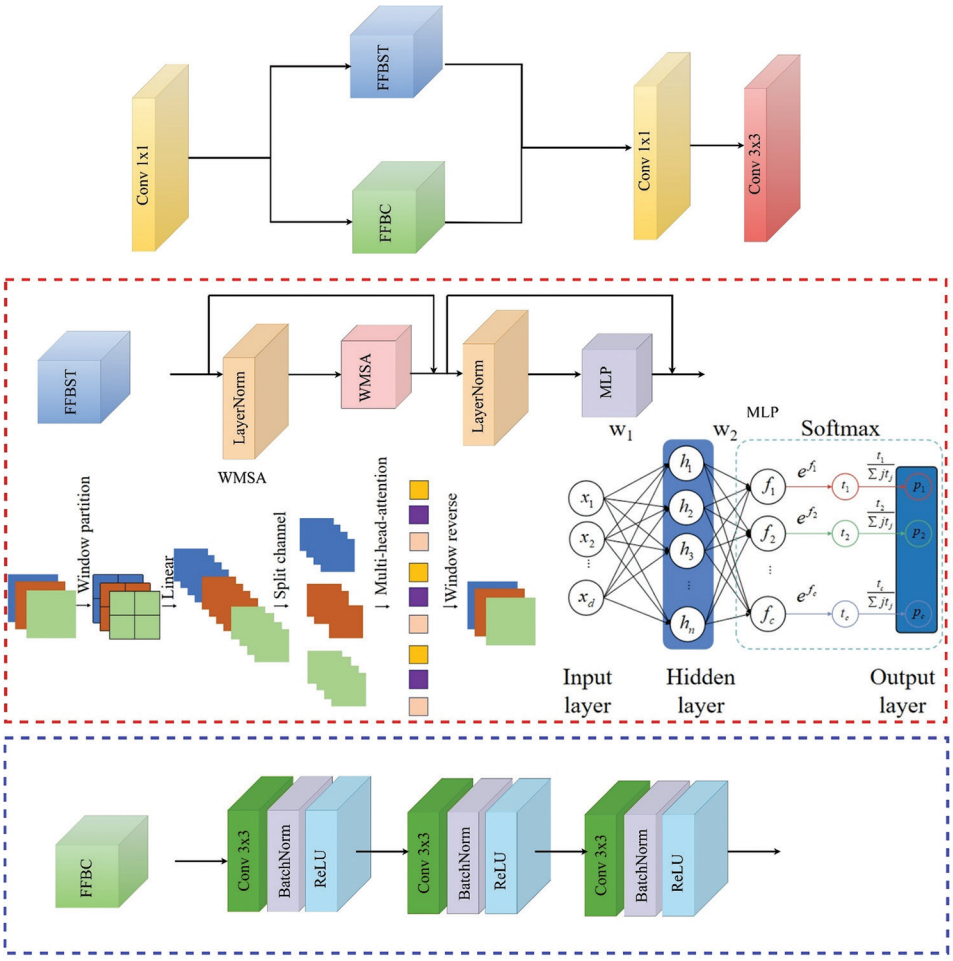

Figure 2. The structure of STCRB

Abbreviations: BatchNorm: Batch normalization; Conv: Convolution; FFBC: Feature fusion block with CNN; FFBST: Feature fusion block with Swin transformer; LayerNorm: Layer normalization; MLP: Multi-layer perceptron; ReLU: Rectified linear unit; STCRB: Shifted window transformer convolutional residual blocks; WMSA: Window-based multi-head self-attention.

channel division and dimension transformation, the feature dimension is further adjusted to (3, [108 × 108], 3, 8 × 8, 96) to meet the computational requirements of a multi-head attention mechanism. Here, the first position "3" represents three matrices (query, key, value) used for adaptive calculation, and the third position "3" represents the number of attention heads used in the multi-head attention mechanism. After the multi-head attention calculation is completed, the window features are restored to the original input size to maintain spatial consistency of the features. A residual structure adds the input features to the WMSA-processed features, mitigating gradient vanishing and preserving original feature information. The MLP is used to further fuse and refine features. Its structure includes an input layer, a hidden layer, and an output layer. The dimensions of the input layer are [10, 10, 32], the hidden layer [10, 10, 128], and the output layer [10, 10, 32]. Through a nonlinear activation function (such as Gaussian error linear unit), the MLP enhances the representational capacity of the features. The entire FFBST can be expressed as:

$$FFB\text{–}ST(X) = MLP(WMSA(LN(X))) + X \quad (4)$$

Where X is the input feature, LN represents LN, $WMSA$ represents the WMSA mechanism, MLP represents a MLP, and the final residual connection ensures the integrity of the feature information.

(ii) FFBC module: Although CNNs have limitations in extracting global features, they excel at extracting local features by leveraging high correlation through shared weights. To fully exploit this advantage, the FFBC module was designed. The module enhanced local feature extraction by stacking three identical

sub-modules. Each sub-module comprised a 3 × 3 convolution (Conv 3 × 3), batch normalization (BN), and a rectified linear unit (ReLU) activation function.[41] The convolution kernel size of 3 × 3 was used to extract local information from the input features. Through sliding window operations, the convolutional layer captured local spatial patterns within the feature map, maintaining translation invariance. BN is used to standardize the features of the convolution output, speeding up the training process and improving model stability. BN reduced internal covariate shift by normalizing the mean and variance of features within each batch. The ReLU activation function is defined as

$$ReLU(x) = max(0, x) \tag{5}$$

And is used to introduce nonlinearity. ReLU helps alleviate the problem of vanishing gradients and accelerates network convergence.

The FFBC module is formed by stacking three identical submodules, and the structure of each submodule can be expressed as:

$$Module(x) = ReLU(BN(Conv.3 \times 3.(x))) \tag{6}$$

Where X is the input feature, $Conv3 \times 3$ represents the 3 × 3 convolution operation, BN represents BN, and $ReLU$ represents the activation function.

The overall process of the FFBC module can be expressed as:

$$FFBC(x) = module(module(module(x))) \tag{7}$$

By extracting and fusing local features across three layers, the FFBC module effectively enhances the network's ability to model local details. This provides richer feature representations for subsequent global feature extraction and tasks like denoising and super-resolution.

3.3. Residual dense blocks

Weak seismic signals, such as those from thin-layer interfaces, are often lost in deep networks. The ResDenB module was introduced to address this issue. By combining residual and dense connections, it facilitates the flow of information and gradients throughout the network, ensuring that subtle features are preserved and enhanced during the decoding process. It acts as a core component for seismic data denoising and missing-trace recovery. Its core design integrates residual learning and dense connections, allowing full utilization of multi-level feature information and thereby improving network performance. As shown in Figure 3, the key component in ResDenB is the residual dense block, which combines the advantages of residual connections and dense connections to effectively extract and transmit feature information. Residual blocks and dense blocks are defined as follows:

$$x \rightarrow x + f(x), \ldots (residual\ block) \tag{8}$$

$$x \rightarrow [x, f_1(x), f_2(x, f_1(x)), \ldots]\ (dense\ block), \tag{9}$$

Here, f, f_1, f_2,..., represent convolutional layers.

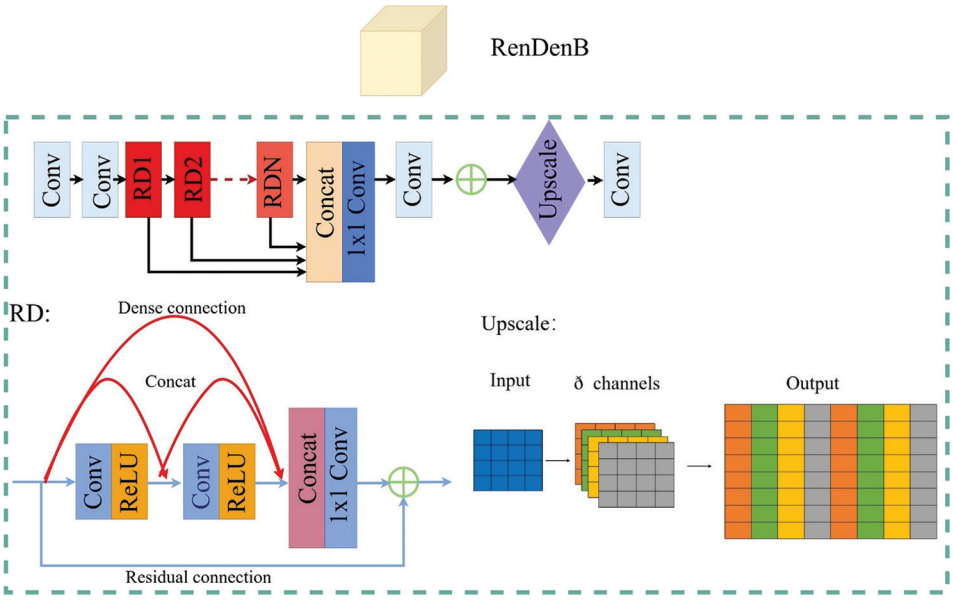

Figure 3. The structure of residual dense blocks (ResDenB)
Abbreviations: Conv: Convolution; RD: Residual dense block.

