Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# A Review of Deep-learning-based Seismic Data Denoising and Its Promising Paradigm Shift to Foundation Models

**Xintong Dong**
Jilin University
**Zhengyi Yuan**
Jilin University
**Changxin Wei**
Jilin University
**Wenshuo Yu**
Jilin University
**Shiqi Dong**
dsq1994@126.com
Northeast Electric Power University
**Jun Lin**
Jilin University

**Research Article**

**Keywords:** Foundation model, Seismic data denoising, Pre-training, Downstream tasks.

**Posted Date:** October 23rd, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-11188693/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:** No competing interests reported.

# A Review of Deep-learning-based Seismic Data Denoising and Its Promising Paradigm Shift to Foundation Models

Xintong Dong[1], Zhengyi Yuan[1], Changxin Wei[1], Wenshuo Yu[1], Shiqi Dong[2*], Jun Lin[1]

[1]Jilin University, State Key Laboratory of Deep Earth Exploration and Imaging, College of Instrumentation and Electrical Engineering, Changchun, China.

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

*Corresponding author(s). E-mail(s): dsq1994@126.com;
Contributing authors: dxt@jlu.edu.cn; yuanzhengyi0224@163.com;
cxwei24@mails.jlu.edu.cn; yuws24@mails.jlu.edu.cn; lin_jun@jlu.edu.cn;

**Abstract**

Denoising is a long-standing and widely-concerned topic in seismic data processing, for it can significantly increase the signal-to-noise ratio of seismic data. Numerous deep learning (DL) methods have shown promising denoising performance, but most of them are task-specific and focus on a certain type of seismic background noise. The real condition that seismic datasets are often contaminated by various types of noises motivates us to explore a well-generalized and versatile DL model for seismic data denoising. Recently, in the fields of computer vision and natural language processing, foundation models (FMs) pre-trained on vast datasets demonstrate outstanding adaptability and generality across diverse downstream tasks. This paradigm offers a promising path to address the challenges faced by task-specific DL denoising models, such as poor generalization, retraining from scratch for different noise, and the lack of labeled data. We first provide a brief review of traditional seismic denoising methods, followed by a comprehensive review of existing DL-based denoising methods categorized by noise type. Furthermore, we conduct a case study on a dedicated seismic denoising foundation model termed SeisDeFM. This is the first study in geophysical

research to develop and validate a seismic denoising foundation model on pre-stack gathers with diverse noise conditions. Experimental results demonstrate that, compared with task-specific DL baselines, SeisDeFM achieves superior denoising performance and cross-noise generalization by the advantages of sufficient pre-training and appropriate downstream adaptation, and it effectively preserves weak reflection events while suppressing complex noise in pre-stack seismic data.

**Keywords:** Foundation model; Seismic data denoising; Pre-training; Downstream tasks.

**Article Highlights.**

1. We review the existing denoising methods and identify the core limitation of task-specific deep learning methods.
2. We present a dedicated seismic denoising foundation model termed SeisDeFM and show a case study on diverse noise attenuation.
3. SeisDeFM achieves superior denoising with small-scale labeled data, enabling efficient large-scale field seismic data processing.

# 1 Introduction

Seismic data denoising plays a pivotal role in the realm of exploration geophysics. Effective denoising methods can suppress seismic background noise and restore high-fidelity signals that reveal detailed information about subsurface structures and physical parameters. During the acquisition of seismic data, effective signals are simultaneously contaminated by various types of noises that can be roughly divided into random and coherent noises. Coherent noise generated by sources, such as air waves, ground rolls, and multiples, often shows specific propagation characteristics. On the contrary, random noise, primarily caused by environmental interferences, exhibits non-coherency in time or offset.

With the increasing demand for high-precision seismic exploration, the geophysical community has gradually recognized the importance of noise reduction and developed various denoising methods to enhance the quality of seismic data. These seismic denoising methods can be broadly classified into four categories: filtering, decomposition-based, sparse-representation, and low-rank approaches.

