ACCSCIENCE PUBLISHING

Journal of Seismic Exploration

ARTICLE

GMLAN: Grouped-residual and multi-scale large-kernel attention network for seismic image super-resolution

Anxin Zhang[1†], Zhenbo Guo[2†*], Shiqi Dong[1†], and Zhiqi Wei[2]

[1]Key Laboratory of Modern Power System Simulation and Control and Renewable Energy Technology (Ministry of Education), School of Electrical Engineering, Northeast Electric Power University, Jilin, China

[2]Bureau of Geophysical Prospecting, China National Petroleum Corporation, Zhuozhou, Hebei, China

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

Abstract

The resolution of seismic images significantly impacts the accuracy of subsequent seismic interpretation and reservoir location. However, the resolution of seismic images often degrades due to the influence of multiple factors, making super-resolution of seismic images essential and critical. We propose a grouped-residual and multi-scale large-kernel attention network (GMLAN) framework, trained on synthetic seismic images to achieve excellent seismic image super-resolution on field seismic data. GMLAN is primarily composed of two modules: The feature extraction module (FEM) and the image reconstruction module (IRM). The FEM consists of two components: Shallow feature extraction (SFE) and deep feature extraction (DFE). The SFE component is designed to capture the basic information of seismic images, such as large-scale structures and morphological features of the strata. The DFE component serves as the cornerstone of the feature extraction process, leveraging residual groups and multi-scale large-kernel attention to distill detailed features from seismic images, such as stratigraphic interfaces, dip angles, and relative amplitudes. Finally, the IRM utilizes sub-pixel convolution, a learnable upsampling technique, to reconstruct super-resolution seismic images while preserving the continuity of seismic features. The framework demonstrates satisfactory performance on both synthetic and field data.

Keywords: Seismic images; Super-resolution; Deep learning; Grouped-residual structures; Malti-scale large-kernel self-attention

[†]These authors contributed equally to this work.

***Corresponding author:**
Zhenbo Guo
(guozhenbo01@cnpc.com.cn)

Citation: Zhang A, Guo Z, Dong S, Wei Z. GMLAN: Grouped-residual and multi-scale large-kernel attention network for seismic image super-resolution. *J Seismic Explor*. doi: 10.36922/JSE025350063

Received: August 25, 2025

Revised: October 17, 2025

Accepted: October 23, 2025

Published online: November 17, 2025

1. Introduction

Geological interpretation is highly dependent on the quality of seismic images. However, seismic images, which are obtained after data acquisition, processing, and imaging, are inevitably influenced by the acquisition environment and data processing methods, resulting in blurred events and noise. Low-resolution seismic images are detrimental to subsequent geological interpretation (e.g., fault detection and reservoir location) and

lead to the loss of valuable details (e.g., folds, faults, and thin layers). Consequently, enhancing the resolution of seismic images has become a critical step in the seismic exploration process.

Over the past few decades, researchers have conducted numerous studies to improve the resolution of seismic images. In general, methods for enhancing resolution can be divided into two categories: Data acquisition[1-6] and data processing.[7] Data acquisition mainly includes broadband acquisition[1-3] and high-density acquisition.[4,5] Broadband acquisition increases the longitudinal resolution of seismic data by recording data over a wider frequency range, while high-density acquisition increases the number of sources and receivers and reduces the processing unit area, effectively suppressing noise and improving the resolution of seismic profiles. However, both broadband acquisition and high-density acquisition require high costs during the data acquisition process. While the data processing methods include, but are not limited to, inversion, denoising, interpolation, attenuation compensation, and deconvolution,[7,8] traditional resolution enhancement methods for seismic images typically involve multiple steps, each introducing corresponding errors.[9] The cumulative effect of these errors across multiple stages often prevents the full restoration of seismic image details in the final high-resolution output. In contrast, end-to-end deep learning approaches can perform super-resolution in a single step, thereby effectively removing the error accumulation inherent in traditional methods. Moreover, deep learning models can automatically learn and update their learnable parameters[10,11] and generally require lower computational resources and costs compared to traditional data processing techniques.

In recent years, traditional resolution enhancement methods for seismic images have become insufficient to meet the demands for rapid and high-precision processing of large amounts of seismic data. With the rapid advancement of graphics processing units (GPUs) and deep learning algorithms, deep learning methods have been widely applied in direct end-to-end processing of seismic images, enabling resolution enhancement in a single step. Among existing approaches, the convolutional neural network (CNN) is a classical method used for image super-resolution reconstruction, extracting local features of seismic images through multiple convolution operations.[12,13] During CNN training, the model learns the texture and edge information of seismic images, thereby restoring useful image content and improving resolution. Li *et al.*[14] proposed using a U-Net to achieve seismic image super-resolution. The encoder of U-Net captures contextual information through convolutional layers, pooling layers, and down-sampling operations, while the decoder gradually restores image features and improves resolution through upsampling and convolution operations, aided by skip connections. Min *et al.*[15] introduced D2Unet, a dual-decoder network based on the U-Net architecture, which simultaneously addresses the restoration of high-resolution image reconstruction and edge detection, thereby enhancing resolution while preserving critical edge information. In recent years, transformer-based networks have also achieved great success in improving the resolution of seismic images.[16,17] The self-attention mechanism, a cornerstone of the transformer architecture, dynamically attends to contextual information across relevant elements within a sequence, effectively capturing long-range dependencies and enhancing both model efficiency and expressiveness through parallel computation. The multi-head attention mechanism further enhances the model's feature extraction capacity by enabling simultaneous learning of input characteristics from multiple perspectives. In addition, in the field of seismic image super-resolution, other notable techniques include generative adversarial networks[18,19] and diffusion models.[20]

In this study, inspired by the CNN and the self-attention mechanism in the transformer, the authors propose a grouped-residual and multi-scale large-kernel attention network (GMLAN) to achieve seismic image super-resolution, combining convolution and multi-scale attention mechanisms. GMLAN is mainly composed of a feature extraction module (FEM) and an image reconstruction module (IRM). The FEM includes shallow feature extraction (SFE) and deep feature extraction (DFE). SFE utilizes convolutional layers to extract prominent large-scale features in seismic images, such as geological structures and shapes. DFE integrates group residuals and multi-scale large-kernel self-attention mechanisms to efficiently extract multi-scale features, reducing computational parameters while capturing detailed information such as layer interfaces and dip angles. The features extracted by SFE and DFE significantly influence the accuracy of subsequent image reconstruction. In the IRM, the fused shallow and deep features are upsampled and reconstructed through sub-pixel convolution to generate the final high-resolution seismic image. The authors construct a synthetic dataset whose features resemble those of field seismic images using a convolutional model and conduct supervised training on the GMLAN network. Consequently, the proposed method serves as a super-resolution processing approach for seismic images that effectively integrates algorithmic design with data. GMLAN provides a structural foundation for multi-scale feature capture and high-frequency recovery, achieved through its grouped-residual structures and multi-scale

large-kernel attention (MLKA) mechanisms. Meanwhile, the synthetic dataset, which is consistent with field data in geological and seismic characteristics, provides sufficient supervised training samples.[21] The effectiveness of the proposed model is validated through tests on both synthetic and field seismic images, demonstrating superior accuracy in recovering folds and faults in field seismic data.

In the comparative experiment section, the proposed approach is compared with U-Net, demonstrating that it not only achieves significant improvements in seismic image super-resolution but also exhibits faster convergence.

2. Methods

2.1. Network architecture

The architecture of the proposed GMLAN is shown in Figure 1. It combines convolution with multiple attention mechanisms and mainly consists of two parts: FEM and IRM. The FEM can be further subdivided into SFE and DFE.

The network first performed SFE on the input low-resolution seismic images, transforming the image space into a higher-dimensional feature space. Subsequently, the DFE was employed to establish a non-linear mapping relationship between low-resolution and high-resolution images, thereby reconstructing higher-frequency texture details. After feature extraction, residual connections were introduced to integrate the extracted shallow features with the deep features. This approach not only facilitates the extraction of high-dimensional image information but also ensures amplitude preservation in super-resolution reconstruction and reduces artifacts. In the final IRM, the fused features were first integrated, and their channels were adjusted using convolutional layers. Subsequently, the integrated features were upsampled via sub-pixel convolution to enhance the edge and texture details of the seismic image, thereby improving the resolution and yielding the final super-resolution reconstructed images.

The proposed GMLAN was trained on synthetic data, and the well-trained GMLAN was applied to the super-resolution prediction of field seismic images.

2.1.1. Feature extraction

The SFE was primarily implemented by a convolutional layer with a 3×3 convolution kernel. It can be expressed by the formula:

$$X_0 = L_{SF}\left(I_{in}\right) \tag{I}$$

Where $I_{in} \in R^{H\times W\times C_{in}}$ and $X_0 \in R^{H\times W\times C_s}$ refer to the low-resolution input image of the network and the output feature map of SFE, respectively. H and W represent the height and width of the input seismic images and feature maps, respectively, C_{in} and C_s denote the number of channels in the input image and the output feature maps of the shallow features, respectively. L_{SF} refers to the main layer of SFE.