The residual dense block significantly enhances signal propagation within the network. It achieves this by combining residual and dense connections, promoting the cascaded utilization of multi-level features. This design not only effectively extracts local features but also enhances the network's representational ability through the feature-reuse mechanism. In addition, the skip connection in residual dense blocks further increases the network depth, improves training stability, and avoids problems such as gradient vanishing and gradient explosion. The Upscale module is a key component for improving spatial resolution. Its goal is to transform low-resolution seismic features into high-resolution features, thereby enhancing seismic data detail. To achieve this, we adopted a combination of the Pixel Shuffle layer and a convolutional layer. Given that temporal sampling in seismic data is usually sufficient, interpolation is performed only in the spatial dimensions. The pixel rearrangement layer is an efficient up-sampling method that improves the resolution of feature maps without introducing additional parameters. It works by redistributing channel information into the spatial dimensions, effectively increasing the spatial resolution.[42]

3.4. Convolutional block attention module

Not all features contributed by the encoder are equally important for reconstructing a clean, high-resolution image in the decoder. Seismic noise is often unevenly distributed, and key geological structures occupy specific spatial locations. The CBAM module addresses this by performing intelligent feature selection, suppressing noisy channels, and emphasizing spatially significant regions like fault lines, thus preventing the transfer of noise through skip connections. By incorporating an attention mechanism into CNNs, CBAM effectively enhances the network's ability to capture important features, improving performance in denoising and super-resolution tasks. As shown in

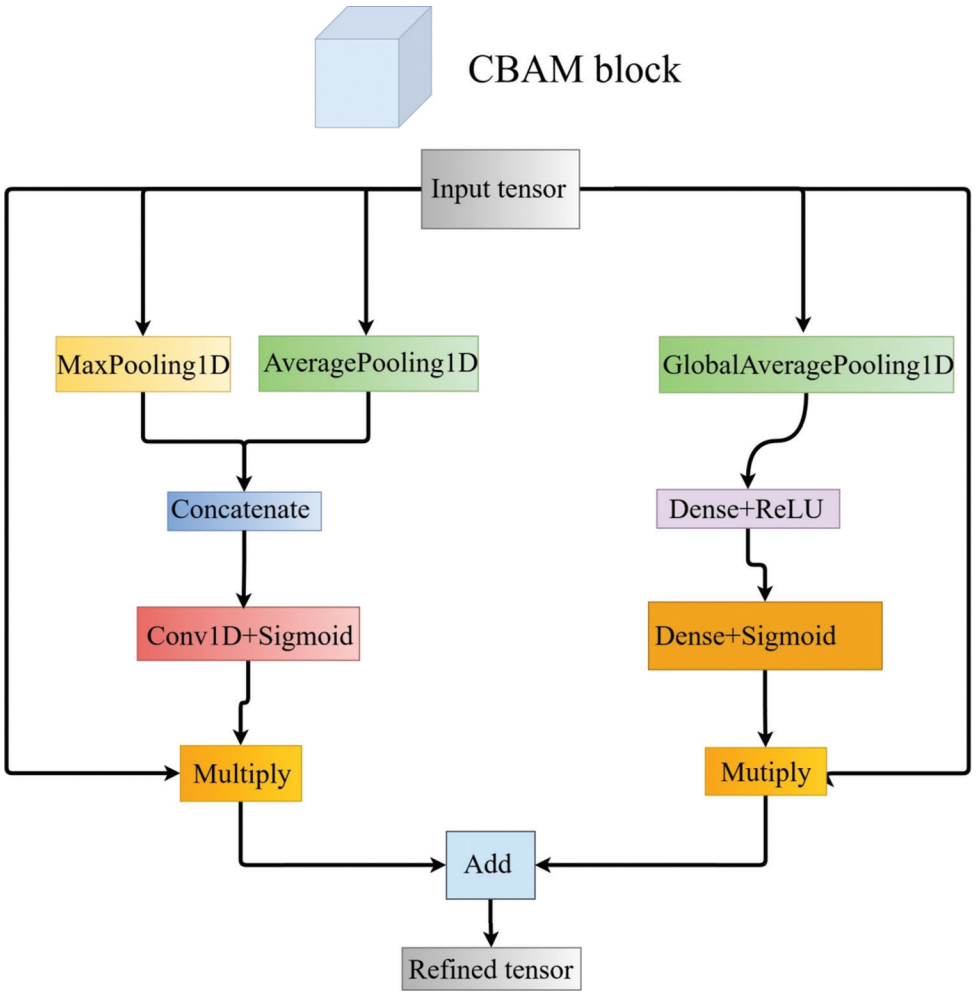

Figure 4. The structure of the convolutional block attention module
Abbreviations: Conv: Convolution.

Figure 4, the CBAM module mainly consists of two parts:

(i) Channel attention: The channel attention module computes global statistics of each channel through global average pooling and global maximum pooling operations. These global features are input into a fully connected layer to generate the channel attention diagram. Channel attention is used to weight the characteristics of each channel, highlighting the important channel characteristics and suppressing unimportant ones.

(ii) Spatial attention: The spatial attention module generates a spatial attention diagram by applying global average pooling and global maximum pooling across the channel dimension. Spatial attention weighs the features at each pixel location, highlighting important spatial regions and suppressing unimportant ones.

In the joint task, the CBAM module and Swin Transformer performed complementary functions. The Swin Transformer modeled global features such as stratum continuity through a hierarchical self-attention mechanism, while CBAM enhanced local detail representation through channel-spatial attention. For example, channel attention suppresses noise-dominated feature channels, while spatial attention focuses on high-frequency areas such as fault edges. Together, they enhance the model's recovery accuracy for complex geological structures.

4. Loss function

The loss function adopted in this study builds on the work of Li *et al.*[43] and combines L1 loss with MS-SSIM loss. This combined loss aims to improve the perceptual quality of restored images and reduce over-smoothing of geological edges. The specific loss function is described in Sections 4.1–4.3.

4.1. L_1 loss

L_1 loss is a commonly used pixel-level loss function that measures the pixel-wise difference between the predicted image and the target image. It has been widely used in image super-resolution tasks. It is defined as follows:

$$L_{L1} = \frac{1}{N}\sum_{i,j}|I_{S\cdot R(i,j)} - I_{H\cdot R(i,j)}|, \tag{10}$$

Where N is the total number of pixels in the image, $I_{SR}(i, j)$ is the pixel value of the super-resolution image at (i, j), and $I_{HR}(i, j)$ is the pixel value of the ground-truth high-resolution image at (i, j).

4.2. Multiscale structural similarity loss

The MS-SSIM loss is a perceptual loss function used to measure the structural similarity between images. It considers image luminance, contrast, and structure, aligning more closely with human visual perception. SSIM is defined as follows:

$$SSIM(x, y) = [l(x, y)]^{\alpha} * [c(x, y)]^{\beta} * [s(x, y)]^{\gamma}, \tag{11}$$

Where $l(x, y)$ represents brightness comparison, $c(x, y)$ represents contrast comparison, $s(x, y)$ represents structure comparison, and α,β, and γ are weighting parameters. The specific calculation formulas are:

$$l(x,y) = \frac{2\mu_x\mu_y + c_1}{\mu_x^{2+}\mu_y^2 + c_1}, \tag{12}$$

$$c(x,y) = \frac{2\sigma_{xy} + c_2}{\sigma_x^{2+}\sigma_y^2 + c_2}, \tag{13}$$

$$s(x,y) = \frac{\sigma_{xy} + c_3}{\sigma_x\sigma_y + c_3}, \tag{14}$$

Where μ_x and μ_y are the mean intensities of images x and y, respectively; σ_x and σ_y are the standard deviations of images x and y, respectively; σ_{xy} is the covariance of images x and y; c_1, c_2, and c_3 are small constants introduced to avoid division by zero. MS-SSIM calculates SSIM at multiple scales and combines the results. Typically, five scales are used for MS-SSIM computation. Its definition is as follows:

$$MS-SSIM(x,y) = [l_M(x,y)]^{aM} * \prod_{j=1}^{M}[c_j(x,y)]^{\beta_j} * [s_j(x,y)]^{\gamma_j}, \tag{15}$$

Where M is the number of scales, usually set to 5.