## 1.1 Traditional seismic denoising methods

Filtering methods assume that events of interest are spatially predictable and then use the differences in various domains (time, frequency, offset, and wavenumber, etc) to separate signals from noise. Canales (1984) introduced the method of random noise attenuation by frequency-space (f-x) predictive filter, showing good performance in several two-dimensional (2D) real examples. Gulunay (1986) further developed the f-x deconvolution based on complex series prediction work proposed by Treitel (1974). Beresford-Smith and Rango (1989) applied a frequency-wavenumber (f-k) filter to

attenuate the ground rolls, but it is prone to destroy partial signals. Abma and Claerbout (1995) extended the time-space (t-x) and f-x predictive techniques to their three-dimensional (3D) versions. The two 3D extensions enhance noise-attenuation performance by using more samples in the predictions and also relax the strict requirement on linear events. Naghizadeh and Sacchi (2012) developed a vector autoregressive model-based f-x denoising method and utilize it to deal with multicomponent seismic data. Chen and Ma (2014) proposed a combination of the f-x predictive filter and empirical-mode decomposition (EMD) to enhance the preservation of effective signals while attenuating noise.

EMD (Huang et al. 1998) is a classical signal processing technique that can decompose input data into a series of intrinsic modes. This technique specializes in analyzing non-linear and non-stationary signals, which are inherent in seismic data (Chen and Ma 2014). Bekara and Van der Baan (2009) proposed a data-adaptive EMD-based method for the reduction of random and coherent noise. Han and van der Baan (2015) applied ensemble EMD to decompose noisy microseismic data and subsequently used adaptive thresholding to remove incoherent components from the resulting intrinsic mode functions. Experimental results showed that this method is effective for denoising both microseismic and reflection seismic records. To accelerate EMD-based denoising implementations, Gomez and Velis (2016) streamlined the computationally expensive polynomial interpolation by introducing a window-averaged sifting operation. Compared to conventional EMD, this improved version reduced computational complexity and required fewer user-defined parameters. Beyond EMD, other decomposition-based methods, such as variational mode decomposition (VMD) and singular value decomposition (SVD), have also been applied to background noise attenuation in seismic data. Bekara and Van der Baan (2007) applied local SVD to enhance the signal-to-noise ratio (SNR) of seismic data. Although this method outperformed the f-x deconvolution and median filter for noise suppression, it was less effective in preserving weak and complex events. To overcome the potential energy attenuation of events and low-resolution decomposition of signals when using EMD, Yu and Ma (2017) developed a novel VMD-based denoising method for random noise. By applying VMD on constant-frequency slices in f-x domain, their approach effectively suppressed random noise while showing improved performance in preserving steep structural slopes.

Sparse-representation approaches assume that seismic data admit a sparse representation in a suitable basis. These methods transform the data into sparsity-promoting domains (e.g., wavelet, Radon, curvelet, contourlet, and shearlet), then remove noise by thresholding the coefficients and applying an inverse transform. Gao et al. (2006) applied the wavelet transform to denoise the random noise in pre-stack seismic data. Compared with wavelet transform, curvelet transform offers superior directional sensitivity and can capture more geometric features. Hennenfent and Herrmann (2006) extended the fast discrete curvelet transform to non-uniformly sampled data and deployed denoising and interpolation tasks to discuss the sparseness of this extended version on seismic data. Subsequently, Liang et al. (2018), Zhang and Baan (2019), and Dong et al. (2019a) developed several denoising methods based on multi-scale and multi-direction shearlet transform. Dictionary learning adaptively learns basis functions from the data itself, rather than using fixed bases in mathematical

transforms, so it can provide a sparser representation for seismic data denoising (Dong et al. 2022). For example, Beckouche and Ma (2014) partition seismic data into patches and adaptively learn a sparse dictionary, showing superiority over wavelet and curvelet transforms in SNR enhancement and weak signal preservation.

