Journal of Applied Mathematics and Physics, 2025, 13(12), 4280-4295
https://www.scirp.org/journal/jamp
ISSN Online: 2327-4379
ISSN Print: 2327-4352

Seismic Multiple Attenuation Method Based on Spectral Subtraction and Convolutional Dual Attention Mechanism

Mingming Fang[1], Zhonghua Ma[1,2*]

[1]College of Science, Tianjin University of Technology and Education, Tianjin, China
[2]Lvliang Vocational and Technical College, Lvliang, China
Email: *2585336915@qq.com

How to cite this paper: Fang, M.M. and Ma, Z.H. (2025) Seismic Multiple Attenuation Method Based on Spectral Subtraction and Convolutional Dual Attention Mechanism. *Journal of Applied Mathematics and Physics*, **13**, 4280-4295.
https://doi.org/10.4236/jamp.2025.1312237

Received: November 18, 2025
Accepted: December 15, 2025
Published: December 18, 2025

Abstract

Multiple reflections in seismic exploration data severely degrade subsurface imaging accuracy and interpretation reliability. Traditional multiple attenuation methods exhibit limited performance under complex geological conditions, while existing deep learning approaches struggle to simultaneously suppress high-energy multiples and preserve the structural continuity and amplitude fidelity of primary reflections. This paper proposes an enhanced U-Net architecture termed Seismic-CBAM-UNet, employing a two-stage cascaded strategy of "frequency-domain preprocessing + attention-enhanced network". The method first applies spectral subtraction in the frequency domain to preprocess raw seismic data, specifically targeting multiples concentrated in the low-frequency band. Subsequently, Convolutional Block Attention Modules are embedded within the network architecture, integrating both channel and spatial attention mechanisms to enable adaptive focusing on geologically significant reflection events. Experimental results demonstrate that our method outperforms Wiener filtering and wavelet thresholding across multiple evaluation metrics: achieving 1.62 dB SNR improvement, 61.3% multiple attenuation rate, 92.10% primary preservation rate, and 0.798 structural similarity index. The proposed approach effectively suppresses multiple interference while precisely maintaining geological structure continuity and amplitude characteristics, providing reliable technical support for high-precision imaging in complex seismic environments.

Keywords

Seismic Multiple Attenuation, Deep Learning, U-Net, Convolutional Block Attention Module, Spectral Subtraction

1. Introduction

Seismic exploration, as a fundamental geophysical technique for hydrocarbon detection [1] [2], mineral prospecting [3]-[5], and investigations of Earth's deep structure [6] [7], seeks to reconstruct subsurface geological architectures and lithological distributions with high accuracy and resolution by analyzing the propagation characteristics of artificially generated seismic waves through subsurface media [8] [9]. However, during practical data acquisition, raw seismic records are inevitably contaminated by strong multiple reflections resulting from complex wavefield interactions [10]. These multiples arise from repeated reflections, refractions, and diffractions of seismic waves between strong reflectors such as the free surface, sea surface, and interfaces of high-velocity layers. Characterized by complex propagation paths, high energy, broad spatial distribution, and substantial overlap with primary reflections in the time-space domain, multiples present major challenges for seismic interpretation [11]. Their presence not only gives rise to "false horizons" or "ghost reflections" in seismic images—potentially leading to serious misinterpretations of subsurface structures—but also interferes with essential processing workflows such as amplitude variation with offset analysis and inversion modeling, thereby significantly reducing the reliability of reservoir property estimation. More critically, as coherent noise with high amplitude, multiples inject spurious energy during seismic stacking and migration, severely degrading both the signal-to-noise ratio and lateral resolution of the final image. This ultimately constrains imaging fidelity and reduces exploration success rates. Consequently, the effective identification and suppression of multiples have emerged as pivotal technical bottlenecks in enhancing seismic data quality, ensuring the accuracy of geological modeling, and mitigating uncertainty in hydrocarbon exploration.

Traditional multiple attenuation techniques—such as predictive deconvolution [12] [13], Radon transform [14]-[16], and τ-p domain filtering [17] [18]—are predominantly developed under specific physical assumptions, such as the periodicity, linear trajectories, or hyperbolic moveout behavior of multiple reflections. Although these conventional approaches have achieved notable success under idealized or geologically simple conditions, their performance deteriorates markedly when encountering complex near-surface structures, highly nonlinear multiple events, or data with low signal-to-noise ratios. The fundamental limitation of these approaches lies in their strong dependence on accurate velocity models or prior wavefield information, which are often difficult to obtain reliably in practical exploration environments. Moreover, these methods are highly sensitive to parameter tuning and typically require extensive manual intervention and expert knowledge, which severely constrain their degree of automation. Furthermore, these traditional techniques generally exhibit limited generalization capability, struggling to adapt to data variations across different survey areas or geological contexts, thereby compromising the robustness and consistency of the entire processing workflow.

In recent years, the rapid evolution of artificial intelligence has introduced transformative paradigms to seismic signal processing. Encoder-decoder fully convolutional architectures, exemplified by U-Net [19], exploit their strong nonlinear modeling capacity and intrinsic end-to-end mapping capability to directly learn complex wavefield separation patterns from large collections of paired noisy and clean data, thereby eliminating the need for explicit physical modeling or labor-intensive parameter tuning. Wang *et al.* [20] propose a deep learning-based method for multiple suppression, using a convolutional encoder-decoder neural network trained with data augmentation to effectively reconstruct primaries and attenuate multiples in marine seismic data. Hu *et al.* [21] propose a lightweight deep learning approach—Mobilenet-Unet with depthwise separable convolution—for intelligent suppression of free-surface-related multiples in marine seismic data, achieving efficient and robust multiple attenuation through data augmentation and synthetic training sets. Geng *et al.* [22] propose a DC-LSTM method that combines data correlation with LSTM networks to accurately suppress multiples while preserving primaries in CMP gathers, outperforming the Radon transform in both synthetic and real data tests. Wang *et al.* [23] propose an unsupervised deep learning method based on adaptive virtual events and ensemble learning to suppress internal multiples without requiring labeled primary or multiple data, effectively addressing the lack of training labels and demonstrating strong performance on both synthetic and field data.