4.3. Mixed loss function

To leverage the advantages of both L_1 and MS-SSIM losses, we use a mixed loss function defined as follows:

$$L_{Mix} = \alpha * L_{MS-SSIM} + (1-\alpha) * L_{l1}, \tag{16}$$

Where $L_{MS-SSIM} = 1-MS-SSIM(I_{SR}, I_{HR})$, and α is the weighting factor of the loss function, empirically set to 0.6 in this study. The mixed loss function significantly improves the quality of the restored images. This improvement is particularly evident in the restoration of detailed features such as faults and stratigraphic interfaces. Compared with using only the L_1 loss, the mixed loss better preserves

geological details and reduces over-smoothing. Crucially, the inclusion of the MS-SSIM component serves as a strong structural constraint, which is vital for addressing the ill-posed nature of the super-resolution problem. By penalizing structural discrepancies across multiple scales, it effectively mitigates the risk of the model generating geologically implausible artifacts (i.e., "hallucinations"), thereby enhancing the fidelity of the output. Consequently, it improves the accuracy of subsequent seismic interpretation.

5. Evaluation metrics

To accurately evaluate the performance of the neural network in noise removal, we adopt PSNR as a quantitative evaluation metric. PSNR is a widely used metric in the field of image processing that measures the ratio between the maximum possible signal power and the power of corrupting noise. A higher PSNR value indicates a lower noise level and better image quality. For neural network denoising, PSNR is calculated using the mean squared error (MSE) between the original clean (noise-free, high-resolution) image and the network's output image. Specifically, PSNR is computed by the following formula:

$$PSNR = 20 * log_{10}\left(\frac{MAX_1}{\sqrt{MSE}}\right), \tag{17}$$

Where MAX_I represents the maximum possible pixel value, and MSE is the MSE between the original image and the processed image, defined as:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(I_{original}(i) - I_{processed}(i))^2, \tag{18}$$

Where N is the total number of pixels, and $I_{original}(i)$ and $I_{processed}(i)$ represent the pixel values of the original image and the processed image at the i^{th} pixel position, respectively.

By calculating PSNR, the ability of the neural network to remove noise while retaining image details and structural information can be quantitatively evaluated. A higher PSNR value generally indicates better denoising performance, meaning the network effectively reduces random noise while maintaining the clarity and accuracy of geological features. This metric is used to compare the denoising performance of different network architectures, loss functions, and training strategies in this study, providing an important reference for model optimization.

6. Training data

To ensure the quality and diversity of the training data, the dataset used in this study was derived from Wu *et al.*[44,45] As shown in Figure 5, the specific steps are as follows:

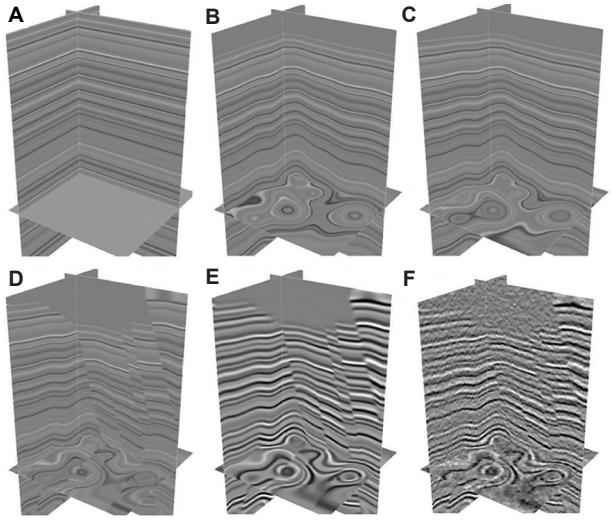

Figure 5. The workflow for creating a 3D synthetic training dataset. (A) Initial flat reflectivity model; (B) Model after adding fold structures using vertical shear and 2D Gaussian functions; (C) Model after further addition of fault structures via plane shear; (D) Final reflectivity model containing both folds and faults; (E) Synthetic high-resolution seismic data volume generated by convolving (D) with a Ricker wavelet; (F) Low-resolution noisy input data created by down-sampling and adding random noise to (E).

(i) Construction of initial reflectivity model: First, an initial reflectivity model (Figure 5A) was generated, which included all flat structures. Subsequently, fold and fault structures were added to the model to simulate the geological characteristics in actual seismic data.

(ii) Adding fold structure: Using a vertical shear model combined with several 2D Gaussian functions to define shear displacement, fold structures were added to the initial model (Figure 5B). This step simulated common geological folding phenomena in real seismic data and increased data complexity.

(iii) Adding fault structures: Plane shear (Figure 5C) was further added to the model to simulate geological faults. The inclusion of these fault structures made the model more representative of real geological conditions and enhanced data diversity.

(iv) Generation of reflectivity model: Through the above steps, a reflectivity model containing folds and faults was obtained (Figure 5D). This model was then used to generate a synthetic seismic data volume.

(v) Generating synthetic seismic data volume: The reflectivity model was convolved with a Ricker wavelet to generate a synthetic seismic data volume (Figure 5E). To simulate the noise interference present in real seismic data, random noise was added to the low-resolution data volume (Figure 5F).

(vi) Extracting 2D slices: 2D slices were extracted from the high-resolution seismic data volume as label images

in the training dataset. Corresponding 2D slices were also extracted from the low-resolution seismic data volume and down-sampled to serve as input images in the training dataset.

(vii) Data enhancement: To increase the diversity of the training data, the extracted 2D slices were augmented using operations such as random cropping and horizontal flipping.

(viii) Data normalization: The input and label seismic images were normalized to the range [0, 1] to ensure the stability and convergence of the training process.

Through the above steps, a high-quality 3D synthetic training dataset was generated, providing a solid foundation for model training and optimization.

7. Dataset partition

Following the method of Li *et al.*,[43] the generated synthetic seismic data were divided into training, validation, and test sets. Specifically, 600 3D seismic data volumes were used for the training set, 75 for the validation set, and 75 for the test set. From each seismic volume, a low-resolution noisy seismic image of size 128 × 128 was extracted as the network input, while a high-resolution noise-free seismic image of size 256 × 256 was used as the network output, as shown in Figure 6.

8. Experiments

The model was implemented in PyTorch (Meta AI Research, USA) and trained on an NVIDIA GeForce RTX 3090 graphics processing unit. The Adam optimizer[46] was used with parameters ($\beta_1 = 0.9$, $\beta_2 = 0.999$). The initial learning rate was set to 1×10^{-4} and halved every 30 epochs. The model was trained for 100 epochs with a batch size of 16. The mixed loss function in **Equation 16** with $\alpha = 0.6$ was used to update the network parameters. The input patches were normalized to [0, 1]. Data augmentation techniques, including random cropping and horizontal flipping, were applied during training. To verify the effectiveness and generalization of the proposed U-STDRNet, both synthetic data and two field datasets were used for testing. The traditional U-Net served as the baseline for comparison. Figure 7 illustrates the overall workflow, comprising three stages: data preparation, model training, and evaluation. First, a 3D synthetic reflectivity model was generated to create paired low-resolution noisy inputs and high-resolution clean labels. These data pairs were then fed into the U-STDRNet model for training. Finally, the trained U-STDRNet was applied to both synthetic and field data for comprehensive evaluation.