## 1.2 DL-based seismic denoising methods

Most conventional denoising methods rely heavily on manually designed prior assumptions. When these assumptions are violated in real-world applications, these methods often suffer from significant performance degradation (Dong et al. 2019a; Yu et al. 2019). Moreover, tedious empirical parameter tuning further hinders their practical deployment (Yu et al. 2019; Zhong et al. 2024). Deep learning (DL) methods have shown great promise in the field of artificial general intelligence. Through end-to-end training, they enable the automatic optimization of a massive number of trainable parameters, thereby learning high-dimensional and nonlinear complex mappings (Dong et al. 2020; Sheng et al. 2025). Therefore, since 2018, researchers in the relevant fields have developed various DL-based denoising methods for seismic data. In this section, we review these methods according to the types of noise they target.

### 1.2.1 Surface waves

Surface waves are a typical type of coherent noise characterized by low velocity and strong energy. Moreover, they often seriously overlap with reflections in both the t-x and f-x domains, which frequently leads to the unintended attenuation of signals during the suppression process. Dong et al. (2019b) and Yu et al. (2019) were among the first to employ convolutional neural networks (CNNs) for surface wave attenuation, achieving denoising performance superior to that of conventional approaches. Kaur et al. (2020) generated training labels using a local time-frequency transform and regularized non-stationary regression, and then fed them into a generative adversarial network (GAN). This approach yields results comparable to those of the two conventional label-generation methods, while automating surface wave attenuation and substantially reducing computational cost.Pham and Li (2022) further proposed a physics-constrained DL framework for surface-wave attenuation that integrates unsupervised decomposition, a supervised f-k domain classifier, and signal-to-noise mapping, thereby effectively suppressing noise while preserving weak reflections. Yang et al. (2023c) proposed a fully convolutional framework that leverages dense and skip connections to improve information propagation, thereby better preserving low-frequency signals overlapping with surface waves. Li et al. (2024b) proposed a conditional diffusion model that jointly generates clean data and surface waves, so as to minimize reflection leakage and outperforming two conventional methods.

### 1.2.2 Random noise

Random noise, another major and ubiquitous type, is a spatio-temporal stochastic process arising from natural or anthropogenic sources (Li et al. 2017). The attenuation of random noise remains a challenging task due to its complex characteristics, including

the absence of fixed frequencies and apparent velocities, non-stationarity, and non-Gaussianity (Zhong et al. 2015).

In the early stages, experts commonly employed CNN-based architectures and trained them in a supervised fashion. For example, Wang et al. (2019) applied data augmentation to generate paired datasets for network training, achieving better denoising performance than conventional methods. The U-Net, proposed by Ronneberger et al. (2015), is a widely used encoder-decoder CNN that employs skip connections to preserve fine-grained spatial information. Zhong et al. (2022) later introduced an enhanced version that incorporates residual learning and reconstruction blocks, which outperformed the standard U-Net in both noise attenuation and signal preservation. Afterwards, Zhao et al. (2023) proposed another improved U-Net by incorporating deformable and dilated convolutions, which outperformed two classic DL-based methods in terms of quantitative analyses and visual quality.

Despite their promising performance, these methods are grounded in supervised learning paradigms that demand a substantial number of high-quality labeled datasets, thereby imposing significant hurdles in dataset sourcing and preparation (Sun et al. 2022; Zhang and Wang 2023). A number of unsupervised or self-supervised learning approaches have recently been proposed to circumvent this limitation. Birnie et al. (2021) introduced a self-supervised blind-spot network that circumvents the need for noisy-clean data pairs. They generated corrupted samples by randomly masking a set of non-adjacent pixels and used the original data as training targets. By recovering these masked pixels from their neighboring ones, this network learns directly from the noisy data without labels, effectively suppressing random noise while minimally distorting the signals. Sun et al. (2022) achieved random noise suppression via a patch-wise self-supervised approach that integrates invariant function and transfer learning, requiring only a single noisy shot gather. Zhang and Wang (2023) proposed an unsupervised recursive deep image prior (RDIP) approach that incorporates an improved quality control criterion for random noise attenuation. The RDIP iteratively retrains the network by using the previous noisy input and its corresponding denoised output as a new input-target pair. This recursive self-training strategy progressively refines the result until convergence, yielding higher denoising accuracy and reduced signal leakage.