Although deep learning-based multiple attenuation techniques have achieved remarkable progress in recent years, a critical challenge persists in complex seismic environments: effectively suppressing high-energy multiples while simultaneously preserving the structural continuity and amplitude fidelity of primary reflections with high precision. To address this issue, this study proposes an enhanced U-Net-based architecture, termed Seismic-CBAM-UNet, which integrates spectral subtraction with attention mechanisms. The proposed method first applies spectral subtraction in the frequency domain to pre-process the raw seismic data, specifically attenuating multiple energy concentrated within the low-frequency band. Subsequently, Convolutional Dual Attention modules are embedded within the U-Net's encoder-decoder framework, incorporating both channel and spatial attention mechanisms. This design enables the network to adaptively emphasize geologically significant reflection events while effectively suppressing residual noise. Consequently, this strategy enhances the network's sensitivity to critical wavefield features and significantly improves the balance between multiple suppression and primary signal preservation.

2. Method

The proposed method, Seismic-CBAM-UNet, adopts a two-stage cascaded strategy—"frequency-domain preprocessing + attention-enhanced network"—to achieve efficient multiple suppression and high-fidelity reconstruction of primary reflections. First, before inputting the data into the network, spectral subtraction

is applied in the frequency domain to preprocess the raw seismic records. This step specifically attenuates the multiple energy concentrated in the low-frequency band [8] [9] while preserving the original phase information to maintain waveform integrity. Subsequently, the preprocessed data is fed into a U-Net-based deep convolutional network, in which a Convolutional Dual Attention module is embedded within every encoder and decoder block. The CDA module integrates both channel and spatial attention mechanisms, enabling the model to adaptively focus on geologically significant reflection events and suppress residual noise. Through this joint design, Seismic-CBAM-UNet achieves a fine-grained balance between multiple attenuation and preservation of primary reflection structures within an end-to-end learning framework.

2.1. Spectral Subtraction Preprocessing

Spectral subtraction is a well-established technique for noise reduction in the frequency domain. The fundamental principle involves subtracting an estimate of the noise spectrum from the magnitude spectrum of the noisy signal within the frequency domain, thereby approximating the spectrum of the clean signal. This method assumes that the noise and the desired signal are statistically independent in the time domain. Under the assumption of short-time stationarity, the noise power spectrum can be estimated from segments where the signal is absent.

Let the original noisy seismic signal be represented by $x(t)$, which can be formulated as:

$$x(t) = s(t) + n(t) \tag{1}$$

where $s(t)$ represents the primary reflection and $n(t)$ accounts for multiples or random noise. By applying the Short-Time Fourier Transform (STFT) to $x(t)$, its complex spectrum is obtained as follows:

$$X(k,m) = S(k,m) + N(k,m) \tag{2}$$

where k refers to the frequency bin index, and m corresponds to the time frame index.

Spectral subtraction operates exclusively on the magnitude spectrum, while the original phase is preserved to prevent waveform distortion. The estimated magnitude spectrum is expressed as:

$$\hat{A}_s(k,m) = \max\left\{\left|X(k,m)\right| - \alpha \cdot \hat{A}_n(k), \beta \cdot \left|X(k,m)\right|\right\} \tag{3}$$

where $\left|X(k,m)\right|$ is the magnitude spectrum of the noisy signal; $\hat{A}_n(k)$ is an estimate of the noise magnitude spectrum; $\alpha \geq 1$ is an over-subtraction factor used to enhance noise suppression and prevent residual noise; $\beta \in [0,1]$ is a spectral floor factor that prevents excessive subtraction and mitigates the occurrence of "musical noise"; and the $\max(\cdot)$ operation ensures a non-negative output.

The final enhanced signal is reconstructed via the inverse STFT (ISTFT):

$$\hat{s}(t) = \mathrm{ISTFT}\left\{\hat{A}_s(k,m) \cdot \mathrm{e}^{j\angle X(k,m)}\right\} \tag{4}$$

where $\angle X(k,m)$ is the original phase spectrum of the noisy signal.

2.2. Seismic-CBAM-UNet Architecture

The Seismic-CBAM-UNet employs the classic U-Net architecture as the baseline model. The U-Net architecture is a symmetric encoder-decoder structure that has been widely employed in tasks such as image segmentation and seismic wavefield separation. The encoder progressively extracts multi-scale features via stacked convolutions and downsampling operations, whereas the decoder incrementally restores spatial resolution through upsampling, simultaneously integrating fine-grained low-level details through skip connections to achieve pixel-wise mapping. This architecture provides robust end-to-end learning capabilities, effectively preserving the structure of primary reflections.

To further augment the network's capacity to capture critical wavefield structures, this work integrates the Convolutional Block Attention Module (CBAM) within the standard U-Net framework. Without substantially increasing computational overhead, as shown in Figure 1, CBAM enables fine-grained reweighting of feature maps by sequentially stacking the Channel Attention Module (CAM) and the Spatial Attention Module (SAM), thereby refining feature representations along two complementary dimensions: channel importance and spatial saliency.