8.1. Synthetic data examples

As mentioned above, 1,200 pairs of 2D images were generated for training. Each input seismic image and its corresponding target image were normalized to the range [0, 1] using the following formula:

$$x^* = \frac{x - x_{min}}{x_{max} - x_{min}}, \tag{19}$$

Where x^* represents the normalized seismic image, and x_{max} and x_{min} denote the maximum and minimum values of each input seismic image, respectively. The images were then augmented using the previously described methods. During model training, the Adam optimizer[46] was used,

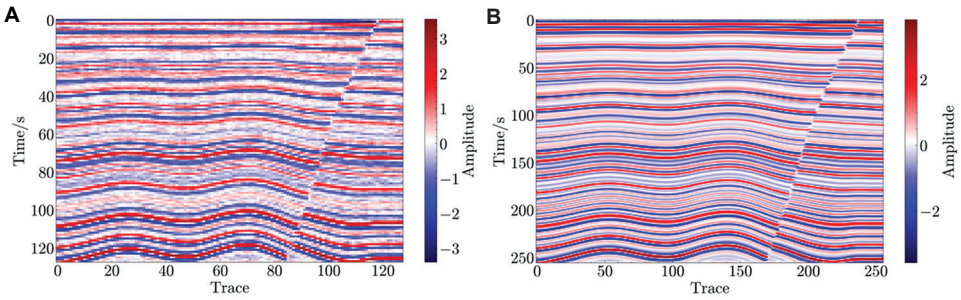

Figure 6. Example of training data. (A) Low-resolution noisy data; (B) High-resolution noise-free label data.

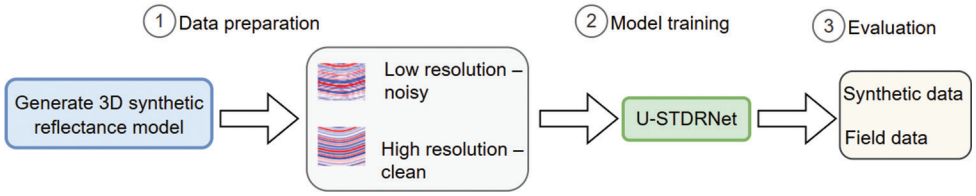

Figure 7. The overall workflow of the proposed method encompassing data preparation, model training, and evaluation phases

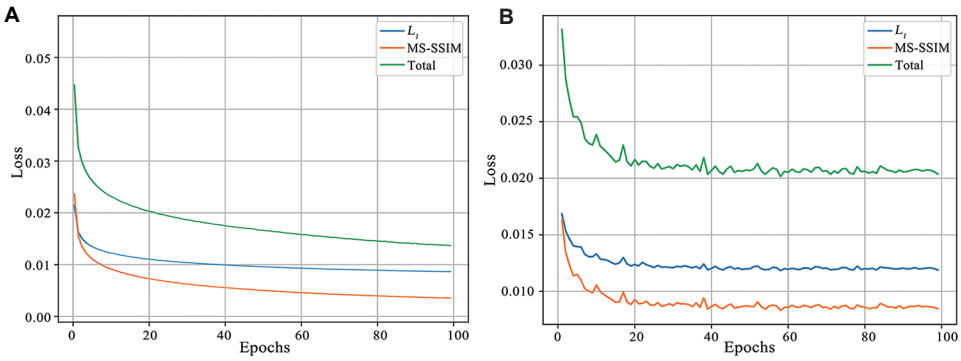

Figure 8. Training and validation loss curve for the U-Net: (A) Training; (B) Validation
Abbreviation: MS-SSIM: Multiscale structural similarity.

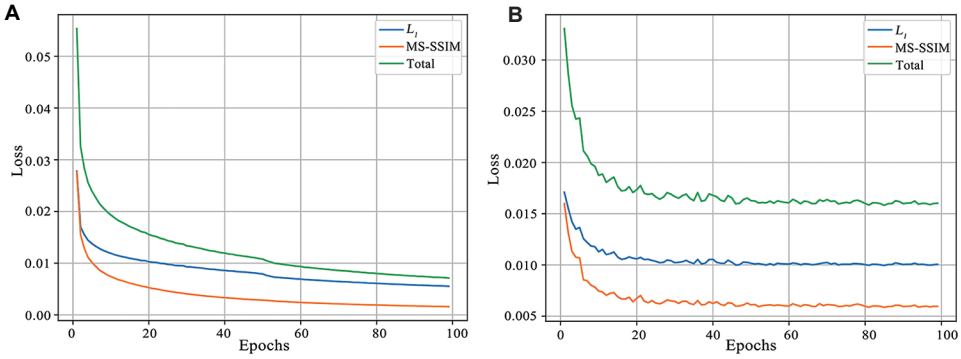

Figure 9. Training and validation loss curve for the U-Net Shifted Window (Swin) Transformer-based dense residual network: (A) Training; (B) Validation
Abbreviation: MS-SSIM: Multiscale structural similarity index.

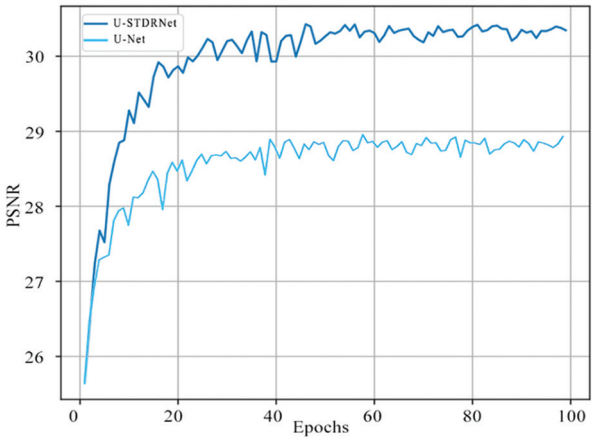

Figure 10. Peak signal-to-noise ratio comparison between U-Net and U-Net Shifted Window Transformer-based dense residual network (U-STDRNet) models

with the initial learning rate set to 1×10^{-4}. The batch size was 16, and a total of 16,000 image patch pairs were extracted from the training dataset. The network was trained for 100 epochs in total.

To evaluate the contribution of each proposed module (STCRB, CBAM, ResDenB) in the network, U-STDRNet and the original U-Net were trained and tested using identical datasets and hyperparameters. Figure 8 shows the training and validation loss curves for the U-Net, while Figure 9 demonstrates the corresponding curves for the U-STDRNet. Analysis of both figures revealed that U-STDRNet achieves lower training and validation loss than U-Net throughout the training process. Incorporating the proposed modules into the baseline U-Net further reduced both the L_1 loss and MS-SSIM loss components, indicating continuous optimization in pixel-level accuracy. In addition, image structure and perceptual quality progressively improved. The network also exhibited faster convergence and higher accuracy. Figure 10 presents the PSNR on the validation sets for both networks, showing that U-STDRNet consistently achieves higher PSNR than U-Net. This demonstrates U-STDRNet's superior ability to suppress random noise.

To quantitatively evaluate the contribution of each proposed module (STCRB, CBAM, ResDenB), seven experimental configurations were designed: Unet, Unet + CBAM, Unet + ResDenB, Unet + STCRB, Unet + STCRB + ResDenB, Unet + STCRB + CBAM, and

Table 1. Quantitative comparison of different models

Model	PSNR	SSIM	L_{mix}
U-Net	28.903	0.907	0.014
U-Net+CBAM	29.301	0.909	0.012
U-Net+ResDenB	29.788	0.910	0.012
U-Net+STCRB	30.120	0.932	0.013
U-Net+STCRB+ResDenB	30.198	0.941	0.010
U-Net+STCRB+CBAM	30.056	0.950	0.009
U-STDRNet	30.523	0.956	0.008

Abbreviations: CBAM: Convolutional block attention module; ResDenB: Residual dense blocks; PSNR: Peak signal-to-noise ratio; SSIM: Multiscale structural similarity index; STCRB: Shifted Window Transformer convolutional residual blocks; U-STDRNet: U-Net Shifted Window Transformer-based dense residual network.

the full U-STDRNet model. Their performance on the synthetic test set, measured by PSNR, SSIM, and L_{mix}, is comprehensively compared in Table 1.

Table 1 provides a detailed ablation study, revealing the impact of different module combinations on model performance across multiple metrics:

(i) Effectiveness of individual modules: The introduction of any single module (CBAM, ResDenB, or STCRB) to the baseline U-Net brought consistent improvements across all three metrics. The STCRB module yielded the most significant boost in PSNR (+1.217 dB) and SSIM (+0.025), underscoring the paramount importance of global feature modeling using the Swin Transformer for capturing stratigraphic continuity. Concurrently, it also reduced the L_{mix} metric by 0.001, indicating better preservation of geological edges. The ResDenB module, focusing on local feature enhancement through dense connections, achieved the second-highest PSNR gain and also improved SSIM. The CBAM module, acting as an adaptive feature selector, provided a moderate PSNR increase and the most notable reduction in the L_{mix} value (from 0.014 to 0.012) among the single-module additions, demonstrating its particular efficacy in refining structural edges and suppressing noise in non-critical areas.