In the DFE part, multi-branch residuals and large-kernel self-attention are the core components, which were executed in four stages. In DFE, four grouped residual and MLKA (GRMLKA) blocks were employed, corresponding to the four stages of DFE. In each stage of DFE, six layers of multi-residual groups were set. Furthermore, MLKA mechanisms were incorporated after the multi-branch residuals in the second, fourth, and sixth layers to enhance the precision and accuracy of feature extraction. Therefore, each DFE stage can actually be divided into three subgroups. It is worth noting that the number of multi-branch residual groups in each DFE stage should be set to an even number because the MLKA module was only added to the even-numbered layers. The operation sequence of each stage is as follows: first, pass through the first multi-branch residual block; then, perform normalization and pass through the feed-forward network (FFN); finally, conduct another normalization operation to complete the first layer of operations. In the second layer, unlike the first layer, a MLKA operation was carried out after the first multi-branch residual group. After this operation, normalization and subsequent processes were performed. This set of operations can be expressed by the following formula:

$$\begin{aligned}
X_{i,1} &= L_{RG}\left(X_{i-1}\right)\\
X_{i,2} &= L_{Norm}\left(X_{i,1}\right) + X_{i-1}\\
X_{i,3} &= L_{FFN}\left(X_{i,2}\right)\\
X_{i,4} &= L_{Norm}\left(X_{i,3}\right) + X_{i,2}\\
X_{i,5} &= L_{RSA-MLKAB}\left(X_{i,4}\right)\\
X_{i,6} &= L_{Norm}\left(X_{i,5}\right) + X_{i,4}\\
X_{i,7} &= L_{FFN}\left(X_{i,6}\right)\\
X_{i,8} &= L_{Norm}\left(X_{i,7}\right) + X_{i,6}
\end{aligned} \tag{II}$$

The entire process of extracting deep features is expressed as:

$$\begin{aligned}
X_i &= L_{GRMLKA}\left(X_{i-1}\right)\\
I_{DF} &= L_{Conv}\left(X_N\right) + X_0
\end{aligned} \tag{III}$$

Where X_i, X_{i-1} refer to the output of each GRMLKA, and the range of i is $1 \le i \le N$. Given that four stages were set in our model, we set $N = 4$. I_{DF} denotes the final output feature image obtained from DFE. L_{GRMLKA} represents the

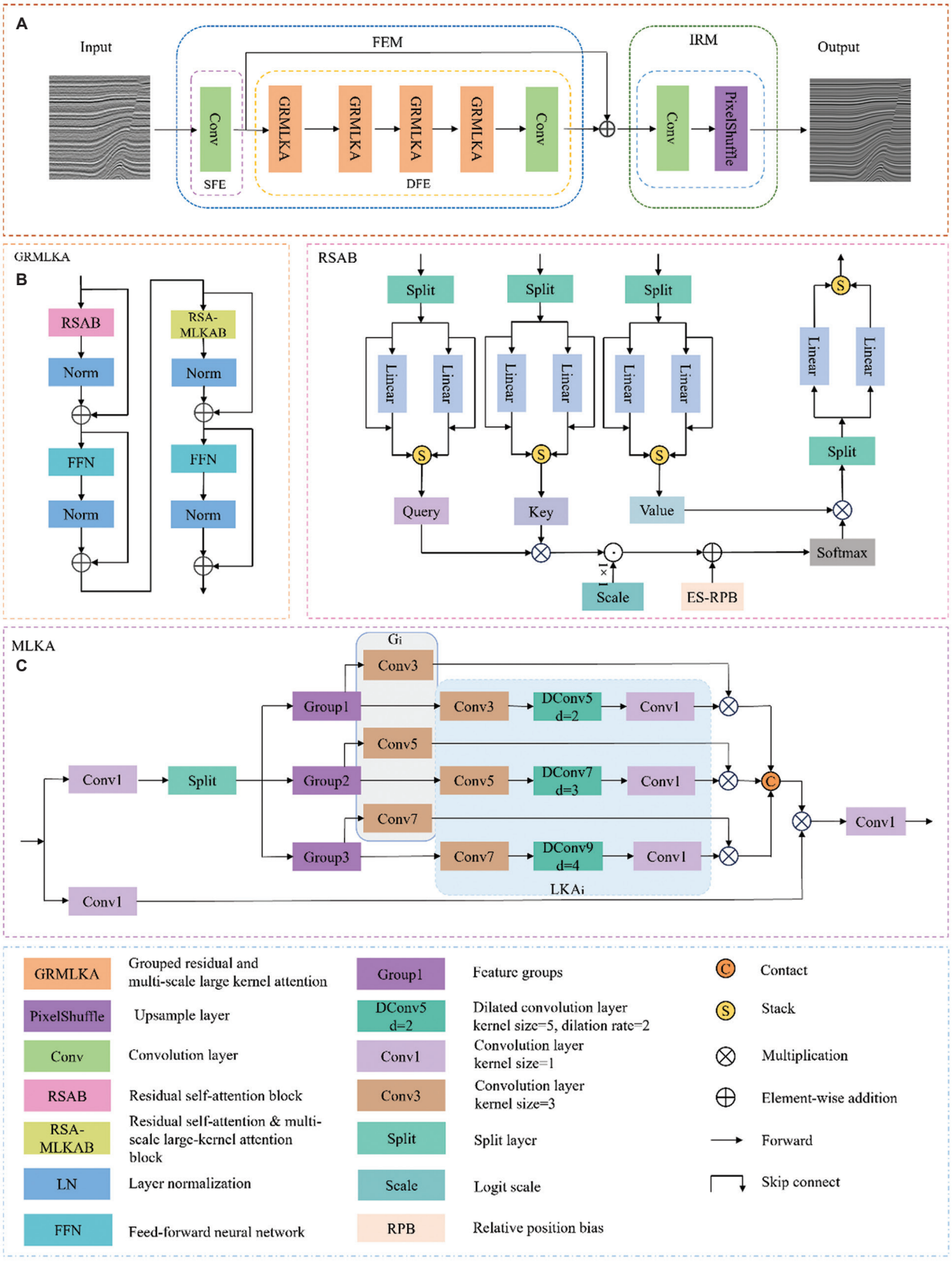

Figure 1. The architecture and components of the grouped-residual and multi-scale large-kernel attention network

four stages within the DFE process, and L_{Conv} represents the last convolutional layer within the DFE. Within each GRMLKA stage, the output feature map of each block is denoted as $X_{i,m}$. In **Equation (II)**, the index m represents integers from 1 to 8. Given that each stage is divided into three subgroups, the range of m is $1 \le m \le 24$. Specifically, when $m = 24$, $X_{i,24}$ is simply denoted as X_i, representing the output of the *i-th* GRMLKA. L_{RG} refers to the residual group layer, L_{Norm} denotes the layer normalization component, L_{FFN} indicates the FFN layer, and $L_{RSA\text{-}MLKAB}$ represents the fusion layer integrating residual groups and large-kernel attention.

The self-attention mechanism is highly effective at capturing long-range spatial dependencies within seismic images.[22] However, the process of generating the query (Q), key (K), and value (V) matrices and calculating their products often leads to computational redundancy.[23] To harness the benefits of the self-attention mechanism while mitigating this redundancy, we incorporated a residual self-attention block (RSAB) into the proposed GMLAN model, as depicted in Figure 1B. Specifically, we partitioned the input feature channels into two distinct groups and applied separate linear transformations and residual calculations to each group. The following equation illustrates the operation of grouping the input feature vectors:

$$\left[A_1, A_2\right] = S\left(A\right) \tag{IV}$$

Where A refers to the input feature vector. The operation S pertains to the grouping of features by channels, which are divided into two distinct groups, denoted as A_1 and A_2, with each group encompassing half of the channels.

We then solved each component and incorporated residual connections, merging the outcomes from the two branches to derive the Q, K, and V matrices. To illustrate, the computation of the Q matrix is detailed in the following formula:

$$Q^{'} = ST\left(A_1 + Linear\left(A_1\right), A_2 + Linear\left(A_2\right)\right) \tag{V}$$

Where

$$Linear\left(A_1\right) = A_1 \times {W_{q1}}^T + b_{q1} \tag{VI}$$

$$Linear\left(A_2\right) = A_2 \times {W_{q2}}^T + b_{q2} \tag{VII}$$

ST refers to the operation of stacking along a specified dimension. The term $Q^{'} \in R^{2\times B\times N\times C/2}$ represents the outcome after the merging operation. B signifies the batch size, N indicates the number of tokens obtained after flattening the input features, and C represents feature channels. *Linear* denotes the application of linear mapping operations to A_1 and A_2 individually. ${W_{q1}}^T$ and ${W_{q2}}^T$ are weight matrices, while b_{q1} and b_{q2} denote relative position bias terms. Following the computation of $Q^{'}$, feature rearrangement and merging operations were conducted to obtain the final Q matrix. The K and V matrices were derived using an analogous approach and were subsequently normalized. The attention was then calculated using the formula provided below:

$$Attention\left(Q,K,V\right) = Softmax\left(\left(Q\times K^T\right)\times Scale + B\right)\times V \tag{VIII}$$

Where Q, K, and V were normalized, *Scale* is a learnable scaling factor that dynamically adjusts the distribution of attention scores, enabling the model to adaptively learn the optimal similarity scale and enhance its focus on significant features. B refers to the continuous relative position bias generated by a multi-layer perceptron, which can effectively accommodate the long-range structural continuity inherent in seismic data. When combined with a grouped residual design, it enabled independent optimization of long-range association computations within each feature group. The attention mechanism implemented in RSAB is similar to cosine attention but differs using a learnable scaling factor instead of a fixed value and by incorporating continuous relative position biases. The resulting attention scores were individually subjected to linear projections and then combined to produce the final output of this module. This not only minimizes cross-channel interference but also halves the computational load, thereby avoiding redundancy in long-range calculations.

The grouped learning and computation operation not only captures interdependencies among different channels but also reduces computational complexity.[24] Specifically, before performing grouped calculations, each group has C channels, and the computational cost is C^2. After grouping, both the number of channels and the computational cost are halved compared to their values before grouping.