The CAM is designed to model the relative importance of feature responses across different channels. By adaptively learning channel-wise weights, CAM

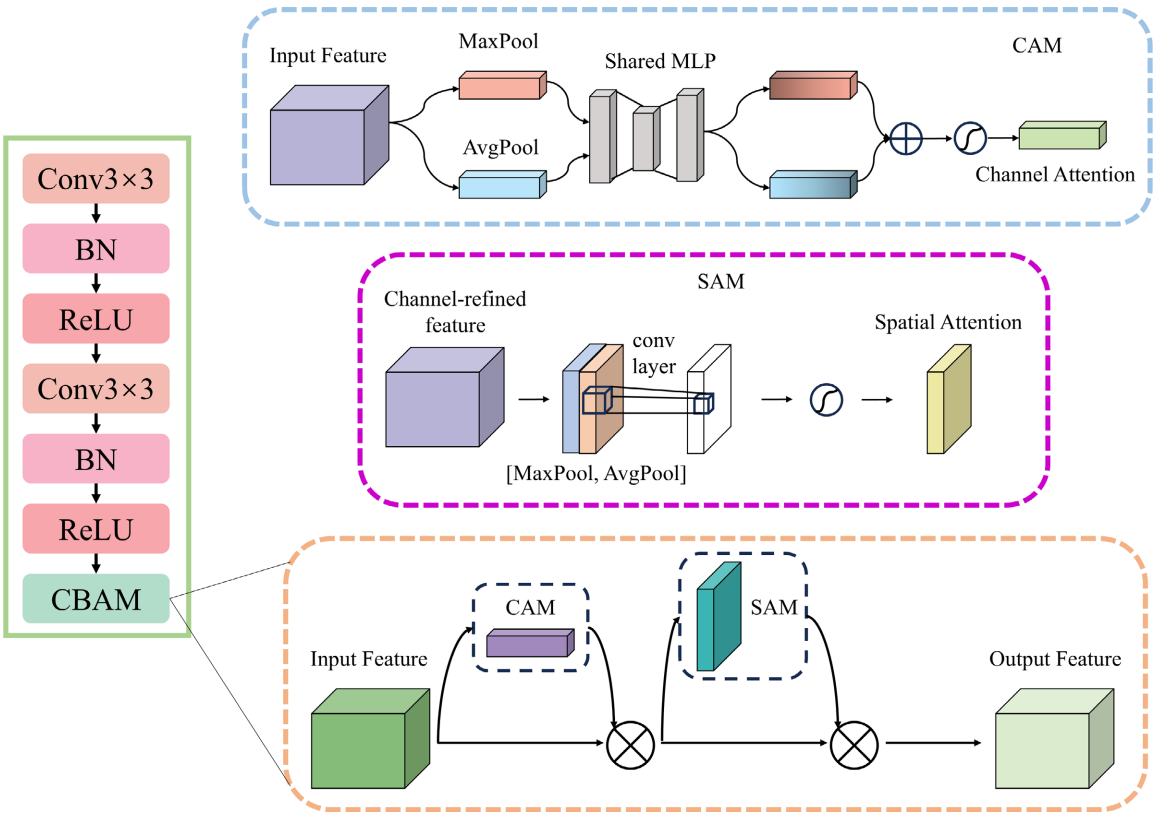

Figure 1. The structure of the CDA module.

enhances the responses of task-relevant channels while effectively suppressing redundant or noisy ones. Specifically, given an input feature map F, the module first applies Global Average Pooling (GAP) and Global Max Pooling (GMP) operations along the spatial dimensions to generate two distinct channel descriptors:

$$F_{\text{avg}}^{c} = \text{GAP}\left(F\right) \in \mathbb{R}^{C\times 1\times 1} \tag{5}$$

$$F_{\max}^{c} = \text{GMP}\left(F\right) \in \mathbb{R}^{C\times 1\times 1} \tag{6}$$

These two pooling operations capture the global semantic information of channel features from complementary perspectives: the "mean channel response" and the "peak channel response," respectively. This, in turn, provides complementary statistical representations. To capture nonlinear inter-channel dependencies, the two pooled results are subsequently fed into a shared-parameter Multilayer Perceptron (MLP). The MLP typically consists of two fully connected layers, with a dimensionality-reduction bottleneck in the intermediate layer, aimed at reducing computational complexity and enhancing cross-channel interaction:

$$W_{\text{avg}} = \text{MLP}\left(F_{\text{avg}}^{c}\right) \tag{7}$$

$$W_{\max} = \text{MLP}\left(F_{\max}^{c}\right) \tag{8}$$

Finally, the channel feature vectors output by the two MLP branches are element-wise summed and subsequently normalized using a Sigmoid activation function to generate the final channel attention weight map:

$$M_{c}\left(F\right) = \sigma\left(\text{MLP}\left(F_{\text{avg}}^{c}\right) + \text{MLP}\left(F_{\max}^{c}\right)\right) \in \mathbb{R}^{C\times 1\times 1} \tag{9}$$

where σ represents the Sigmoid function, with an output range of (0, 1), and is employed to adaptively rescale the original feature map along the channel dimension:

$$F' = F \otimes M_{c}\left(F\right) \tag{10}$$

This design enhances the sensitivity of the channel attention mechanism to diverse feature patterns by integrating complementary information from both average and max pooling, thereby improving the model's ability to discriminate and select critical channels. Meanwhile, the shared MLP structure effectively captures complex inter-channel dependencies without introducing excessive parameters, thereby reflecting the module's lightweight design and efficiency.