(ii) Synergistic effects of module combinations: The combination of modules revealed insightful interactions. The integration of STCRB and ResDenB (U-Net+STCRB+ResDenB) achieved a high PSNR of 30.198 dB and a significantly improved SSIM of 0.941, demonstrating that global context and local feature reuse are highly complementary. However, the combination of STCRB and CBAM (U-Net+STCRB+CBAM) resulted in the highest SSIM (0.950) and a further reduced L_{mix} value (0.009), highlighting that global modeling coupled with attentive feature selection is exceptionally powerful for enhancing structural perceptual quality and edge accuracy. This suggests that, after establishing a strong global context, intelligently weighting features is more beneficial for structural fidelity than simply adding more network capacity.

(iii) Superiority of the full model (U-STDRNet): The proposed full U-STDRNet model, which integrated all three modules (STCRB, ResDenB, and CBAM), achieved the best performance across all metrics—the highest PSNR (30.523 dB), the highest SSIM (0.956), and the lowest L_{mix} value (0.008). This demonstrates a clear synergistic effect, where each module addresses a distinct aspect of the problem—STCRB ensured global coherence, ResDenB enhanced fine detail recovery, and CBAM dynamically prioritized salient features and suppressed noise. The lowest L_{mix} value confirms that U-STDRNet is most effective at preserving the sharpness and integrity of crucial geological structures like faults and stratigraphic interfaces, which is essential for accurate seismic interpretation.

Figure 11 presents a visual comparison of the results on the synthetic test set. At first glance, the outputs from U-Net (Figure 11A) and U-STDRNet (Figure 11B) appeared similar, as both effectively enhanced resolution. However, a standard visual comparison is insufficient to judge geological fidelity, which is critical for interpretation. To provide a precise, localized assessment of structural fidelity, we generated local SSIM maps (Figure 11C and 11D). These maps highlight areas of structural error. In these maps, brighter colors (yellow) represent high structural similarity (high fidelity) to the ground truth, while darker colors (green) represent low similarity (i.e., structural errors or artifacts). The SSIM map for U-Net (Figure 11C) reveals significant dark green patches concentrated along the complex fault zones and dipping events (highlighted by the red box), indicating that U-Net failed to accurately preserve the fidelity of these critical geological structures. In stark contrast, the SSIM map for U-STDRNet (Figure 11D) is almost uniformly bright yellow across the entire section, including the complex fault zone. This demonstrates superior and consistent structural fidelity, showing that our model correctly reconstructed the geological features where the U-Net failed.

This clear visual evidence directly supports the quantitative metrics in Table 1, where U-STDRNet achieved a much higher overall SSIM (0.956) compared to U-Net (0.907). Combined with the lowest L_{mix} value (0.008) from the ablation study, this confirms our model's exceptional performance in preserving geological edges and minimizing structural distortion. These results provide compelling

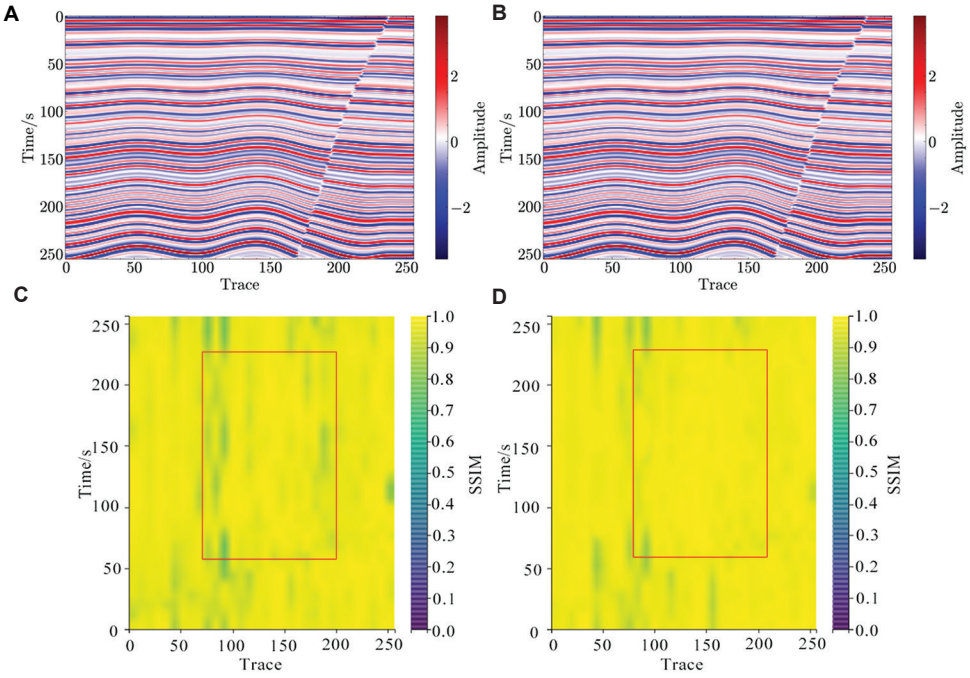

Figure 11. Results and local similarity on synthetic data for U-Net and U-STDRNet. (A) U-Net result; (B) U-STDRNet result; (C) Local SSIM versus ground truth (U-Net); (D) Local SSIM versus ground truth (U-STDRNet)
Abbreviations: SSIM: Structural similarity index; U-STDRNet: U-Net shifted window transformer-based dense residual network.

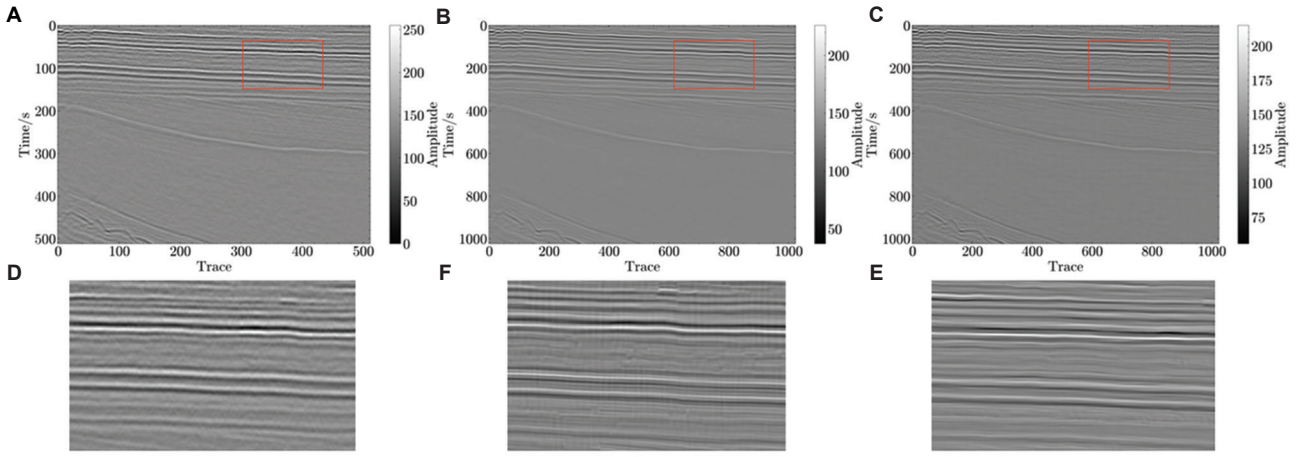

Figure 12. Comparison of the first field dataset. (A) Raw data; (B) Result of U-Net processing; (C) Result of U-STDRNet processing; (D-F) Enlarged images corresponding to the red borders.

evidence that the integration of global modeling (STCRB), local feature enhancement (ResDenB), and attentive feature selection (CBAM) in U-STDRNet collectively enables a more geologically faithful reconstruction.