In the DFE stage, in addition to the grouped residual attention, we introduced a MLKA mechanism, as depicted in Figure 1C. Convolution kernels of varying sizes enabled the capture of seismic image features at multiple scales, thereby enhancing the interconnections among different receptive fields. Within the MLKA module, the output from the preceding stage was initially subjected to convolution to augment the number of input feature channels, facilitating the subsequent division of feature channels. Assuming the input features have n channels, the output features after this operation will have $2n$ channels, thereby increasing the non-linearity of the seismic image features. This step can be formulated as follows:

$$\left[X_{1:n}, X_{n+1:2n}\right] = C\left(X_{RSABO}\right) \tag{IX}$$

Where X_{RSABO} represents the output features from the RSAB. The operation C denotes the process of augmenting the feature channels and partitioning them into two identical groups. The notations $X_{1:n}$ and $X_{n+1:2n}$ refer to the two segments of the partitioned features, where 1:n indicates the range of the first half of the feature channels after dimensionality increase, and $X_{1:n}$ can alternatively be denoted as X_a. Similarly, $X_{n+1:2n}$ signifies the range of the latter half of the feature channels, and thus is also referred to as X_b. Both X_a and X_b have the same number of feature channels as X_{RSABO}.

Consequently, the input features, following the augmentation of feature channels, were split into two equivalent segments. Each segment underwent a convolutional layer to adjust the number of output feature channels, aligning them with the input requirements of subsequent operations. One segment was further subdivided into three parts along the feature channels for the upcoming large kernel attention (LKA_i) operation. These steps can be mathematically formulated as follows:

$$F_{Gn} = S\left(L_{C1}\left(X_a\right)\right) \tag{X}$$

Here, L_{C1} denotes the convolutional layer with a kernel size of 1 in Figure 1C, while F_{Gn} signifies different feature groups, with the subscript n indicating the group identifier, which can extend up to 3. Within the LKA_i module, there were three analogous pathways that processed the corresponding feature groups independently, namely Group1 to Group3 as depicted in the figure. Each pathway's LKA comprised three distinct convolutional layers. The initial layer consisted of standard convolutional layers with kernel sizes of 3 × 3, 5 × 5, and 7 × 7, enabling the extraction of local image features at varying scales. The smaller the convolution kernel, the more detailed the geological structures captured in the extracted feature map, indicating a richer content of high-frequency information. The subsequent layer employed dilated convolution with different kernel sizes and dilation rates; this differs from the first layer by expanding the receptive field of the convolutional kernel, thereby allowing it to capture broader global features.[25] By integrating small-kernel convolutions with dilated convolutions, the model can effectively capture both low-frequency information, such as large-scale geological structures in seismic data, and high-frequency features. This integration compensates for the insufficiencies in learning high-frequency components that can occur with large-kernel convolutions alone.[26] The final convolutional layer was utilized to modulate the feature channels and enhance the non-linearity of the output features. The procedure of LKA can be mathematically represented as follows:

$$LKA_i\left(F_{Gn}\right) = L_{C1}\left(L_{DC}\left(L_C\left(F_{Gn}\right)\right)\right) \tag{XI}$$

Where L_C refers to the initial standard convolutional layer within the LKA_i module, and L_{DC} signifies the dilated convolutional layer. However, convolution with large kernels may introduce artifacts. To address this issue, we incorporated spatial gating (G_i) into the network to enhance the image's feature representation capabilities.[27] The configuration of G_i mirrored that of the initial convolutional layer in the respective branches of LKA_i. The convolutional layers with different kernel sizes in the spatial gating focused on the weak signal features of varying intensities in the seismic images. After passing through G_i, the grouped features were integrated with the output features from LKA_i. The resulting three sets of integrated features were concatenated along the channel dimension to yield a feature map X_a', which matches X_a in terms of channel count. The final output of the MLKA module can also be viewed as an attention computation, which is mathematically expressed as follows:

$$O_{MLKA} = L_{C1}\left(CAT\left(LKA_i\left(F_{Gn}\right) \times G_i\left(F_{Gn}\right)\right)\right) \tag{XII}$$

Where CAT denotes the operation of concatenating along the feature channels, and O_{MLKA} refers to the overall output of the MLKA module.

The application of the MLKA module within the DFE not only facilitated the extraction of local features of seismic images at various scales but also leveraged dilated convolutions to expand the receptive field without additional computational overhead. This approach effectively addresses the issue of missing long-range feature mappings that can arise from focusing solely on local feature extraction.

2.1.2. Image reconstruction

In the IRM, we employed a 3 × 3 convolutional layer to adjust the channels of the fused deep and shallow features. Finally, through the upsampling layer, an image with dimensions twice those of the input image in each direction was generated, effectively reconstructing the image to a super-resolution version $I_{SR} \in R^{H \times W \times C_{out}}$.

I_{SR} refers to the reconstructed seismic super-resolution image, which also serves as the final output of the entire network, while H, W, C_{out} denote the height, width, and number of channels of the output image, respectively.

The entire image reconstruction process can be represented by the following formula:

$$I_{SR} = L_{upsample}\left(L_{Conv}\left(I_{DF} + I_{SF}\right)\right) \tag{XIII}$$

Where I_{SR} denotes the reconstructed super-resolution image, I_{SF} refers to the extracted shallow features, and I_{DF} represents the features extracted by the DFE module. $L_{upsample}$ refers to the upsampling layer in the reconstruction part, where we employed sub-pixel convolution for upsampling. Traditional interpolation techniques, such as bilinear and nearest neighbor interpolation, can enlarge low-resolution images to high-resolution counterparts. However, these methods often introduce image artifacts and blurring. In contrast, sub-pixel convolution, which utilizes convolutional operations and channel recombination, can enlarge low-resolution images without distortion. Furthermore, it significantly enhances image clarity and detail, thereby preserving a greater amount of detailed information.

2.2. Modeling of the synthetic dataset

For the training of GMLAN, we used synthetic seismic images. The generation of the synthetic dataset follows the method proposed by Wu *et al.*[28] A total of 2,000 pairs of synthetic data were generated, with 1,600 pairs allocated for training and the remaining 400 pairs reserved for validation and testing. Given that the GMLAN employed supervised training, it required a substantial amount of paired low-resolution and high-resolution image data. The high-resolution images served as the ground truth, while the low-resolution images were used as input. However, the availability of high-resolution field seismic images was limited, which restricted the amount of data that could serve as ground truth. In addition, the cost of obtaining high-resolution data was substantial. Therefore, in the proposed model training, a large number of synthetic seismic images were generated for training.

The synthetic images mainly simulated the distribution of geological structures based on the target data. First, a three-dimensional reflectivity model was established, and fold and fault structures were added to all horizontal layers. The frequency band of the high-resolution image was wider than that of the low-resolution image. Therefore, to obtain the corresponding high-resolution and low-resolution data, convolution operations were performed between the three-dimensional reflectivity models (with added folds and faults) and wavelets of different frequencies. Specifically, the three-dimensional reflectivity models were convolved with a high-frequency Ricker wavelet (35–55 Hz) to obtain the high-frequency seismic volumes, and then with a low-frequency Ricker wavelet (10–25 Hz) to obtain the corresponding low-frequency seismic volumes. Two-dimensional seismic slices were extracted from the high-frequency seismic volumes to serve as labels for actual training. Colored noise was added to the low-frequency seismic volumes to simulate real-world interference. Two-dimensional slices were then extracted from the low-frequency seismic volume using the same method and downsampled to obtain the low-resolution data required for training. The downsampled low-resolution seismic data and the high-resolution labels were paired. Finally, using the above-described method, 2,000 pairs of synthetic seismic data were generated and allocated for training, validation, and testing with an 8:1:1 ratio.

2.3. Loss function

In this study, we used a combined loss function consisting of the L1 loss and the multi-scale structural similarity index measure (MS-SSIM):

$$L_{Mix} = a \times L_{MS-SSIM} + (1-a) \times L_1 \tag{XIV}$$

Where

$$L_{MS-SSIM} = 1 - MS-SSIM(x,y) \tag{XV}$$

L_{Mix} refers to the combined loss function, L_1 denotes the L1 loss function, and $L_{MS\text{-}SSIM}$ signifies the MS-SSIM loss. The parameter a represents the weight of the $L_{MS\text{-}SSIM}$ component. As the sum of the weights for L_1 and $L_{MS\text{-}SSIM}$ equals 1, the weight assigned to L_1 was set to $1-a$. *MS-SSIM(x, y)* denotes the MS-SSIM, which was used to assess the similarity between the target image x and the output image y produced by the GMLAN across multiple scales. The L1 loss function is less sensitive to outliers, thereby demonstrating greater robustness in image reconstruction tasks. Meanwhile, MS-SSIM focuses on the structural information and perceptual quality of seismic images, evaluating image similarity across multiple scales. This multi-scale evaluation enables the model to capture richer details and enhance performance across different scales.[29]

Multi-scale structural similarity is an improvement over the structural similarity index measure (SSIM), which evaluates the SSIM at different scales and aggregates these values to obtain MS-SSIM:

$$MS-SSIM(x,y) = \left[l_M(x,y)\right]^{\alpha_M} \times \prod_{j=1}^{M} \left[c_j(x,y)\right]^{\beta_j} \times \left[s_j(x,y)\right]^{\gamma_j} \tag{XVI}$$

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{XVII}$$

Where M represents the number of scales (set to 5 in this study). α_M is the weight of the fifth scale, which equals 0.1333; α, β, and γ are the weights vector for each part, set to $\alpha = \beta = \gamma$ [0.0448, 0.2856, 0.3001, 0.2363, 0.1333], x and y are the output image of GMLAN and the label image, respectively; μ, σ^2, σ_{xy} represent the mean, variance,

and covariance of the corresponding images. C_1 and C_2 are stabilizing terms used to prevent division by zero. Structural similarity takes into account the luminance $l(x,y)$, contrast $c(x,y)$, and structural $s(x,y)$ information of the image, providing a more comprehensive evaluation that aligns more closely with the human visual system's perception of image quality.