These aforementioned methods are built upon CNN architectures and are therefore inherently constrained by the local receptive fields of convolutional kernels. To overcome this limitation, more advanced frameworks have been adopted for seismic random noise attenuation. Transformer (Vaswani et al. 2017), with self-attention mechanism, can capture more global features. Therefore, Li et al. (2024a) adopted Swin-Transformer (Liu et al. 2021) to suppress random noise and generated higher SNR than competing methods. Zhang et al. (2024) combined Swin-Transformer with GAN to simultaneously interpolate and denoise seismic data. However, Transformer-based architectures often suffer from quadratic computational complexity, incurring substantial time and computational costs during both training and testing. Recently, a novel architecture named Mamba (Gu and Dao 2024) can model global dependencies with linear complexity, significantly accelerating the training process. Chen et al. (2025) combined the Mamba architecture with fast Fourier convolutions to

suppress seismic random noise in both time and frequency domains. This approach achieved superior denoising performance compared to Transformer-based methods, while incurring substantially lower computational cost.

### 1.2.3 Erratic noise

Erratic noise refers to a highly non-Gaussian component of seismic background noise, commonly originating from natural sources (e.g., air blasts, wind, rain, and poor surface conditions), cultural interferences (e.g., power-line noise), acquisition glitches (e.g., recording or parity errors), and quality-control artifacts (Trickett et al. 2012). It usually appears as localized high-amplitude anomalies in seismic records, with its noise level varying from traces, time windows and shot gathers (Wang et al. 2022c).

In recent years, geophysical scholars have made considerable efforts to develop DL-based methods for erratic noise suppression. Wang et al. (2022c) attenuated erratic noise by introducing an attention-based CNN together with a shuffled noise training data generation strategy, allowing the network to better focus on noisy regions while eliminating the need for clean training data. A joint-guided denoising network was proposed by Zhong et al. (2024) to attenuate erratic noise by enriching feature interactions in CNNs. This network shows good denoising performance under extremely low SNR conditions.

To increase the flexibility of DL-based erratic noise attenuation without the need for paired noisy and clean training data, Qian et al. (2022, 2024) turned to unsupervised learning strategies. In their earlier work (Qian et al. 2022), they constructed a robust unsupervised deep convolutional auto-encoder capable of removing both erratic and Gaussian noise. Subsequently, they extended their approach to 3D version by proposing a robust tensor deep learning framework that integrates tensor sparse representation with a tensor neural network, thereby effectively exploiting the 3D spatial correlations of seismic data. However, most of unsupervised denoising methods are prone to converging to sub-optimal solution and often suffer from high computational cost. To address this issue, Bao et al. (2025) further designed a lightweight unsupervised framework based on implicit neural representation. This framework incorporated positional encoding and specialized activation functions to accelerate the procedure of network optimization.

Most of the aforementioned erratic noise attenuation approaches are purely data-driven, with no explicit incorporation of physical constraints. Such a data-centric paradigm inevitably compromises signal fidelity, often leading to waveform distortion and degraded robustness. As a result, whether the reconstructed signals truly honor the underlying physics remains an open question. To alleviate this, Xie et al. (2025) designed a physics-informed neural network to simultaneously remove random and erratic noises. The effective recovery of high-frequency signals on both synthetic and field data highlights the contribution of physical information.

### 1.2.4 Multiples

In field seismic data, the recorded events include both primaries and multiples. Primaries are seismic waves that have undergone exactly one reflection from subsurface interfaces, whereas multiples result from two or more reflections before reaching the

receiver (Verschuur 2013; Durall et al. 2024). Multiples do not correspond to true subsurface interfaces and are generally eliminated prior imaging and interpretation (Li et al. 2021; Tao et al. 2022; Wang et al. 2022a).