The SAM is designed to model the distribution of importance across various spatial locations in the feature map. By generating a spatial attention mask, SAM directs the network's focus towards more discriminative regions, thereby enhancing the spatial awareness of the features. Given the intermediate feature map $F' \in \mathbb{R}^{C\times H\times W}$ output from the Channel Attention Module, SAM first applies GAP and GMP along the channel dimension to produce two 2D spatial descriptor maps:

$$F_{\text{avg}}^{s} = \text{AvgPool}_{\text{channel}}\left(F'\right) \in \mathbb{R}^{1\times H\times W} \tag{11}$$

$$F_{\max}^{s} = \text{MaxPool}_{\text{channel}}\left(F'\right) \in \mathbb{R}^{1\times H\times W} \tag{12}$$

These operations extract spatial saliency cues from the perspectives of "mean channel response at each spatial location" and "maximum channel response at each spatial location," respectively, while preserving the spatial structure of the original feature map and reducing the channel dimension to 1. This results in compact input representations, which are used for subsequent spatial attention modeling. The two spatial descriptor maps are then concatenated along the channel dimension to form a two-channel spatial feature map:

$$F_{cat}^{s} = \text{Concat}\left(F_{\text{avg}}^{s}, F_{\max}^{s}\right) \in \mathbb{R}^{2\times H\times W} \tag{13}$$

This concatenation effectively integrates the complementary spatial information captured by the two pooling strategies, thereby enhancing the model's ability to represent diverse spatial patterns. Subsequently, a standard convolutional layer is applied to the concatenated feature map to model the spatial context, resulting in a single-channel spatial attention map:

$$M_s\left(F'\right) = \sigma\left(\text{Conv}_{7\times 7}\left(F_{cat}^{s}\right)\right) \in \mathbb{R}^{1\times H\times W} \tag{14}$$

where σ represents the Sigmoid activation function, normalizing the output values to the range (0, 1), thereby yielding an attention weight distribution across spatial locations. Finally, this weight map is applied to the input feature map F' via element-wise multiplication, achieving adaptive rescaling along the spatial dimension:

$$F'' = F' \otimes M_s\left(F'\right) \tag{15}$$

This design dynamically identifies and enhances semantically critical regions while suppressing background noise and irrelevant areas by integrating spatial statistics from both average and max pooling, coupled with a standard convolutional layer for local context modeling. The selection of the convolution kernel size effectively balances the preservation of local details with a sufficiently large receptive field, thereby enhancing the robustness and generalization capacity of the spatial attention mechanism.

As shown in **Figure 1**, the Convolutional Dual-Attention (CDA) module is designed as a unified composite block that integrates feature extraction and attention-based refinement. Specifically, the CDA module represents the complete sequence of two consecutive 3 × 3 convolutional layers, each followed by batch normalization and ReLU activation, and is immediately followed by the CBAM attention mechanism. In this architecture, the standard double-convolution block of the original U-Net is entirely replaced by the CDA module. The input feature map is first processed by the dual convolutions to extract high-level representations, which are then adaptively recalibrated by the CBAM before being passed to the subsequent downsampling or upsampling stage. CBAM adaptively recalibrates feature weights along both the channel and spatial dimensions, allowing the network to dynamically emphasize important features while suppressing irrelevant information. The convolutional layers are implemented without bias terms, instead relying on batch normalization for bias correction. Additionally, 3 × 3 con-

volutional kernels with unit padding are employed to maintain the spatial dimensions of the feature maps. By integrating an attention mechanism into conventional convolutional feature extraction, this module enhances the model's sensitivity to critical information and improves the effectiveness of feature representation. Given an input feature map $\mathbf{X}$, the CDA module is computed as:

$$\mathbf{X}_1 = \mathrm{Convl}_{3\times3}\left(\mathrm{BN}\left(\mathrm{ReLU}\left(\mathbf{X}\right)\right)\right) \tag{16}$$

$$\mathbf{X}_2 = \mathrm{Conv2}_{3\times3}\left(\mathrm{BN}\left(\mathrm{ReLU}\left(\mathbf{X}_1\right)\right)\right) \tag{17}$$

$$\mathbf{X}_{\mathrm{out}} = \mathrm{CBAM}\left(\mathbf{X}_2\right) \tag{18}$$

where $\mathrm{Convl}_{3\times3}$ and $\mathrm{Conv2}_{3\times3}$ denote the first and second convolutional operations, BN and ReLU represent batch normalization and ReLU activation, and $\mathbf{X}_{\mathrm{out}}$ denotes the final output of the CDA module, which serves as the input for the subsequent pooling or upsampling operations.

As shown in Figure 2, Seismic-CBAM-UNet employs a symmetric encoder-decoder architecture. In the encoder, the input layer first passes through a CDA module with 64 channels, followed by four successive max-pooling operations for progressive downsampling. After each downsampling step, the number of channels progressively doubles (64 → 128 → 256 → 512 → 1024), with a CDA module applied at each stage to perform feature extraction and attention weight computation. The encoder progressively captures high-level semantic features of seismic signals, concurrently reducing spatial resolution. In the decoder, reconstruction begins at the deepest layer with 1024 channels and progresses through four upsampling operations to gradually restore spatial resolution. After each upsampling step, the number of channels is halved (1024 → 512 → 256 → 128 → 64), with a CDA module similarly applied to ensure high-quality feature reconstruction. The primary role of the decoder is to recover fine-grained seismic signal structures from abstract representations. Between the encoder and decoder, four skip connections are established, linking feature maps at corresponding resolution levels. These connections fuse low-level detail information from the encoder with high-level semantic information from the decoder, effectively mitigating potential

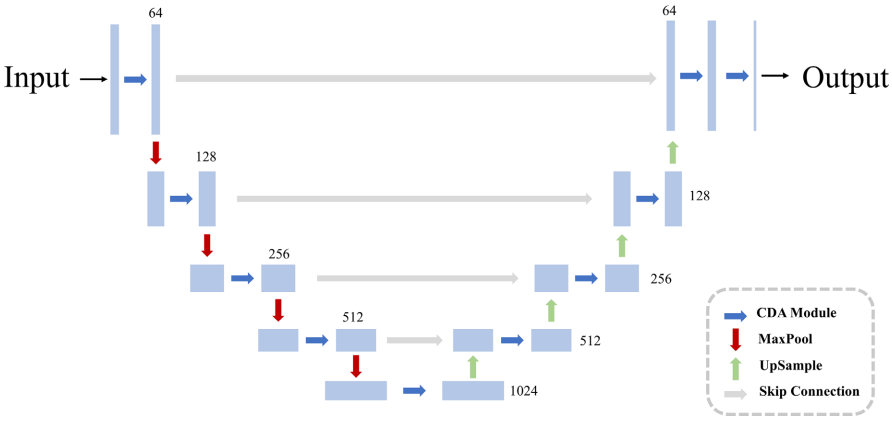

Figure 2. The structure of the Seismic-CBAM-UNet.

information loss during downsampling, and playing a crucial role in preserving the fine structures of seismic reflection events.