Among them, brightness, contrast, and structural similarity can be expressed as:

$$l(x,y)=\frac{2\mu_x\mu_y+C_1}{\mu_x^2+\mu_y^2+C_1} \tag{XVIII}$$

$$c(x,y)=\frac{2\sigma_x\sigma_y+C_2}{\sigma_x^2+\sigma_y^2+C_2} \tag{XIX}$$

$$s(x,y)=\frac{\sigma_{xy}+C_2/2}{\sigma_x\sigma_y+C_2/2} \tag{XX}$$

After pre-training tests, we set $a = 0.6$.

2.4. Evaluation metrics

The peak signal-to-noise ratio (PSNR) is an indicator used to measure the quality of image reconstruction. It calculates the pixel-level error between the ground truth and the network's output based on the mean squared error (MSE). PSNR provides an intuitive numerical value describing the accuracy of the reconstructed image. Generally, the larger the PSNR value, the higher the quality of the reconstructed image. The PSNR can be expressed as follows:

$$PSNR=10\times\log_{10}\left(\frac{MAX^2}{MSE}\right) \tag{XXI}$$

Where

$$MSE=\frac{1}{\mathrm{n}}\sum_{i=1}^{n}\left(y_i-x_i\right)^2 \tag{XXII}$$

Where MAX denotes the maximum value of the image data (set to 1), MSE is the MSE, n denotes the sample size (i.e., the number of seismic images used for training). y_i represents GMLAN's prediction for the i-th seismic image, whereas x_i denotes the true value of the i-th seismic image. However, PSNR only describes the pixel-level error between images and cannot fully reflect the perception of the human visual system, especially in terms of image details and textures.

Therefore, to address this limitation, the SSIM was introduced as another evaluation indicator during training.

The range of SSIM value is [−1, 1]; a value closer to 1 indicates a higher similarity between the super-resolution reconstructed image and the labeled image.

2.5. Experimental setup

The experiment was conducted using PyCharm (PyCharm Community Edition, version 2024.3, JetBrains, Czech Republic) and the PyTorch (version 2.5.1, Meta Platforms, Inc., United States) GPU framework. A Compute Unified Device Architecture parallel computing framework was implemented (version 12.4, NVIDIA Corporation, United States), and training was performed on two NVIDIA RTX 4090 GPUs, each with 24 GB of memory.

To enhance training efficiency, the data were normalized during training. The input data were normalized to the range of [−1, 1] using the following method:

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

Where x denotes the variable to be normalized, which, in the context of the experiment, corresponds to the input training data. $x_{\max}$ and $x_{\min}$ represent the maximum and minimum values of x, respectively, and x^* indicates the normalized result.

To ensure the accuracy of the predicted images, the GMLAN outputs were denormalized before calculating the evaluation metrics:

$$x'=x^*\times\left(x_{\max}-x_{\min}\right)+x_{\min} \tag{XXIV}$$

Where x' signifies the denormalized output.

The parameter configurations for the experiment are detailed in Table 1.

3. Numerical tests

3.1. Comparison method

In this section, we compare the super-resolution performance of the proposed method with that of the U-Net. U-Net is highly flexible and can be adapted to various super-resolution tasks and datasets by

Table 1. Training parameters of different methods

Hyperparameter	U-Net	GMLAN (the proposed model)
Optimizer	Adam	Adam
Batch size	8	8
Epoch	400	400
Learning rate	[^{1}arn^{-3}, 1rni^{-4}, 1rni^{-5}]	[^{11}rn^{-3}, 1rni^{-4}, 1rni^{-5}]
Epochs for learning rate decay	[200, 300]	[200, 300]
Input channels	1	1
Parameters counts	17,409,537	3,377,140

Abbreviations: Adam: Adaptive moment estimation;
GMLAN: Grouped-residual and multi-scale large-kernel attention network.

adjusting parameters such as network depth, width, and convolution kernel size. It has been widely applied in numerous image processing tasks, including image super-resolution reconstruction, where it has demonstrated high effectiveness and adaptability.

To ensure the fairness of the comparison experiment and the persuasiveness of the results, we trained a U-Net on the same dataset and kept the hyperparameters (such as the number of epochs, learning rate, optimizer, etc.) consistent with those used in the proposed model. The specific parameter settings are shown in Table 1. After completing parameter and data preparation, we trained GMLAN and U-Net for the same number of epochs. The comparison of the training process is shown in Figure 2.

As depicted in Figure 2A, the loss curve of the U-Net model exhibits a rapid decline during the initial 50 epochs, after which the rate of decrease gradually slows. After 200 epochs, the loss reached a stable state. In contrast, the proposed approach demonstrated a significantly faster reduction in loss at the start of training, and the loss value consistently remained lower than that of the U-Net model throughout the entire training process.

As illustrated in Figure 2B and C, the proposed method consistently outperforms U-Net in terms of both PSNR and SSIM, particularly in the later stages of training. The SSIM value of U-Net increased rapidly during the initial 50 epochs, then slowed, and ultimately stabilized at approximately 0.85. In contrast, the proposed model maintained a stable SSIM value of around 0.91. Regarding PSNR, the U-Net's value eventually stabilized at 19 dB, whereas the proposed approach achieved a value exceeding 21 dB. Therefore, considering the loss function and these two metrics comprehensively, the proposed method demonstrated significant superiority over U-Net in all training indicators.

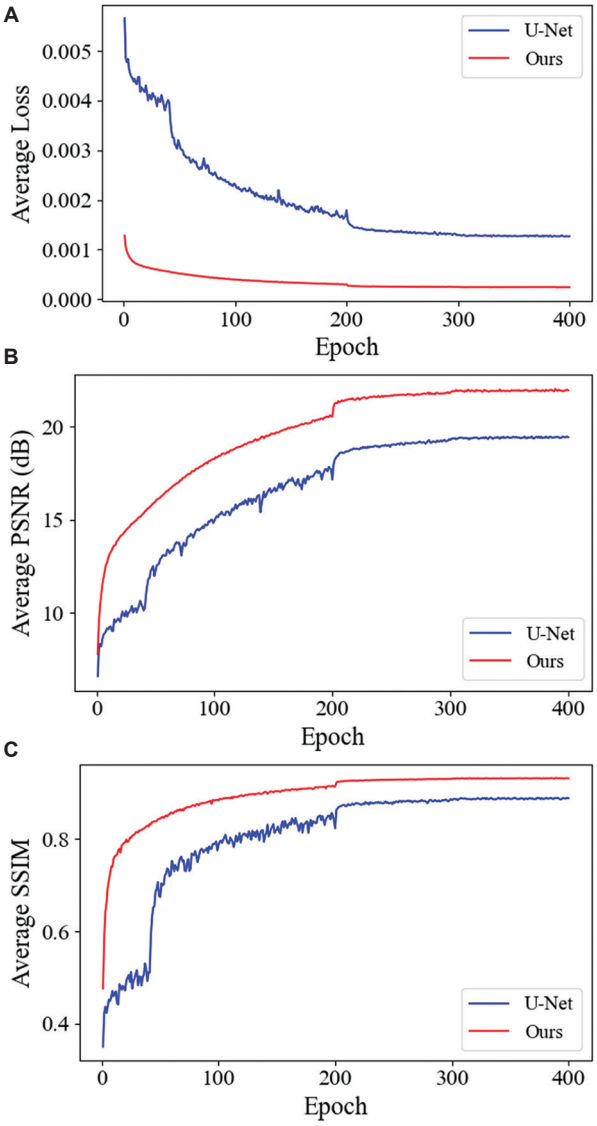

Figure 2. Comparison plot of average loss (A), peak signal-to-noise ratio (PSNR; B), and structural similarity index measure (SSIM; C) achieved by different methods with identical hyperparameters

3.2. Synthetic data testing

First, we evaluated the performance of GMLAN on the test dataset. As shown in Figure 3, at the positions indicated by the red arrows, the proposed model outperforms U-Net in fault recovery. The fault edges recovered by the proposed approach are clear and easily observable, while the fault edges within the solid red box recovered by U-Net appear smooth and blurred, resulting in the loss of edge features. In the area within the dashed yellow box, the interfaces between different layers are blurred in the prediction produced by the U-Net. We extracted the 40th trace from the data in Figure 3B-D to more effectively demonstrate the signal restoration performance. The amplitudes derived from both the proposed method and the U-Net closely resemble those of the ground truth. Nevertheless, the curve produced by the proposed model aligns more precisely with the ground truth, suggesting that the seismic images reconstructed by the proposed approach contain more detailed amplitude information. The coherence attribute serves as a highly effective and extensively utilized tool in the identification of structural and stratigraphic anomalies within seismic data.[30,31] To effectively highlight the variations in strata and the fault prediction results obtained by different methods, we have incorporated a coherence analysis, as depicted in Figure 4. The red solid-line frame within the image distinctly illustrates that the proposed approach outperforms Unet in fault recovery. This visual representation serves to underscore