Depending on their propagation paths, multiples are broadly classified into surface-related and internal multiples (Weglein 1999). Most DL-based demultiple methods have focused on the former. Among these, U-Net-based approaches constitute a major category. Li et al. (2021) were among the first to apply U-Net to surface-related multiple removal, achieving a substantial improvement in SNR over conventional linear regression methods. Durall et al. (2024) employed a U-Net to learn the move-out discrimination between primaries and multiples, rather than directly predicting the multiples themselves. Tests on synthetic and field datasets demonstrate that this method effectively suppresses multiples while preserving the amplitude of primaries. The physics-guided DL framework (PGDL) proposed by Zhang and Verschuur (2026) is another U-Net-based method, which integrates the original recorded full wavefield and wave-equation-based multiple predictions as a dual-channel input. PGDL overcomes the limitation that most neural networks typically treat seismic records merely as images, disregarding the physical propagation of seismic waves.

To reduce the dependence on labeled true primaries or multiples, researchers have explored various unsupervised and self-supervised methods. Wang et al. (2022b) proposed an unsupervised deep neural network based on ensemble learning (UDNNEL), which integrates a residual network, a simple U-Net, and an attention-based network. The UDNNEL jointly optimizes the three networks by integrating their six generated outputs, allowing the model to aggregate the advantages of each learner. Li et al. (2023) introduced an unsupervised fast iterative shrinkage thresholding algorithm network (FISTA-Net) for surface-related multiple removal, where the FISTA iterations are unfolded into a neural network and the shrinkage step is replaced by a U-Net. Subsequently, Qi et al. (2025) enhanced FISTA-Net by adopting a semi-supervised strategy that leverages both limited labeled and pseudo-labeled data, achieving markedly better performance than the original unsupervised version.

Several studies have also addressed the more challenging internal multiples. For example, Liu et al. (2022) and Wang et al. (2023) proposed U-Net-based supervised and unsupervised methods for internal multiple suppression, respectively. To further enhance the suppression performance, Zhang et al. (2025) incorporated U-Net with Transformer to fuse both local and global features, yielding better performance and efficiency than existing conventional and CNN-based methods.

Despite the success of these demultiple methods, some researchers have adopted a different perspective, treating multiples as useful signals and exploiting them together with primaries for joint imaging, thus obviating the need for attenuation (Verschuur et al. 2016; Lu 2021). You et al. (2025) developed Deepimaging, a DL-based multiple imaging framework. This framework leverages stochastic velocity models as the training foundation, introduces a novel DL-driven approach to represent one-way wave propagation operators, and ultimately establishes an intelligent full-wavefield imaging system.

### 1.2.5 DAS denoising

Over the past decade, distributed optical-fiber acoustic sensing (DAS) has been proven as a highly promising technology for seismic data acquisition. In contrast to conventional electronic geophones, DAS leverages optical fibers as sensing elements to measure strain induced by seismic waves, offering both high spatial and temporal resolution (Nesterova et al. 2026; Zhang et al. 2026). In seismic exploration, DAS has gained acceptance as an alternative to conventional downhole geophones, owing to its exceptional resilience to extreme temperature and pressure. This enables the acquisition of high-fidelity and densely-sampled VSP data with full-well coverage (Lellouch et al. 2020; Yang et al. 2023b).

In DAS-VSP data, in addition to conventional seismic background noise, signals are also contaminated by multiple instrument-related noises that are inherent to fiber-optic sensing systems. These instrument-related noises, including fading, horizontal, and checkerboard noises (Li et al. 2022), are absent from other seismic datasets, and their underlying mechanisms remain poorly understood. At that time, no effective suppression techniques were available, creating a critical bottleneck that severely hindered the widespread deployment of DAS acquisition technology.