3. Experiments

3.1. Dataset and Implementation Details

The model was trained and evaluated using a private seismic dataset consisting of 500 paired samples. Each seismic section was cropped to a spatial resolution of 512 × 512 pixels. The implementation was carried out using the PyTorch framework on an NVIDIA 4090 GPU.

For the training strategy, we employed the Adam optimizer with an initial learning rate of 1×10^{-3} and a weight decay of 1×10^{-4} to ensure stable convergence and mitigate overfitting. The batch size was set to 16, and the network was trained for a total of 200 epochs. To further optimize the training process, we utilized the ReduceLROnPlateau scheduler to dynamically adjust the learning rate when the validation loss ceased to decrease.

3.2. Loss Function

This paper employs a combination of Mean Squared Error (MSE) and L1 loss as the objective function, formulated as follows:

$$L = L_{MSE} + 0.1 \times L_{L1} \tag{19}$$

$$L_{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2 \tag{20}$$

$$L_{L1} = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - \hat{y}_i\right| \tag{21}$$

where y_i denotes the ground-truth primary-wave seismic trace, $\hat{y}_i$ represents the network's predicted output, and N is the total number of samples in the data.

3.3. Evaluation Metrics

This paper employs the Primary Preservation Rate (PPR), SNR Improvement, Structural Similarity Index (SSIM), and Multiple Attenuation Rate (MAR) as evaluation metrics, which are calculated as follows:

$$\mathrm{PRP} = \left(1 - \frac{\left\|\hat{y} - y_{clean}\right\|_2}{\left\|y_{clean}\right\|_2 + \varepsilon}\right) \tag{22}$$

$$\Delta\mathrm{SNR} = \mathrm{SNR}\left(\hat{y}, y_{clean}\right) - \mathrm{SNR}\left(y_{dirty}, y_{clean}\right) \tag{23}$$

$$\mathrm{SNR}\left(x, x_{ref}\right) = 10 \cdot \log_{10}\left(\frac{\left\|x_{ref}\right\|_2^2}{\left\|x - x_{ref}\right\|_2^2 + \varepsilon}\right) \tag{24}$$

$$\mathrm{SSIM}\left(x, y\right) = \frac{\left(2\mu_x\mu_y + C_1\right)\left(2\sigma_{xy} + C_2\right)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)} \tag{25}$$

$$\mathrm{MAR} = \left(1 - \frac{\left\|\hat{y} - y_{clean}\right\|_2}{\left\|y_{dirty} - y_{clean}\right\|_2 + \varepsilon}\right) \tag{26}$$

where $\hat{y}$ is the network's predicted output, y_{clean} is the ground-truth primary-wave label, y_{dirty} is the input image, and $\varepsilon = 10^{-8}$ is a numerical stability term.

4. Results

To evaluate the effectiveness of the proposed method, comparative experiments were conducted against wavelet thresholding and Wiener filtering. The experimental results are summarized in Table 1. As shown in Table 1, the proposed method consistently achieves the best performance across all four core evaluation metrics, demonstrating its clear superiority as a preprocessing strategy for seismic multiple attenuation. Specifically, the ΔSNR attains +1.62 dB, corresponding to a net gain of 0.69 dB relative to Wiener filtering. This improvement not only indicates that the energy distribution of the model output more closely approximates that of the true primary wave, but also signifies a fundamental shift in the processing paradigm—from "introducing distortion" to "effectively enhancing" signal fidelity. Moreover, the MAR attains 61.3%, indicating that Seismic-CBAM-UNet, when trained with spectral subtraction preprocessing, can effectively suppress the energy of multiples embedded in the raw data. Furthermore, the Primary Preservation Rate (PPR) increases to 92.10%, and the Structural Similarity Index (SSIM) attains 0.798, confirming that the model effectively preserves the amplitude, phase, and continuity of primary reflections while suppressing noise—without introducing common artifacts such as over-smoothing or structural discontinuities. The foundation of this comprehensive performance gain lies in the strong correspondence between the spectral subtraction mechanism in the frequency domain and the physical characteristics of seismic multiples. In marine and shallow-land seismic surveys, multiples are typically concentrated within the low-frequency band, exhibiting high energy and relatively stable time-frequency distributions, with spectral energy predominantly residing in the 5 - 20 Hz range. By contrast, useful primary reflections are primarily distributed within the mid-to-high frequency range (20 - 80 Hz). By applying the STFT to project the signal into the time-frequency domain, spectral subtraction adaptively attenuates each frequency bin according to an estimated noise spectrum, thereby enabling precise "peak clipping" of multiple energy—particularly in the low-frequency region. Crucially, this method preserves the original phase spectrum and modifies only

Table 1. Quantitative comparison of multiple attenuation performance among different methods.

Method	PRP (%)	ΔSNR (DB)	SSIM	MAR (%)
Wiener filtering	91.62	+0.93	0.789	52.4
wavelet thresholding	91.85	+1.35	0.791	58.7
Ours	92.10	+1.62	0.798	61.3

the magnitude spectrum, thereby maximally maintaining the temporal waveform integrity during reconstruction. This effectively prevents phase-induced distortions, such as alignment warping or travel-time shifts—factors that are critical for subsequent geological interpretation and horizon delineation.