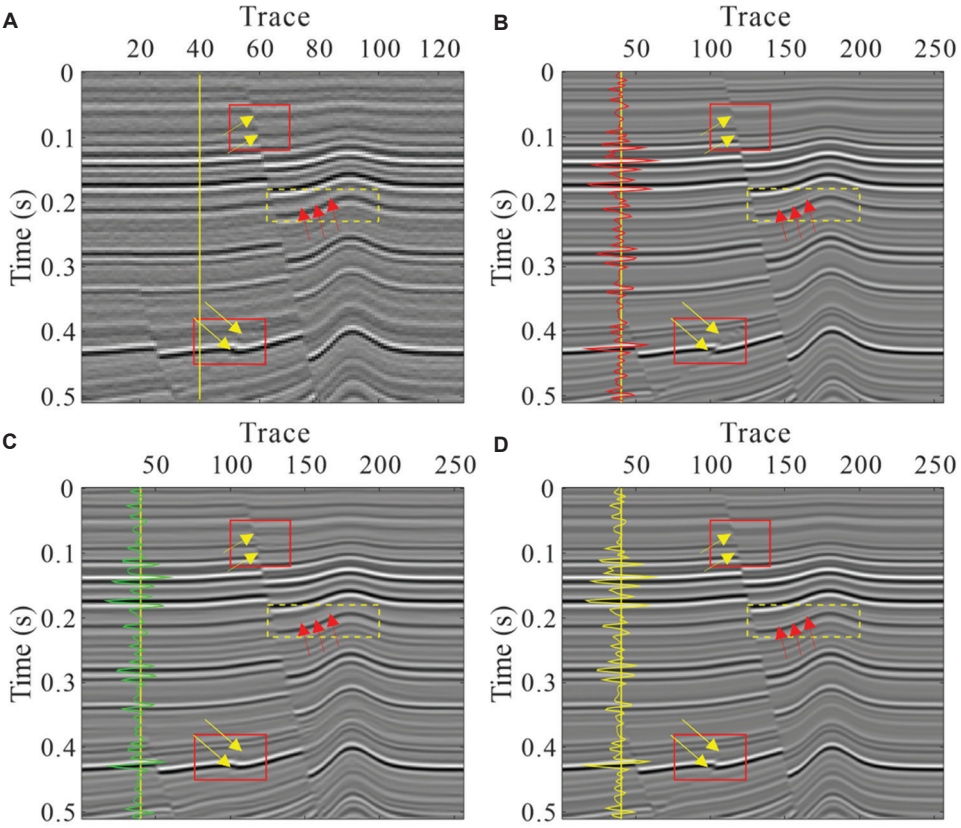

Figure 3. Comparison of super-resolution reconstruction results on synthetic data for different networks. (A) The low-resolution seismic image. (B) The ground truth of one synthetic dataset. (C) The super-resolution result predicted by the U-Net. (D) The super-resolution results predicted by the grouped-residual and multi-scale large-kernel attention network. The yellow line represents the 40th trace. In panels B, C, and D, the red, green, and yellow curves denote the amplitude of the corresponding seismic images at the 40th trace, respectively.

the enhanced capability of the proposed approach in accurately delineating geological features, which is a critical aspect of seismic image analysis. Figure 5 presents the spectra corresponding to the seismic images shown in Figure 3. Across most frequency bands, the proposed method closely matches the spectral curves of the ground truth, particularly in the mid-to-high frequency range (50–150 Hz). This consistency indicates that the proposed model excels at recovering the frequency components of low-resolution seismic images. The U-Net also approximates the ground truth with reasonable accuracy across some frequencies; however, it shows discrepancies at specific frequency points, especially in the high-frequency band. These discrepancies suggest that the U-Net is less effective than the proposed approach in recovering high-frequency components. In addition, Table 2 provides the evaluation metrics for the super-resolution prediction results of the synthetic data using the U-Net and the proposed method. The results demonstrate that the proposed model achieves significantly superior performance compared to U-Net in terms of PSNR, SSIM, and root mean square error. These findings corroborate the results presented in Figures 3-5.

3.3. Field data testing

To further verify the performance of GMLAN on field data, we selected a seismic image from the Netherlands F3 seismic survey for testing. The data, consisting of 512 traces contaminated by noise, were used as the input low-resolution seismic images. There is no ground truth for the field data. As shown in Figure 6, Figure 6A presents the field data from the Netherlands F3 block, Figure 6B illustrates the prediction result of the U-Net, and Figure 6C displays the prediction result of the proposed approach. Figures 6D-F are the coherence maps corresponding to Figure 6A-C, respectively. From the seismic images, it is evident that the prediction result of U-Net (Figure 6B) exhibits excessive smoothness at the fault locations indicated by the red arrows within the blue dashed box, leading to the loss of critical details. In contrast, the proposed method (Figure 6C) effectively restores the minor faults within the stratigraphic layers with high clarity. Within the two red solid boxes, the regions indicated by the red arrows show the recovered fault edges. The fault edges reconstructed by U-Net are less distinct compared to those recovered by

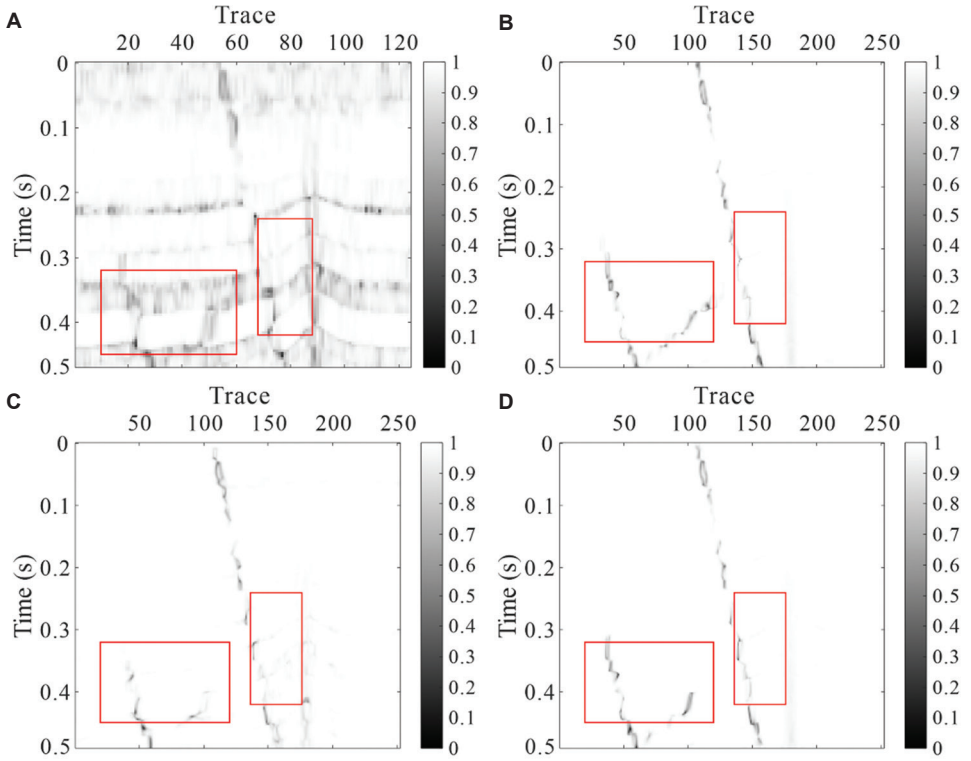

Figure 4. Comparison of seismic coherence results on synthetic data. (A) Coherence map of the low-resolution synthetic data. (B) Coherence map of the ground truth of the synthetic image. (C) Coherence map of the super-resolution result predicted by U-Net. (D) Coherence map of the super-resolution result predicted by the grouped-residual and multi-scale large-kernel attention network.

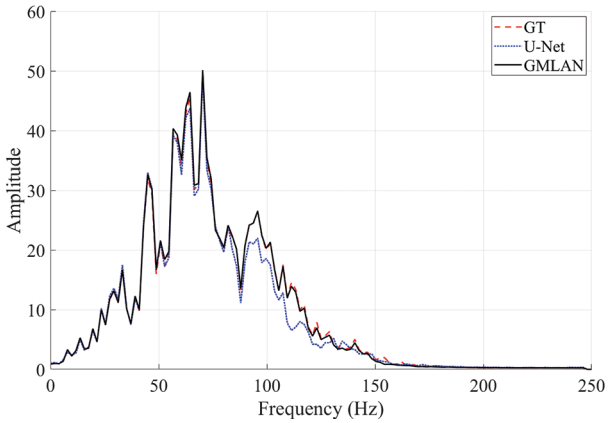

Figure 5. Spectral comparison for a synthetic seismic image
Abbreviations: GMLAN: Grouped-residual and multi-scale large-kernel attention network; GT: Ground truth.