8.2. Field data examples

In this section, two field datasets are used for testing. The complexity and diversity of real data place higher demands on model performance. The following is a detailed description of the processing results for these two datasets. The first field dataset is characteristic of a large, stable sedimentary region, such as a cratonic basin. As shown in the raw data (Figure 12A and D), the geological structure is relatively simple, dominated by flat-lying to sub-horizontal continuous reflectors. The primary challenge in this dataset is not structural complexity, but the presence of significant random noise that obscures the details of thin-layer interfaces. This data effectively tested the model's ability

to suppress noise while enhancing subtle stratigraphic features and improving event continuity.

Figure 12 shows the results for the first field dataset. Figure 12B and C show the seismic images processed using the original U-Net and U-STDRNet, respectively.

Figure 12E shows the seismic image processed using the original U-Net. Compared to the raw data (Figure 12D), the U-Net improves image resolution to some extent. However, it still exhibits shortcomings in restoring fine details and suppressing noise. Some geological features are over-processed by the U-Net, and new artifacts are introduced.

Figure 12F shows the seismic image processed by U-STDRNet. Compared to U-Net, U-STDRNet demonstrates superior detail feature recovery and noise suppression. Stratigraphic interfaces appear clearer. Waveform fluctuations and variations are smoother, and noise is effectively suppressed while original geological characteristics are preserved.

To quantitatively evaluate the resolution improvement, we restricted the analysis to 0–100 Hz and compared spectra within this band (Figure 13). Analysis of the 0–100 Hz spectrum (Figure 13) shows that U-STDRNet (red line) achieves better energy recovery than U-Net (green line) across low, intermediate, and high frequencies. This improvement is particularly pronounced in the high-frequency band, indicating that U-STDRNet better preserves the overall stratigraphic structure while more effectively enhancing details and achieving higher resolution. Therefore, for seismic data super-resolution, U-STDRNet exhibited superior resolution enhancement and geological information recovery capabilities compared to the traditional U-Net.

The second field dataset presents a more complex geological challenge, originating from a tectonically active area, such as a fold-and-thrust belt (e.g., on the margin of a basin). The raw section (Figure 14A and D) is dominated by complex structures, including steeply dipping strata and significant faulting. The reflectors show variable dips and discontinuities. Processing this data effectively requires the model to not only suppress noise but also maintain structural integrity across faults and accurately image the dipping events, testing the model's robustness and generalization to complex geological settings.

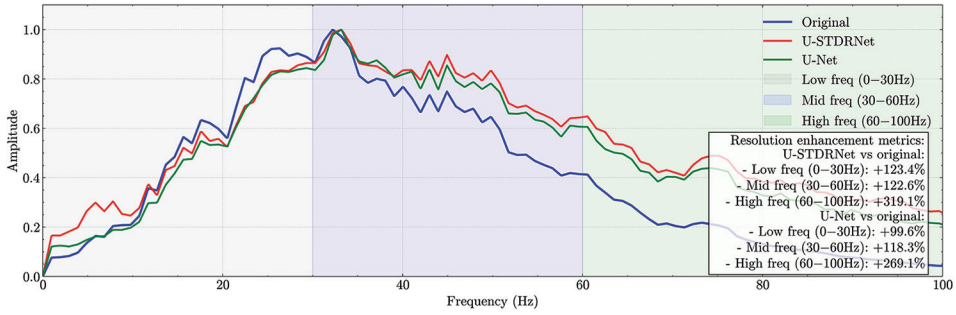

Figure 13. Spectrum comparison I (0–100 Hz)
Abbreviations: freq: Frequency; U-STDRNet: U-Net Shifted Window Transformer-based dense residual network.

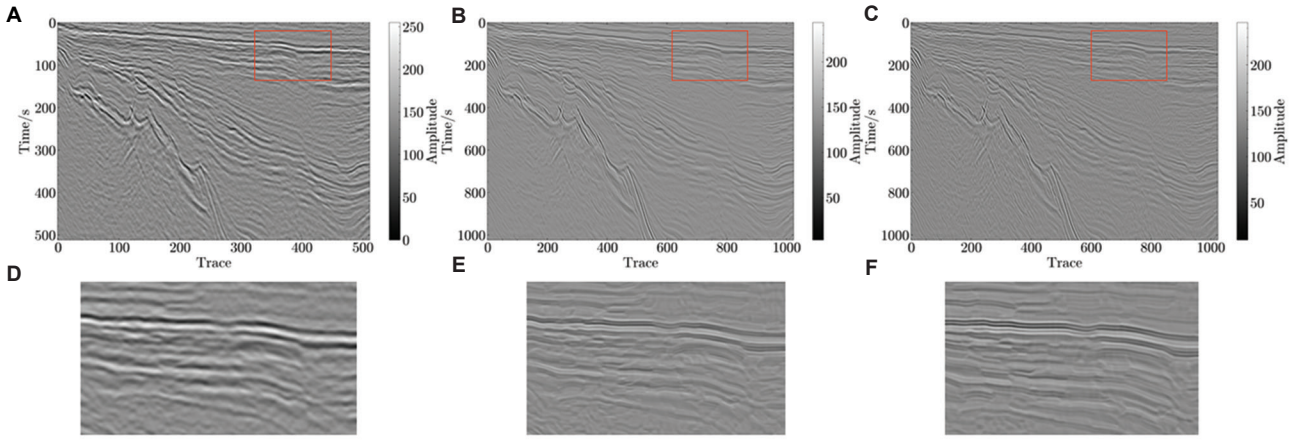

Figure 14. Comparison of the second field dataset. (A) Raw data; (B) Result of U-Net processing; (C) Result of U-STDRNet processing; (D-F) Enlarged images corresponding to the red borders.

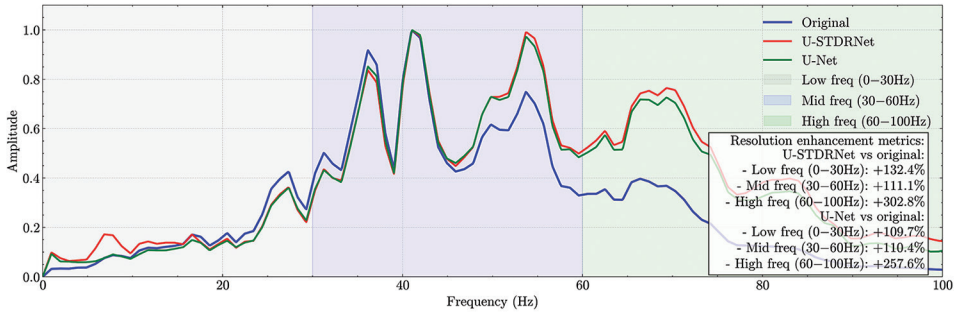

Figure 15. Spectrum comparison II (0–100 Hz)
Abbreviations: freq: Frequency; U-STDRNet: U-Net Shifted Window Transformer-based dense residual network.

As shown in Figure 14, we processed the second field dataset. Figure 14B and C, respectively, show the seismic images processed using the original U-Net and the U-STDRNet.

Figure 14E illustrates the result processed by the U-Net. When handling complex geological structures, the U-Net exhibits limited ability to recover detailed features and suppress noise. Some distinct features are smoothed out, stratigraphic interfaces remain unclear, and noise is still evident.

Figure 14F depicts the output of U-STDRNet. U-STDRNet restores key geological features, such as stratigraphic interfaces, more clearly. These features are significantly more distinct in the processed image.

The corresponding spectral comparison is shown in Figure 15. It is crucial to note that the original data spectrum (blue line) exhibits severe spectral oscillations, characterized by sharp, noisy peaks and deep spectral troughs (nulls). This indicates that the original signal is of low quality, with certain frequency bands already severely attenuated, likely due to geophysical (Q-filtering) effects. In contrast to this compromised original spectrum, within the 0–100 Hz range (Figure 15), U-STDRNet outperformed U-Net in recovering and enhancing signal energy in the low-frequency band. This indicates superior preservation and enhancement of the overall stratigraphic structure. Energy recovery in the mid-frequency band is more complete, demonstrating stronger resolution of stratigraphic details and better revelation of mid-scale geological characteristics. Energy recovery in the high-frequency band is most pronounced, indicating the greatest resolution improvement. This highlights U-STDRNet's enhanced capability for revealing small-scale geological features such as thin layers and faults.