Figure 3 illustrates the superior performance of Seismic-CBAM-UNet in suppressing seismic multiples. The original noisy data contains strong and chaotic interfering waves that severely obscure weak yet geologically significant reflection events in the background. In particular, in the first and second rows, the noise energy is so dominant that the primary geological structures are almost completely obscured. After processing with the proposed method, the noise is substantially suppressed. Compared with the original noisy data, the background becomes markedly cleaner, and the signal-to-noise ratio is greatly improved. Continuous and coherent reflection events representing true geological interfaces—previously buried beneath noise—such as the curved reflector in the first row and the horizontal and dipping reflectors in the second and third rows, are now clearly revealed. Furthermore, when compared with the ideal ground-truth data, the proposed method not only effectively removes noise but, more importantly, accurately identifies and preserves geologically meaningful information without introducing perceptible artifacts. This finding clearly demonstrates the model's exceptional capability in maintaining structural fidelity.

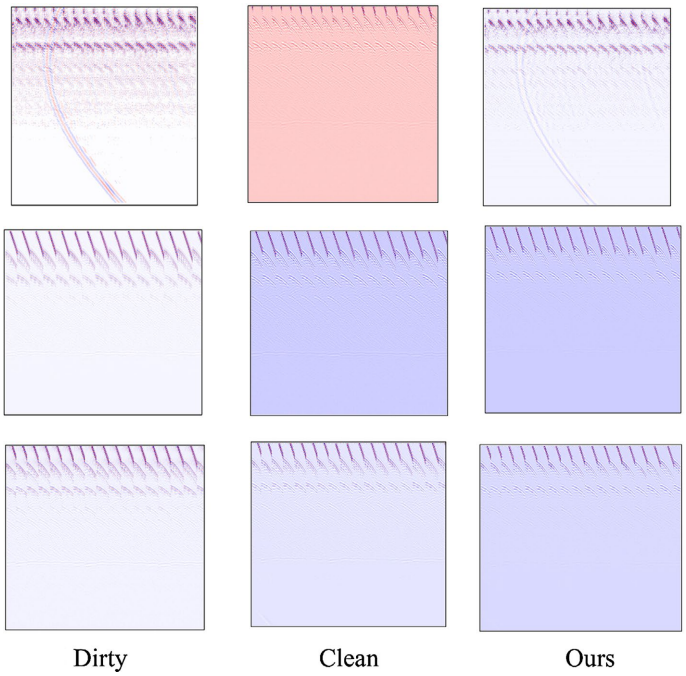

Figure 3. Comparison of denoising performance on seismic data sections.

Figure 4 compares the original multiple-wave data, data processed by three different preprocessing methods, and the primary-wave target data. In the original multiple-wave data, dense arc-shaped multiples are clearly visible in the upper section, forming distinct "false reflectors" that severely obscure the underlying

valid reflections. In the middle and lower sections, although the reflectors are generally continuous, they suffer from intense amplitude fluctuations and local discontinuities due to multiple interference. After spectral subtraction preprocessing, the energy of the arc-shaped multiples in the upper section is significantly attenuated—nearly eliminated—allowing the deeper reflectors beneath to become clearly visible. The reflection structures in the middle and lower sections are well preserved, with stabilized amplitudes and restored continuity of reflectors. Overall, the image "cleanliness" is greatly enhanced, closely resembling the primary-wave target. Following wavelet-threshold preprocessing, the suppression of multiples is slightly less effective than spectral subtraction: faint band-like artifacts remain in the upper section, though their intensity is substantially reduced. The middle and lower reflectors remain well-defined with sharp edges, particularly excelling in preserving high-frequency details. After Wiener filtering preprocessing, the image appears smoother overall. Multiples in the upper section are partially attenuated but not as thoroughly as with the previous two methods. Reflectors in the middle and lower sections become slightly blurred, especially at steeply dipping interfaces, where edge sharpness is compromised.

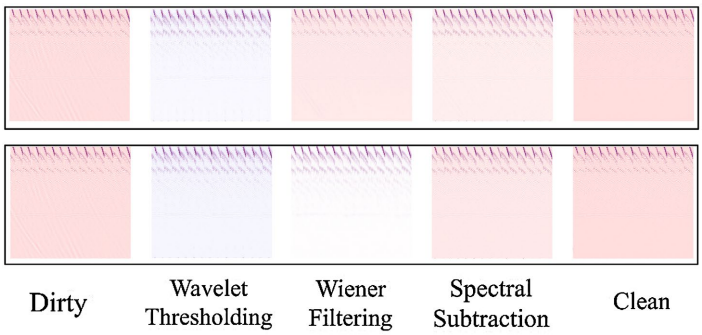

Figure 4. Comparison of multiple suppression results using different preprocessing methods.

Figure 5 displays the improvement maps, which quantify the performance gains of each preprocessing method over using the raw data directly, using a blue-to-red colormap. The improvement map corresponding to spectral subtraction is almost entirely dominated by deep blue: the vast majority of pixels—from top to bottom and left to right—show significant positive enhancements. Particularly in the upper region where multiples are most energetic, the blue saturation is extremely high, indicating that this method achieves its strongest noise suppression precisely where it is most needed. This localized superiority substantially exceeds improvements in other areas and strongly corroborates spectral subtraction's status as the optimal preprocessing approach.

While this study primarily benchmarks against traditional signal processing methods to validate the fundamental efficacy of the proposed architecture, it is valuable to contextually compare the Seismic-CBAM-UNet with existing deep learning paradigms cited in the introduction [20]-[23].