GMLAN. In Figure 6A, the position indicated by the blue arrow within the red solid boxes reveals a faintly discernible fault structure in the input low-resolution seismic image. However, as depicted in Figure 6B, the U-Net result fails to accurately restore this fault structure and instead incorrectly reconstructs it as a continuous stratigraphic layer, thereby introducing continuity artifacts into the

Table 2. Comparison of evaluation metrics for super-resolution results

Method	PSNR (dB)	SSIM	RMSE
U-Net	13.68	0.68	0.25
GMLAN (the proposed model)	19.42	0.89	0.12

Abbreviations: GMLAN: Grouped-residual and multi-scale large-kernel attention network; PSNR: Peak signal-to-noise ratio; RMSE: Root mean square error; SSIM: Structural similarity index measure.

super-resolution seismic image. In contrast, Figure 6C illustrates that GMLAN clearly restores the fault structure. This comparison with U-Net on field seismic images underscores the superior detail fidelity and structural accuracy of GMLAN. Similarly, the coherence maps in Figure 6 clearly demonstrate that, at the corresponding positions to the seismic images, the stratigraphic structure details reconstructed by the proposed model exhibit notably enhanced resolution and precision. The objective of super-resolution reconstruction for seismic images was to preserve low-frequency signals without degradation while simultaneously expanding the high-frequency bands. Figure 7 presents the average spectra of the low-resolution field seismic images and the super-resolution results reconstructed using the two different methods. In

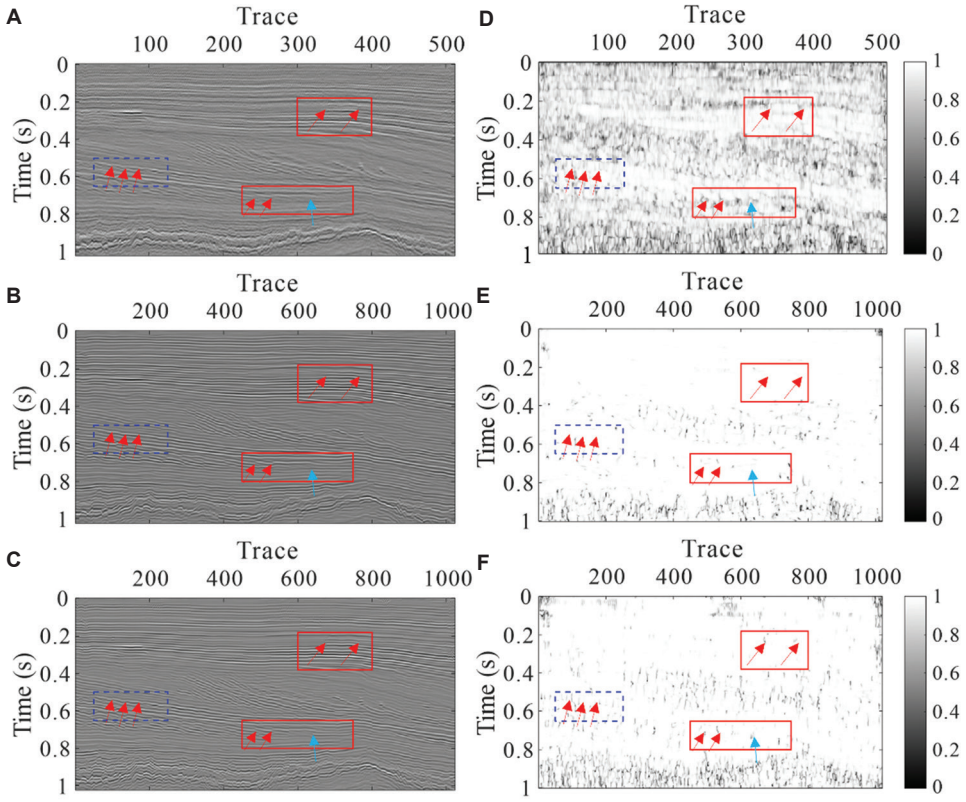

Figure 6. Comparison of super-resolution results on field data. (A) The low-resolution field data. (B) The reconstruction result of the U-Net. (C) The super-resolution reconstruction result of the proposed model. (D) Coherence map corresponding to the low-resolution field data in panel A. (E) Coherence map corresponding to the U-Net reconstruction result in panel B. (F) Coherence map corresponding to the reconstruction result of the proposed approach in panel C.

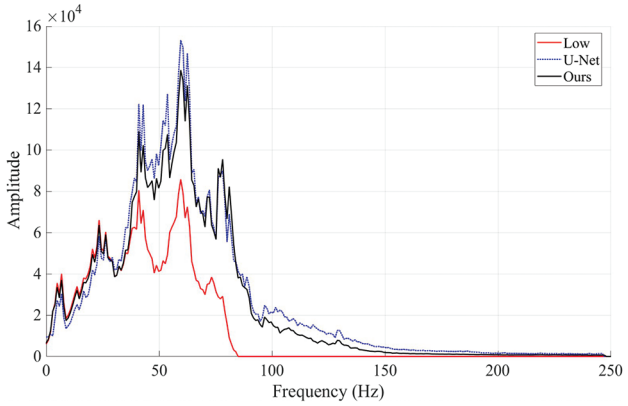

Figure 7. Comparison of average spectra of super-resolution results on field seismic images obtained by different methods

the low-frequency range, the proposed approach exhibits a higher degree of correlation with the low-frequency components of the low-resolution seismic images. In contrast, the proposed method not only effectively expands the high-frequency band but also maintains a significant superiority in preserving the consistency of low-frequency information.

In this section, we conduct a comparative analysis between GMLAN and U-Net across various datasets. These comparisons consistently demonstrate the superiority of the proposed method in the context of seismic image super-resolution reconstruction. Furthermore, the proposed model also demonstrates robust performance on field data, highlighting its strong generalization capability and reliability when applied to diverse datasets.

4. Discussion

4.1. Ablation test

In this section, the indispensable contributions of the network components are verified through ablation experiments, including the configuration of the number of stages for the MLKA and the DFE.

First, we removed all MLKA modules from the network. This variant is referred to as the grouped residual attention network (GRAN), which was trained using the same parameter settings and training data. In addition, we conducted experiments to explore the impact of different numbers of DFE stages on network performance. The experiments in this section are summarized as follows:

(i) GRAN, without MLKA, used to verify the impact of large-kernel attention on network performance; (ii) $MLKA_2$, a network with two DFE stages; (iii) $MLKA_3$, a network with three DFE stages; (iv) $MLKA_5$, a network with five DFE stages. Given that the proposed network incorporated four DFE stages, it was designated as $MLKA_4$.

After determining the number of experiments to be conducted, we set identical hyperparameters for all four networks and trained them using the same dataset. The evaluation metrics for the training results are presented in Table 3.

From the evaluation results, it can be observed that the network without MLKA exhibited the poorest performance. Although $MLKA_5$ performed slightly better than $MLKA_4$, the improvement in PSNR and SSIM was marginal. This minor enhancement came at the cost of an additional 3,528 seconds of training time, making it computationally inefficient. In contrast, $MLKA_4$ demonstrated a slight advantage over $MLKA_3$, achieving a 1.62db increase in PSNR. This enhancement is considered reasonable given the moderate increase in training time (an additional 3,528 s). Overall, the incorporation of MLKA and the implementation of the four DFE stages prove to be the optimal configuration.

Table 3. Comparison of evaluation index results from different methods

Method	PSNR (dB)	SSIM	Training time (s)
GRAN	19.20	0.8879	70,104
$MLKA_2$	19.90	0.9080	80,160
$MLKA_3$	20.32	0.9228	83,688
$MLKA_5$	21.98	0.9324	90,744
The proposed model (GMLAN/$MLKA_4$)	21.94	0.9321	87,216

Note: The subscript attached to "MLKA" denotes the number of deep feature extraction stages.
Abbreviations: GMLAN: Grouped-residual and multi-scale large-kernel attention network; MLKA: Multi-scale large-kernel attention; PSNR: Peak signal-to-noise ratio; SSIM: Structural similarity index measure.

To further illustrate the rationale and effectiveness of the proposed network design, we applied the four methods to the field seismic data for testing. Specifically, we selected a sample from the Netherlands F3 seismic survey as the low-resolution input for the network prediction. The prediction results of each network are shown in Figure 8.

The yellow dashed boxes in Figure 8 highlight each model's capability to recover stratigraphic structures. In Figure 8B-D, the recovered strata appear overly smooth, leading to a loss of detailed information. In contrast, Figures 8E and F demonstrate the recovery of complex stratigraphic structures, preserving richer details. The regions indicated by the yellow arrows within the red solid boxes in Figure 8 show that these areas are reconstructed as discontinuous strata (Figure 8B-D). However, as shown in Figure 8A, these positions should represent continuous stratigraphic structures, accurately depicted in Figure 8E and F. Compared with Figure 8E, the super-resolution result in Figure 8F demonstrates that the proposed approach achieves satisfactory outcomes even with fewer DFE stages.

To further compare the reconstruction performance of different networks, we used field data from the Kerry

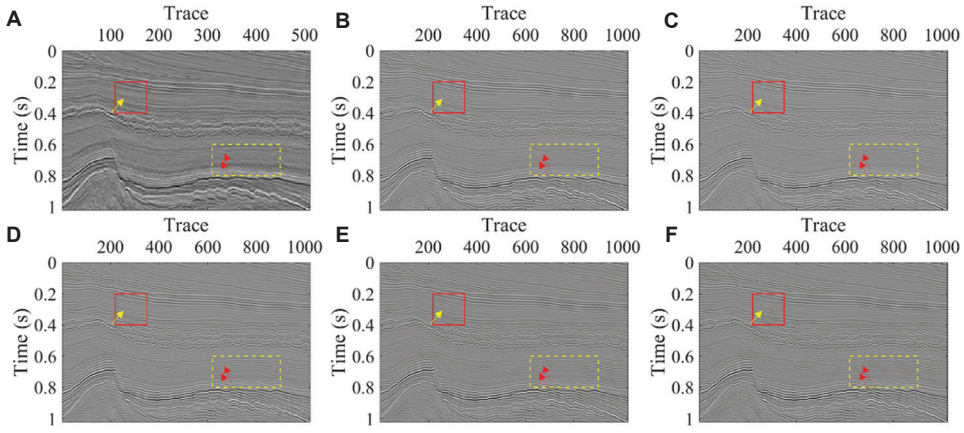

Figure 8. F3 field data prediction for networks with different structures. (A) Low-resolution field seismic image. (B) Super-resolution result by the grouped residual attention network. (C) Super-resolution result by $MLKA_2$. (D) Super-resolution result by $MLKA_3$. (E) Super-resolution result by $MLKA_5$. (F) Super-resolution result by $MLKA_4$.
Note: The subscript attached to "MLKA" denotes the number of deep feature extraction stages.
Abbreviation: MLKA: Multi-scale large-kernel attention.

block in New Zealand and extracted 512 traces for testing. Figure 9 displays the super-resolution results of these networks.