To further strengthen the evaluation of field data processing, we introduced non-reference quantitative metrics, as ground-truth labels were unavailable. We employed image entropy to measure noise suppression (lower is better) and the Tenengrad gradient-based metric to quantify image sharpness and detail preservation (higher is better). Table 2 compares these metrics for the original data, the U-Net result, and our U-STDRNet result across both field datasets. The results showed a clear and consistent trend. For Field Data 1, the original data had high entropy (7.15). U-Net provided a moderate reduction (6.82), but our U-STDRNet achieved the lowest entropy (6.51), confirming its superior noise suppression capabilities. Concurrently, U-STDRNet also achieved the highest Tenengrad value (48.9), significantly improving upon U-Net (41.3) and the original (32.4), which indicates a much clearer restoration of stratigraphic interfaces. This advantage was even more pronounced in the complex Field Data 2. U-STDRNet again achieved the lowest entropy (6.79), demonstrating robust noise control even in structurally complex areas. More importantly, it scored the highest on the Tenengrad metric (79.5), demonstrating its exceptional ability to sharpen complex features such as dipping strata and faults, far exceeding U-Net (68.2). These results provide strong quantitative evidence, complementing the visual and spectral analyses, that U-STDRNet is more effective in both noise suppression and detail recovery for practical field data applications.

Table 2. Quantitative comparison of field datasets using no-reference metrics

Dataset	Method	Image entropy	Tenengrad
Field data 1	Original raw	7.15	32.4
	U-Net	6.82	41.3
	U-STDRNet	6.51	48.9
Field data 2	Original raw	7.32	55.7
	U-Net	7.01	68.2
	U-STDRNet	6.79	79.5

Abbreviation: U-STDRNet: U-Net Shifted Window Transformer-based dense residual network.

8.3. Result analysis

Comparing results on synthetic and field data shows that U-STDRNet provides clear advantages in seismic image super-resolution and denoising tasks. In detailed feature recovery, the U-STDRNet can better recover fine geological features in seismic images, which are crucial in seismic interpretation and can help geologists identify and analyze geological structures more accurately. In noise processing, the U-STDRNet performed better in noise suppression, effectively reducing random noise in the image and improving its clarity and quality, whereas noise remained more prevalent in the U-Net processed images. U-STDRNet delivers consistent performance across field datasets, demonstrating strong generalization to varying geological conditions and data characteristics. Comparison of PSNR values across models confirms U-STDRNet's superior overall performance, achieving the highest PSNR and indicating the best image quality.

9. Study limitations

Despite the proposed model's promising performance, this study has limitations that pave the way for future research. A primary concern in any DL-based image enhancement task is fidelity—ensuring the model does not generate false geological events. While our use of a mixed loss function with MS-SSIM and a U-Net architecture with skip connections was explicitly designed to mitigate this risk by enforcing structural consistency with the input data, a more rigorous validation is desirable. A gold-standard validation, as suggested by the community, would involve testing the model on synthetic datasets generated from well logs and rock-physics models. This would provide an undisputed ground truth to definitively verify that the enhanced details are geophysically meaningful. This represents a crucial next step in our future work. In addition, the current training data primarily addressed random noise, leaving the model's efficacy on coherent noise (e.g., multiples) less explored. Future work will also focus on incorporating physics-informed constraints and more diverse data augmentation strategies to improve robustness. Finally, extending the current 2D framework to a full 3D architecture is a key direction for better modeling of spatial continuity and for increasing the practical impact of U-STDRNet in modern seismic interpretation workflows.

10. Conclusion

In this paper, we present U-STDRNet, a novel DL framework designed specifically for the joint super-resolution and denoising of seismic images. The core innovation of U-STDRNet lies in its seismically aware architecture, which synergistically combines the global contextual modeling capabilities of the Swin Transformer (STCRB), the local feature enhancement and weak signal recovery prowess of ResDenB, and the adaptive feature refinement provided by the CBAM. Unlike a straightforward aggregation of existing components, the integration was carefully designed to address the distinct challenges of seismic data: preserving stratigraphic continuity, enhancing subtle geological features, and suppressing noise without introducing artifacts.

Experimental results on both synthetic and field datasets demonstrated that U-STDRNet significantly outperforms the baseline U-Net model. It showed superior performance in recovering complex structures and suppressing noise compared to state-of-the-art traditional methods such as signal-and-noise orthogonalization[47] and non-stationary predictive filtering,[48] particularly in terms of automation and feature preservation. It achieves higher PSNR and SSIM values and, more importantly, produces superior visual outcomes with clearer stratigraphic interfaces, better-restored fault and fold continuities, and more effective noise suppression. The ablation studies confirm the indispensable role of each proposed module in achieving these gains.

Acknowledgments

None.

Funding

This work was supported in part by the National Key Research and Development Program of China under Grant 2024YFB4007100, in part by the National Major Science and Technology Projects of China under Grant 2024ZD1004300, in part by National Natural Science Foundation of China under Grant 42304133 and 42574175, in part by Key Project of the Education Department of Hubei Province (Grant No. D20241304), and in part by Key project from the Hubei Research Center for Basic Disciplines of Earth Sciences under Grant HRCES-202401.

Conflict of interest

The authors declare no conflicts of interest.

Author contributions

Conceptualization: Min Bai, Guangtan Huang
Formal analysis: Juan Wu
Investigation: Mingliao Wu, Juan Wu, Haiyu Li
Methodology: Juan Wu, Zhixian Gui
Writing–original draft: Mingliao Wu
Writing–review & editing: Mingliao Wu, Guangtan Huang

Availability of data

The data generated or analyzed during this study are available from the corresponding author upon reasonable request.

References

1. Bianco MJ, Gerstoft P, Olsen KB, Lin FC. High-resolution seismic tomography of long beach, CA using machine learning. *Sci Rep.* 2019;9:14987.

   doi: 10.1038/s41598-019-50381-z

2. Zhang Y, Wang Y, Yin J. Single point high density seismic data processing analysis and initial evaluation. *Oil Geophys Prospect*. 2010;45(2):201-207.

   doi: 10.13810/j.cnki.issn.1000-7210.2010.02.008

3. Rebert T, Sablon R, Vidal N, Charrier P, Soubaras R. Improving pre-salt imaging with variable-depth streamer data. In: *SEG Technical Program Expanded Abstracts 2012*. United States: SEG; 2012. p. 1-5.

   doi: 10.1190/segam2012-1067.1

4. Wu X, Gan L, Yuan S, Rui D. A preliminary study on wellbore flow interpretation of fiber optic vibration signals based on K-means clustering algorithm. *SN Appl Sci*. 2022;4(8):233.

   doi: 10.1007/s42452-022-05117-6

5. Wang Y, Wang J, Wang X, Sun W, Zhang J. Broadband Processing Key Technology Research and Application on Slant Streamer. In: *International Geophysical Conference, Beijing, China, 24-27 April 2018*. United States: SEG; 2018. p. 135-138.

   doi: 10.1190/IGC2018-034

6. Mu XR, Mao Q, Huang JP. Stable attenuation-compensated reverse time migration and its application to land seismic data. *Petrol Sci*. 2023;20(5):2784-2795.

   doi: 10.1016/j.petsci.2023.03.014

7. Wang Z, Yue WZ, Zhu YM, Ji NX, Fan SS. Near-wellbore 3D velocity imaging inversion method based on array acoustic logging data. *Petrol Sci*. 2025;22:002341.

   doi: 10.1016/j.petsci.2024.06.007

8. Wang C, Chen L, Li L, *et al*. Combining unscented Kalman filter and wavelet neural network for anti-slug. *Petrol Sci*. 2023;20(6):3752-3765.

   doi: 10.1016/j.petsci.2023.05.008

9. Li L, Wang ZZ, Yin SD, *et al*. Selection and application of wavelet transform in high-frequency sequence stratigraphy analysis of coarse-grained sediment in rift basin. *Petrol Sci*. 2024;21(5):3016-3028.

   doi: 10.1016/j.petsci.2024.06.020

10. Liao QZ, Xue L, Lei G, Liu X, Sun SY, Patil S. Statistical prediction of waterflooding performance by K-means clustering and empirical modeling. *Petrol Sci*. 2022;19(3):1139-1152.

    doi: 10.1016/j.petsci.2021.12.032

11. Bai M, Huang G, Wang H, Chen Y. Seismic signal enhancement based on the low-rank methods. *Geophys Prospect*. 2020;68(9):2783-2807.

    doi: 10.1111/1365-2478.13028

12. Huang G, Bai M, Zhao Q, Wang H, Chen W, Chen Y. Erratic noise suppression using iterative structure-oriented space-varying median filtering with sparsity constraint. *Geophys Prospect*. 2021;69(1):101-121.

    doi: 10.1111/1365-2478.13032

13. Wu J, Bai M, Zhang D, Wang H, Huang G, Chen Y. Fast and robust low-rank approximation for five-dimensional seismic data reconstruction. *IEEE Access*. 2020;8:175501-17512.

    doi: 10.1109/ACCESS.2020.3026020

14. Liu W, Cheng Q, Liu L, Wang Y, Zhang J. Accelerating high-resolution seismic imaging by using deep learning. *Appl Sci Basel*. 2020;10(7):2502.

    doi: 10.3390/app10072502

15. Yu S, Ma J, Wang W. Deep learning for denoising. *Geophysics*. 2019;84(6):V333-V350.

    doi: 10.1190/geo2018-0668.1

16. Ronneberger O, Fischer P, Brox T. *U-Net: Convolutional Networks for Biomedical Image Segmentation; Proceedings of the Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015*. Cham, F: Springer International Publishing; 2015.