Unlike standard encoder-decoder models [20] or lightweight architectures like

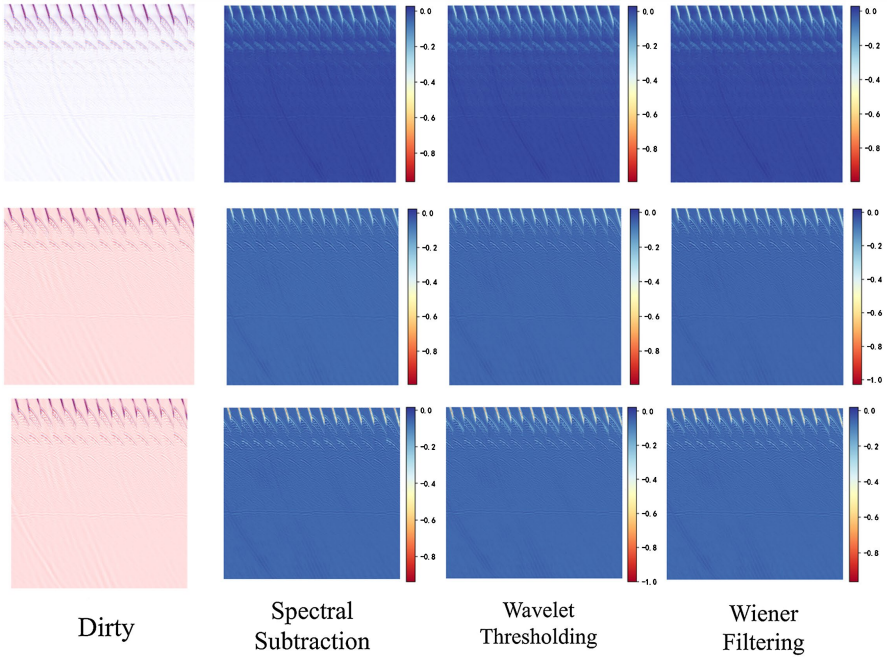

Figure 5. Improvement maps showing the performance gain of different preprocessing methods.

MobileNet-Unet [21] which rely purely on data-driven feature extraction, our method explicitly integrates domain knowledge and attention mechanisms. The spectral subtraction step incorporates the physical characteristic of low-frequency multiple concentration, acting as a physics-guided prior that simplifies the mapping task for the network. Furthermore, compared to sequence-based models like DC-LSTM [22], the embedded Convolutional Dual Attention modules allow our network to capture long-range spatial dependencies and subtle structural details more effectively in the 2D image domain. Although unsupervised methods [23] offer advantages when labels are scarce, our supervised strategy, enhanced by the dual-attention and frequency-domain preprocessing, ensures superior amplitude fidelity and structural preservation when high-quality training data is available.

5. Discussion

Despite the superior performance demonstrated in the experiments, the proposed Seismic-CBAM-UNet has potential limitations. A primary constraint arises from the spectral subtraction preprocessing stage, which relies on the assumption that multiple energy is predominantly concentrated in the low-frequency band (5 - 20 Hz) while primary reflections occupy higher frequencies.

In geological environments with significant frequency overlap—such as deep reservoirs where primary signals are naturally attenuated to lower frequencies—the spectral subtraction step might inadvertently suppress useful primary energy alongside the multiples. Although the subsequent Convolutional Dual Attention mechanisms are designed to recover and focus on structural details using spatial features, the initial loss of spectral information could potentially limit the final

reconstruction fidelity in such extreme scenarios. Future iterations of this method could explore adaptive spectral weighting or end-to-end architectures that learn frequency decomposition implicitly to mitigate this dependency.

6. Conclusions

This paper presents Seismic-CBAM-UNet, a novel seismic multiple attenuation method that integrates spectral subtraction with convolutional dual attention mechanisms to effectively address the fundamental challenge of balancing multiple suppression and primary signal preservation in complex seismic environments. Experimental results demonstrate superior performance compared to traditional methods across key metrics including SNR improvement, multiple attenuation rate, primary preservation rate, and structural similarity. The method's core advantages include: 1) Leveraging the physical characteristic that seismic multiples concentrate in low-frequency bands through targeted spectral subtraction preprocessing, effectively attenuating dominant multiple energy; 2) Embedding CBAM attention mechanisms to enable adaptive recognition and enhancement of geologically significant reflection features while suppressing residual noise; and 3) Preserving original phase information to avoid waveform distortion and ensure reliable seismic interpretation.

The method demonstrates excellent performance not only on synthetic data but also shows strong generalization capability on real seismic data, providing an efficient and robust solution for seismic data processing workflows. Future work will explore the application of this approach to three-dimensional seismic data processing, further optimize the network architecture for more complex geological scenarios, and investigate unsupervised or weakly-supervised learning strategies to reduce dependency on large volumes of labeled training data.

Funding

This paper is partially supported by Fundamental Research Program of Shanxi Province (Grant No. 202303021211245).

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Sun, L., Fang, C., Sa, L., Yang, P. and Sun, Z. (2015) Innovation and Prospect of Geophysical Technology in the Exploration of Deep Oil and Gas. *Petroleum Exploration and Development*, **42**, 454-465. https://doi.org/10.1016/s1876-3804(15)30038-0

[2] Howard, R.J., Wells, C.J., Michot, T.C. and Johnson, D.J. (2014) Effects of Disturbance Associated with Seismic Exploration for Oil and Gas Reserves in Coastal Marshes. *Environmental Management*, **54**, 30-50. https://doi.org/10.1007/s00267-014-0274-2

[3] Malehmir, A., Durrheim, R., Bellefleur, G., Urosevic, M., Juhlin, C., White, D.J., *et al.* (2012) Seismic Methods in Mineral Exploration and Mine Planning: A General Over-

view of Past and Present Case Histories and a Look into the Future. *Geophysics*, **77**, WC173-WC190. https://doi.org/10.1190/geo2012-0028.1