The yellow dashed boxes in Figure 9 highlight the stratigraphic recovery results of the different networks. The positions indicated by the red arrows reveal that Figure 9B-D fail to clearly delineate the detailed information of the stratigraphic structures, resulting in overly smooth reconstructions. In contrast, Figures 9E and F successfully recover each stratigraphic layer with greater clarity. The red solid boxes indicate the fault recovery regions, with the yellow arrows pointing to small fault structures. Figure 9B-D fail to recover these minor fault structures, instead reconstructing them as continuous stratigraphic layers with folded structures. Conversely, Figure 9E and F exhibit the highest accuracy in fault recovery.

Overall, the test and comparative analyses based on field data demonstrate that MLKA is crucial for improving the resolution of seismic images and reducing artifacts. Furthermore, the combined performance across different

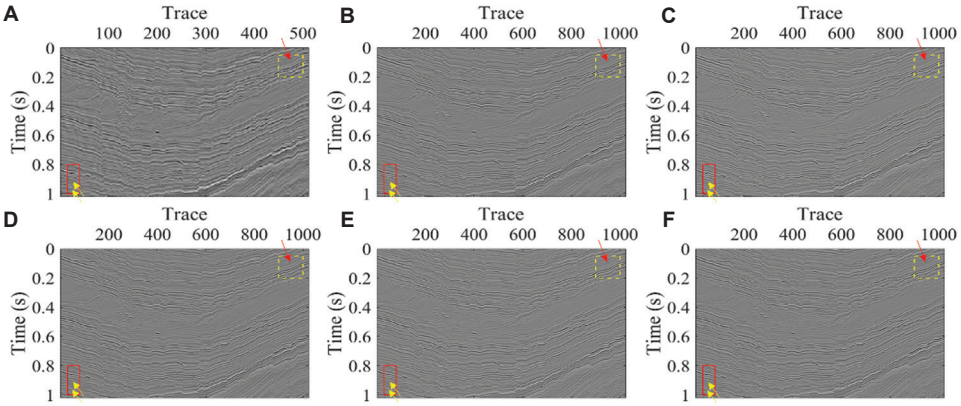

Figure 9. Kerry field data prediction for networks with different structures. (A) Low-resolution field seismic image. (B) Super-resolution result by the grouped residual attention network. (C) Super-resolution result by $MLKA_2$. (D) Super-resolution result by $MLKA_3$. (E) Super-resolution result by $MLKA_5$. (F) Super-resolution result by the proposed $MLKA_4$.
Note: The subscript attached to "MLKA" denotes the number of deep feature extraction stages.
Abbreviation: MLKA: Multi-scale large-kernel attention.

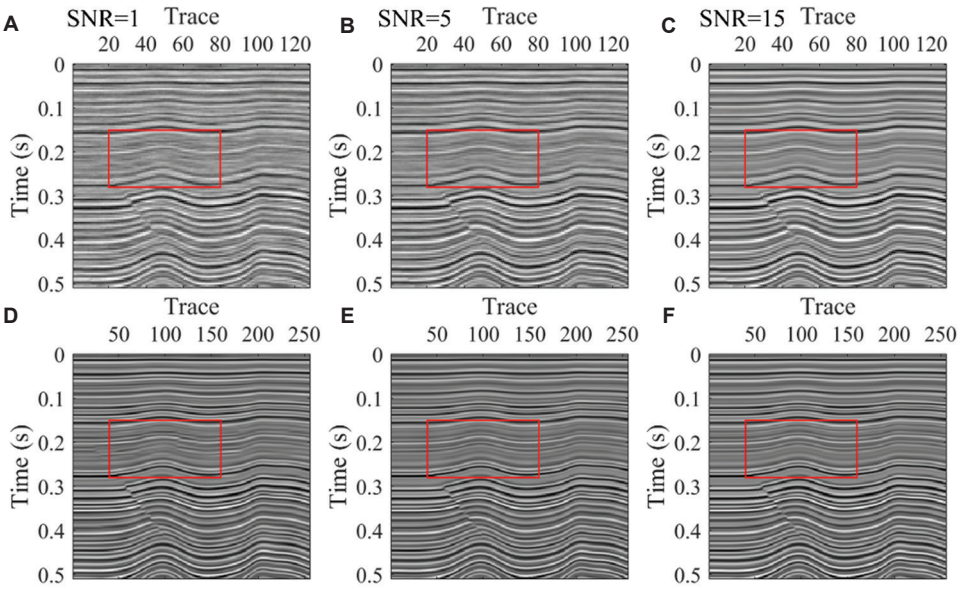

Figure 10. Comparison of denoising results for noisy synthetic seismic data at different signal-to-noise ratio (SNR) levels. (A) Noisy seismic data with SNR = 1 dB. (B) Noisy seismic data with SNR = 5 dB. (C) Noisy seismic data with SNR =15 dB. (D) Denoised result corresponding to panel A. (E) Denoised result corresponding to panel B. (F) Denoised result corresponding to panel C.

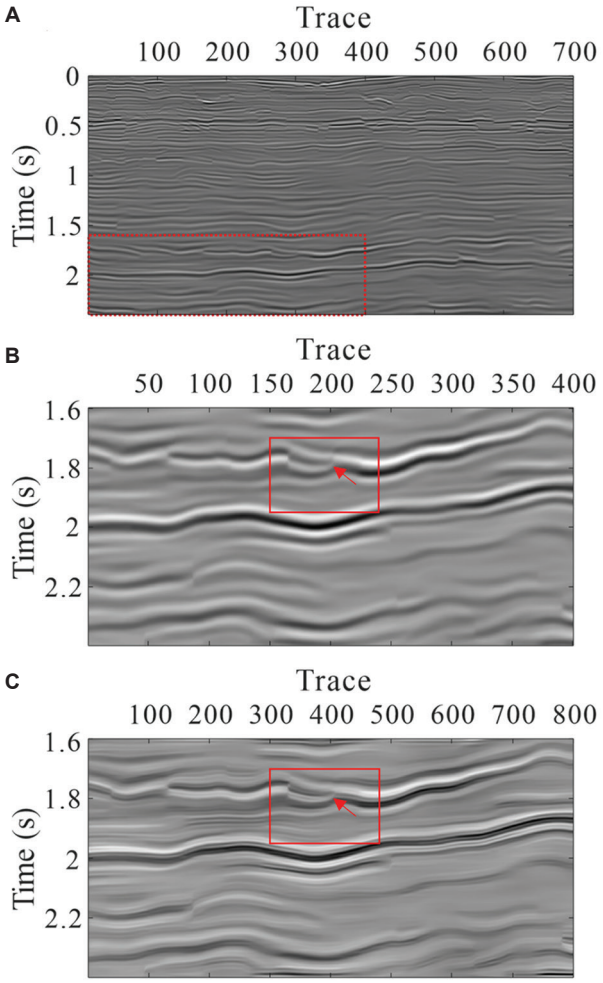

Figure 11. Test results of deep seismic data super-resolution. (A) Seismic data from the North Sea Volve Field in Norway. (B) Low-resolution deep seismic data. (C) Super-resolution result generated by the proposed approach.

DFE stage configurations suggests that a four-stage setup in the DFE component offers the best balance between accuracy and efficiency. This ablation study validates the interpretability and robustness of the proposed method.

4.2. Testing on noise robustness

We performed noise robustness assessments using synthetic seismic data engineered to exhibit distinct signal-to-noise ratio (SNR) characteristics. Specifically, we subjected high-resolution synthetic seismic data to a process of simple downsampling, followed by the introduction of colored noise at different intensities, as illustrated in Figure 10. The data represent SNR levels of 1 dB, 5 dB, and 15 dB, corresponding to strong, medium, and weak noise conditions, respectively.

On examination of the red solid-line boxes in Figure 10D-F, it is evident that the denoising performance at the strong noise level (1 dB) is less effective compared to the medium (5 dB) and weak (15 dB) noise levels. Notably, Figure 10D exhibits discontinuities in layer boundaries and the presence of artifacts, indicative of the challenges posed by high noise levels. Although the performance in low-SNR scenarios has not yet reached an optimal level, the majority of geological strata, including faults, have been successfully and clearly reconstructed. Figure 10E and F demonstrate a clearer restoration of the layer structure, highlighting the enhanced denoising capabilities of the proposed model under conditions of moderate and low noise, respectively. This comparative analysis underscores the adaptability and robustness of the proposed denoising technique in mitigating noise and preserving seismic data integrity across a spectrum of noise environments.

4.3. Testing on deep structures

In the preceding sections, we conducted super-resolution tests on field seismic data and denoising tests on synthetic seismic data. To showcase the adaptability of the proposed methodology, we applied it to the super-resolution of deep seismic data. In this segment, we selected deep seismic data from the North Sea Volve Field in Norway for the experiment, as depicted in Figure 11. The red dashed box in Figure 11A delineates the specific region of the deep seismic image that was extracted. This region is depicted in Figure 11B. The results illustrated in the figure indicate that the proposed model is effective in restoring layer boundaries. Within the red solid box, the minor fault indicated by the red arrow has also been clearly restored.