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

17. Wang Y, Lu W, Liu J, Zhang M, Miao Y. Random seismic noise attenuation based on data augmentation and CNN. *Chin J Geophys Chin*. 2019;62(1):421-433.

    doi: 10.6038/cjg2019M0385

18. Dong X, Zhong T, Wang H, Wu N, Li Y, Yang B. The denoising of desert seismic data acquired from tarim basin based on convolutional adversarial denoising network. *Chin J Geophys*. 2022;65(7):2661-2672.

    doi: 10.6038/cjg2022P0279

19. Bai T, Zhao H, Wang Z. *A U-Net Based Deep Learning Approach for Seismic Random Noise Suppression; Proceedings of the IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*; 2022.

    doi: 10.1109/IGARSS46834.2022.9880193

20. Halpert AD. Deep learning-enabled seismic image enhancement. In: *SEG Technical Program Expanded Abstracts 2018*. United States: SEG; 2018. p. 2081-2085.

    doi: 10.1190/segam2018-2996943.1

21. Jiang J, Ren H, Zhang M. A convolutional autoencoder method for simultaneous seismic data reconstruction and denoising. *IEEE Geosci Remote Sens Lett*. 2022;19:1-5.

    doi: 10.1109/LGRS.2021.3073560

22. Ma H, Yao H, Li Y, Wang H. Deep residual encoder-decoder networks for desert seismic noise suppression. *IEEE Geosci Remote Sens Lett*. 2020;17(3):529-533.

    doi: 10.1109/LGRS.2019.2925062

23. Wang F, Chen S. Residual learning of deep convolutional neural network for seismic random noise attenuation. *IEEE Geosci Remote Sens Lett*. 2019;16(8):1314-1318.

    doi: 10.1109/LGRS.2019.2895702

24. Zhao Z, Rao Y. Seismic random noise suppression based on SP-DnCNN neural network. *Chin J Geophys Chin*. 2024;67(10):3841-3850.

    doi: 10.6038/cjg2023R0483

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

    doi: 10.1109/TGRS.2019.2926772

26. Zhang S, Zhang L, Qin X. Seismic signal denoising using variational mode decomposition and a denoising convolutional neural network. *J Seim Explor*. 2025;34(2):025260030.

    doi: 10.36922/JSE025260030

27. Liu N, Wang J, Gao J, Chang S, Lou Y. Similarity-informed self-learning and its application on seismic image denoising. *IEEE Trans Geosci Remote Sens*. 2022;60:1-13.

    doi: 10.1109/TGRS.2022.3210217

28. Peng T, Liu Y, Liu C, Chen C, Chen J. Seislet weighted stacking method with high resolution for low SNR seismic data. *Chin J Geophys Chin*. 2025;68(3):1087-1101.

    doi: 10.6038/cjg2024R0866

29. Chen G, Liu Y, Zhang M, Zhang H. Dropout-based robust self-supervised deep learning for seismic data denoising. *IEEE Geosci Remote Sens Lett*. 2022;19:1-5.

    doi: 10.1109/LGRS.2022.3167999

30. Liu B, Yue J, Zuo Z, *et al*. Unsupervised deep learning for random noise attenuation of seismic data. *IEEE Geosci Remote Sens Lett*. 2022;19:1-5.

    doi: 10.1109/LGRS.2021.3057631

31. Shao D, Zhao Y, Li Y, Li T. 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

32. Shao J, Wang Y, Liang X, Xue Q, Liang E, Shi S. Siamese network based noise elimination of artificial seismic data recorded by distributed fiber-optic acoustic sensing. *Chin J Geophys Chin*. 2022;65(9):3599-3609.

    doi: 10.6038/cjg2022P0919

33. Wang S, Hu W, Yuan P, *et al*. A self-supervised deep learning method for seismic data deblending using a blind-trace network. *IEEE Trans Neural Netw Learn Syst*. 2023;34(7):3405-3414.

    doi: 10.1109/TNNLS.2022.3188915

34. Xu Z, Luo Y, Wu B, Meng D, Chen Y. Deep nonlocal regularizer: A self-supervised learning method for 3-D seismic denoising. *IEEE Trans Geosci Remote Sens*. 2023;61:1-17.

    doi: 10.1109/TGRS.2023.3329303

35. Geetha K, Hota MK, Karras DA. A novel approach for seismic signal denoising using optimized discrete wavelet transform via honey badger optimization algorithm. *J Appl Geophys*. 2023;219:105236.

    doi: 10.1016/j.jappgeo.2023.105236

36. Yang B, Bai M, Wu J, *et al*. Seismic data denoising using convolutional sparse coding with an efficient alternating direction multipliers minimization algorithm. *J Appl Geophys*. 2025;233:105610.

    doi: 10.1016/j.jappgeo.2024.105610

37. Zhang P, Wang Q, Liu Y, Chen C. A combined denoising method for Q-factor compensation of poststack seismic data. *J Appl Geophys*. 2024;229:105500.

    doi: 10.1016/j.jappgeo.2024.105500

38. Woo S, Park J, Lee JY, Kweon IS. *CBAM: Convolutional Block Attention Module*. Berlin: Springer; 2018.

    doi: 10.1007/978-3-030-01234-2_1

39. Ding M, Zhou Y, Chi Y. Seismic signal denoising using Swin-Conv-UNet. *J Appl Geophys*. 2024;223:105355.

    doi: 10.1016/j.jappgeo.2024.105355

40. Gao L, Shen H, Min F. Swin transformer for simultaneous denoising and interpolation of seismic data. *Comput Geosci*. 2024;183:105510.

    doi: 10.1016/j.cageo.2023.105510

41. Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In: *International Conference on Machine Learning Research*; 2015.

    doi: 10.48550/arXiv.1502.03167

42. Shi W, Caballero J, Huszar F, *et al*. *Real-Time Single Image and Video Super-Resolution using an Efficient Sub-Pixel Convolutional Neural Network; Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; 2016.

    doi: 10.1109/CVPR.2016.207

43. Li J, Wu X, Hu Z. Deep learning for simultaneous seismic image super-resolution and denoising. *Ieee Trans Geosci Remote Sens*. 2021;3057857.

doi: 10.1109/TGRS.2021.3057857

44. Wu X, Hale D. 3D seismic image processing for faults. *Geophysics*. 2016;81(2):IM1-IM11.

doi: 10.1190/geo2015-0380.1

45. Wu X, Liang L, Shi Y, Fomel S. FaultSeg3D: Using synthetic data sets to train an end-to-end convolutional neural network for 3D seismic fault segmentation. *Geophysics*. 2019;84(3):IM35-IM45.

doi: 10.1190/geo2018-0646.1

46. Kingma D, Ba J. *Adam: A Method for Stochastic Optimization*. Berlin: Springer; 2014.

doi: 10.48550/arXiv.1412.6980

47. Huang G, Zhang D, Chen W, Chen Y. Accelerated signal-and-noise orthogonalization. *IEEE Trans Geosci Remote Sens*. 2022;60:1-9.

doi: 10.1109/TGRS.2021.3054839

48. Huang G, Bai M, Wang H, Liu X, Chen Y. Frequency-space-dependent smoothing regularized nonstationary predictive filtering. *IEEE Trans Geosci Remote Sens*. 2022;60:1-9.

doi: 10.1109/TGRS.2021.3064945