[4] Eaton, D.W., Adam, E., Milkereit, B., Salisbury, M., Roberts, B., White, D., *et al.* (2010) Enhancing Base-Metal Exploration with Seismic Imaging. *Canadian Journal of Earth Sciences*, **47**, 741-760. https://doi.org/10.1139/e09-047

[5] Malehmir, A., Maries, G., Bäckström, E., Schön, M. and Marsden, P. (2017) Developing Cost-Effective Seismic Mineral Exploration Methods Using a Landstreamer and a Drophammer. *Scientific Reports*, **7**, Article No. 10325. https://doi.org/10.1038/s41598-017-10451-6

[6] Seredkina, A.I. (2021) The State of the Art in Studying the Deep Structure of the Earth's Crust and Upper Mantle beneath the Baikal Rift from Seismological Data. *Izvestiya, Physics of the Solid Earth*, **57**, 180-202. https://doi.org/10.1134/s1069351321020117

[7] Hao, T., You, Q., Liu, L., Lv, C., Xu, Y., Li, Z., *et al.* (2013) Joint Land-Sea Seismic Survey and Research on the Deep Structures of the Bohai Sea Areas. *Acta Oceanologica Sinica*, **32**, 13-24. https://doi.org/10.1007/s13131-013-0383-4

[8] Yilmaz, Ö., Doherty, S.M. and Yilmaz, O. (2001) Seismic Data Analysis: Processing, Inversion, and Interpretation of Seismic Data. Society of Exploration Geophysicists, 231-233.

[9] Verschuur, D.J. (2013) Seismic Multiple Removal Techniques: Past, Present and Future. EAGE.

[10] Kelamis, P.G. and Verschuur, D.J. (2000) Surface-Related Multiple Elimination on Land Seismic Data—Strategies via Case Studies. *Geophysics*, **65**, 719-734. https://doi.org/10.1190/1.1444771

[11] Shi, Y. and Xing, X.L. (2011) Investigation Progress on Surface-Related Multiple Suppression: Review and Outlook. *Progress in Geophysics*, **26**, 2046-2054.

[12] Wei, K., Zhang, H. and Zeng, F. (2025) Multiples Suppression in Common-Offset GPR Data Based on Correlation-Predictive Deconvolution. *Pure and Applied Geophysics*, **182**, 1617-1636. https://doi.org/10.1007/s00024-025-03701-6

[13] Zhang, Y., Wang, D., Hu, B., Zhang, J., Gong, X. and Chen, Y. (2024) Enhanced Offshore Wind Farm Geophysical Surveys: Shearlet-Sparse Regularization in Multi-Channel Predictive Deconvolution. *Remote Sensing*, **16**, Article 2935. https://doi.org/10.3390/rs16162935

[14] Pan, Y., Qiang, Y., Liang, W., Huang, W., Wang, N., Wang, X., *et al.* (2024) A Transcranial Multiple Waves Suppression Method for Plane Wave Imaging Based on Radon Transform. *Ultrasonics*, **143**, Article ID: 107405. https://doi.org/10.1016/j.ultras.2024.107405

[15] Xie, J., Wang, X., Wang, X., Wu, D., Zeng, H. and Jin, B. (2021) Multiple-Suppression Method Using the λ-F Domain High-Resolution Parabolic Radon Transform with Curvature Magnification. *Applied Geophysics*, **21**, 169-178. https://doi.org/10.1007/s11770-021-0883-5

[16] Ma, J., Zhao, K. and Liao, Z. (2025) A Fast Amplitude Preserving Three-Parameter 3D Parabolic Radon Transform and Its Application on Multiple Attenuation. *Petroleum Science*, **22**, 163-177. https://doi.org/10.1016/j.petsci.2024.06.011

[17] Lu, J., Zhang, R., Li, Q. and Hong, X. (2024) Multiple Suppression by Hyperbolic Vector Median Filtering of Deep-Water Ocean-Bottom PS Waves. *IEEE Transactions on Geoscience and Remote Sensing*, **62**, 1-11. https://doi.org/10.1109/tgrs.2024.3357739

[18] Yang, J., Chen, J., Zeng, Q., Yan, Z., Liu, H. and Wang, X. (2025) Multiple Wave Suppression Method for Shallow Water Seismic Data with Strong Impedance Contrast Layer in the South Yellow Sea Laoshan Uplift Area. *Marine Geophysical Research*, **46**, Article No. 30. https://doi.org/10.1007/s11001-025-09588-1

[19] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W. and Frangi, A., Eds., *Medical Image Computing and Computer-Assisted Intervention—MICCAI* 2015, Springer, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28

[20] Wang, K., Hu, T., Wang, S. and Wei, J. (2022) Seismic Multiple Suppression Based on a Deep Neural Network Method for Marine Data. *Geophysics*, **87**, V341-V365. https://doi.org/10.1190/geo2021-0206.1

[21] Hu, G., Li, Y., Yang, S., Zhang, H., Liu, X., Li, Y., *et al.* (2025) Intelligent Suppression of Marine Seismic Multiples Using Deep Learning Methods. *Journal of Ocean University of China*, **24**, 967-978. https://doi.org/10.1007/s11802-025-6016-7

[22] Geng, H., Peng, S., Cui, X., He, T., Xu, N. and Du, W. (2024) Accurate Multiple Wave Suppression Using Data Correlation and Long Short-Term Memory Networks. *IEEE Geoscience and Remote Sensing Letters*, **21**, 1-5. https://doi.org/10.1109/lgrs.2024.3487618

[23] Wang, K., Hu, T., Zhao, B. and Wang, S. (2023) An Unsupervised Learning Method to Suppress Seismic Internal Multiples Based on Adaptive Virtual Events and Joint Constraints of Multiple Deep Neural Networks. *IEEE Transactions on Geoscience and Remote Sensing*, **61**, 1-18. https://doi.org/10.1109/tgrs.2023.3243106