5. Conclusion

In this study, we propose GMLAN to achieve improved super-resolution for seismic images by integrating the benefits of group residual learning and large-kernel attention mechanisms. GMLAN comprises two primary components: feature extraction and image super-resolution reconstruction. Following the feature extraction phase, residual connections were employed to integrate the geological morphology and ground inclination characteristics of the seismic images. Subsequently, the output was fed into the IRM to produce super-resolution seismic images. A suite of comparative and ablation experiments demonstrated that the proposed network can significantly enhance and restore the high-frequency details of seismic images while preserving low-frequency information, indicating that the proposed approach is both highly effective and precise for seismic image super-resolution. Furthermore, the proposed method demonstrated superior noise robustness under different noise conditions. However, its efficacy in low SNR

scenarios has yet to be optimized. In future work, the authors plan to address this issue through two approaches: (i) implementing a multi-scale framework for adaptive noise analysis and suppression tailored to seismic data across various SNR levels; and (ii) developing an adaptive SNR enhancement strategy that fine-tunes super-resolution reconstruction parameters based on the input seismic images' SNR. By applying these strategies, the authors aim to substantially boost the proposed model's performance in low SNR environments and further enhance the super-resolution of seismic images.

Acknowledgments

None.

Funding

This work was supported by BGP's Science and Technology project (01-04-02-2024).

Conflict of interest

The authors declare that they have no competing interests.

Author contributions

Conceptualization: Anxin Zhang, Zhenbo Guo
Formal analysis: Shiqi Dong, Zhiqi We
Investigation: Zhiqi Wei
Methodology: Shiqi Dong, Zhenbo Guo
Software: Anxin Zhang, Shiqi Dong
Validation: Anxin Zhang, Zhenbo Guo
Visualization: Zhenbo Guo, Zhiqi Wei
Writing–original draft: Anxin Zhang, Zhenbo Guo
Writing–review & editing: Shiqi Dong, Zhiqi Wei

Availability of data

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

References

1. Soubaras R, Dowle R, Sablon R. Broadseis: Enhancing interpretation and inversion with broadband marine seismic. *CSEG Recorder*. 2012:37(7):40-46.

2. 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: Society of Exploration Geophysicists; 2012. p. 1-5.

   doi: 10.1190/segam2012-1067.1

3. Wang Y, Wang J, Wang X, Sun W, Zhang J. Broadband processing key technology research and application on slant streamer. *International Geophysical Conference, Beijing, China, 24-27 April 2018. Society of Exploration Geophysicists and Chinese Petroleum Society*. United States: Society of Exploration Geophysicists; 2018. p. 135-138.

   doi: 10.1190/IGC2018-034

4. Zhang YG, Wang Y, Yin JJ. Single point high density seismic data processing analysis and initial evaluation. *Shiyou Diqiu Wuli Kantan Oil Geophys. Prospect*. 2010;45:201-207.

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

5. Xiao F, Yang J, Liang B, *et al*. High-density 3D point receiver seismic acquisition and processing - a case study from the Sichuan Basin, China. *First Break*. 2014;32(1):81-90.

   doi: 10.3997/1365-2397.32.1.72598

6. Zhang H, Bao X, Zhao H, *et al*. High-precision deblending of 3-D simultaneous source data based on prior information constraint. *IEEE Geosci Remote Sens Lett.* 2025;22:1-5.

   doi: 10.1109/LGRS.2025.3526972

7. Shang XM, Diao R, Feng YP, Zhao CX. The application of spectral modeling method to high resolution processing of seismic data. *Geophys Geochem Explor*. 2014;38(1):75-80.

   doi: 10.11720/j.issn.1000-8918.2014.1.13

8. Wang D, Yuan S, Liu T, Li S, Wang S. Inversion-based non-stationary normal moveout correction along with prestack high-resolution processing. *J Appl Geophy*. 2021;191:104379.

   doi: 10.1016/j.jappgeo.2021.104379

9. Wu X, Ma J, Si X, *et al*. Sensing prior constraints in deep neural networks for solving exploration geophysical problems. *Proc Natl Acad Sci U S A*. 2023;120(23):e2219573120.

   doi: 10.1073/pnas.2219573120

10. Mousavi SM, Beroza GC, Mukerji T, Rasht-Behesht M. Applications of deep neural networks in exploration seismology: A technical survey. *Geophysics*. 2023;89(1):WA95-WA115.

    doi: 10.1190/geo2023-0063.1

11. Yuan S, Yu Y, Sang W, Xie R, Zhou C, Chen S. Seismic horizon picking using deep learning with multiple attributes. *IEEE Trans Geosci Remote Sens.* 2025;63:1-16.

    doi: 10.1109/TGRS.2025.3581462

12. Zeng D, Xu Q, Pan S, Song G, Min F. Seismic image super-resolution reconstruction through deep feature mining network. *Appl Intell*. 2023;53(19):21875-21890.

    doi: 10.1007/s10489-023-04660-y

13. Zhou R, Zhou C, Wang Y, Yao X, Hu G, Yu F. Deep learning with fault prior for 3-D seismic data super-resolution. *IEEE Trans Geosci Remote Sens.* 2023;61:1-16.

    doi: 10.1109/TGRS.2023.3262884

14. Li J, Wu X, Hu Z. Deep learning for simultaneous seismic image super-resolution and denoising. *IEEE Trans Geosci*

Remote Sens. 2022;60:1-11.

doi: 10.1109/TGRS.2021.3057857

15. Min F, Wang L, Pan S, Song G. D2UNet: Dual decoder U-net for seismic image super-resolution reconstruction. *IEEE Trans Geosci Remote Sens*. 2023;61:1-13.

doi: 10.1109/TGRS.2023.3264459

16. Liang J, Cao J, Sun G, Zhang K, Van Gool L, Timofte R. SwinIR: Image Restoration using Swin Transformer. *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). Montreal*, *BC*, *Canada*. 2021. p. 1833-1844.

doi: 10.1109/ICCVW54120.2021.00210

17. Zhong T, Zheng K, Dong S, Tong X, Dong X. Enhancing the resolution of seismic images with a network combining CNN and transformer. *IEEE Geosci Remote Sens Lett*. 2025;22:1-5.

doi: 10.1109/LGRS.2024.3495659

18. Zhong T, Yang F, Dong X, Dong S, Luo Y. SHBGAN: Hybrid bilateral attention GAN for seismic image super-resolution reconstruction. *IEEE Trans Geosci Remote Sens*. 2024;62:1-12.

doi: 10.1109/TGRS.2024.3492142

19. Lin L, Zhong Z, Cai C, Li C, Zhang H. SeisGAN: Improving seismic image resolution and reducing random noise using a generative adversarial network. *Math Geosci*. 2024;56(4):723-749.

doi: 10.1007/s11004-023-10103-8

20. Xiao Y, Li K, Dou Y, Li W, Yang Z, Zhu X. Diffusion models for multidimensional seismic noise attenuation and superresolution. *Geophysics*. 2024;89(5):V479-V492.

doi: 10.1190/geo2023-0676.1

21. Yuan S, Xu Y, Xie R, Chen S, Yuan J. Multi-scale intelligent fusion and dynamic validation for high-resolution seismic data processing in drilling. *Pet Explor Dev*. 2025;52(3):680-691.

doi: 10.1016/S1876-3804(25)60596-9

22. Zhou G, Zhi H, Gao E, *et al*. DeepU-Net: A parallel dual-branch model for deeply fusing multiscale features for road extraction from high-resolution remote sensing images. *IEEE J Sel Top Appl Earth Obs Remote Sens*. 2025;18:9448-9463.

doi: 10.1109/JSTARS.2025.3555636

23. Zhou Y, Li Z, Guo CL, Bai S, Cheng MM, Hou Q. SRFormer: Permuted self-attention for single image super-resolution. In: *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. United States: IEEE; 2023. p. 12734-12745.

doi: 10.1109/ICCV51070.2023.01174

24. Li Y, Deng Z, Cao Y, Liu L. GRFormer: Grouped residual self-attention for lightweight single image super-resolution. In: *Presented at: Proceedings of the 32nd ACM International Conference on Multimedia; 2024; Melbourne VIC, Australia*. United States: Cornell University.

doi:10.1145/3664647.3681554

25. Yan Z, Zi-Xin W, Lin-qI C, Hong-Li D. Research on microseismic event localization based on convolutional neural network. *JSE*. 2024;33(6):1-32.

26. Xu Y, Yuan S, Zeng H, *et al*. Frequency-dependent multiscale network for seismic high-resolution processing. *Geophysics*. 2025;90(4):V297-V312.

doi: 10.1190/geo2023-0682.1

27. Wang Y, Li Y, Wang G, Liu X. Multi-Scale Attention Network for single Image Super-Resolution. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*; 2024. p. 5950-5960.

28. Wu X, Geng Z, Shi Y, Pham N, Fomel S, Caumon G. Building realistic structure models to train convolutional neural networks for seismic structural interpretation. *Geophysics*. 2019;85(4):WA27-WA39.

doi: 10.1190/geo2019-0375.1

29. Zhao H, Gallo O, Frosio I, Kautz J. Loss functions for image restoration with neural networks. *IEEE Trans Comput Imaging*. 2017;3(1):47-57.

doi: 10.1109/TCI.2016.2644865

30. Lou Y, Zhang B, Wang R, Lin T, Cao D. Seismic fault attribute estimation using a local fault model. *Geophysics*. 2019;84(4):O73-O80.

doi: 10.1190/geo2018-0678.1

31. Yan B, Wang T, Ji Y, Yuan S. Multidirectional coherence attribute for discontinuity characterization in seismic images. *IEEE Geosci Remote Sens Lett*. 2022;19:1-5.

doi: 10.1109/LGRS.2022.3151686