One of the earliest DL-based DAS-VSP denoising studies was by Dong et al. (2020), who proposed an end-to-end CNN with a novel mean square error loss that incorporates an energy ratio matrix. While this method effectively suppresses both random and coherent noises, signal leakage observed in the residual records indicates that signal preservation still requires further improvement. To address this issue, Dong and Li (2020) later developed a convolutional adversarial denoising network (CADN) based on game theory. In CADN, a generator and a discriminator engage in an adversarial process that drives the reconstructed wavefield's probability distribution to asymptotically approach the target, thereby achieving more accurate signal recovery.

The attention mechanism, a widely-used tool in neural networks, dynamically assigns weights to feature maps to focus on the most relevant information during feature extraction (Bahdanau et al. 2014). For DAS-VSP denoising, this mechanism is particularly beneficial, as it can prioritize informative signals over various noise patterns, proving especially effective for recovering weak signals. Li et al. (2022) proposed an attention-based progressive multi-stage denoising network for DAS-VSP data. These cascaded stages progressively refine the denoised results, while the channel and stage-interactive attention mechanisms enhance the feature representation of this network. To handle complex DAS noise, Yang et al. (2023b) integrated dense connections and a kernel-wise attention mechanism into a multi-scale network. Ablation studies validated that the attention module significantly enhances multi-scale feature fusion and consequently boosts denoising performance.

Beyond attention mechanisms, several studies have also turned to unsupervised or self-supervised frameworks that impose fewer constraints on labeled data, thus offering greater flexibility in data preparation. Yang et al. (2023a) pioneered an unsupervised denoising network with a U-Net-like architecture, which outperforms both conventional and DL-based counterparts. Subsequently, Saad et al. (2024) developed an unsupervised DL model for DAS-VSP denoising, using band-pass filtered data

and continuous wavelet transform (CWT) coefficients as auxiliary guidance. Building on this strategy, Saad and Alkhalifah (2026) further introduced a self-supervised conditional diffusion model, in which the coarsest-scale CWT serves as conditioning information to steer the reverse diffusion process. Notably, the two approaches (Saad et al. 2024; Saad and Alkhalifah 2026) exploit CWT coefficients as an effective prior for signal preservation, and their practical feasibility has been demonstrated through extensive evaluations on field datasets.

The above studies have mostly targeted single-task denoising improvement, but recent efforts have begun exploring integrated frameworks that address multiple DAS-VSP processing tasks jointly. Notably, Cheng et al. (2025a) proposed a multi-task preprocessing model (MTPM) that unifies denoising, interpolation, and wavefield separation into a single CNN-Transformer network with a two-stage training regimen. Unlike conventional step-by-step pipelines, this approach performs the three tasks concurrently via one-step prediction, effectively preventing cumulative signal attenuation and yielding enhanced overall processing performance.

## 1.3 Foundation model for Pre-stack Seismic Data Denoising

Although existing DL-based seismic denoising methods perform well on specific noise types, most are designed for a single noise scenario and fail to generalize to mixed-noise environments. In field seismic exploration, signals are often simultaneously contaminated by multiple types of noises (Hlebnikov et al. 2021; Cheng et al. 2024). Moreover, even the same type of noise can exhibit significant distribution discrepancies across different survey areas, making it difficult for task-specific models to be directly applied to new areas. Therefore, it is imperative to develop a seismic denoising foundation model (FM) capable of handling diverse noise conditions across different survey areas.

In recent years, FMs have emerged as an important research direction in the field of artificial intelligence. Unlike task-specific DL models, FMs are usually pre-trained on large-scale datasets to learn more general data representations, and then fine-tuned with a small amount of labeled data or task-specific data for different downstream tasks (Awais et al. 2025). Unlike task-specific models, FMs offer superior feature extraction and stronger generalization. In addition, FMs generally have the capability of modeling long-range dependencies (Huang et al. 2023), allowing them to capture global features of different modalities.

Recent advances in FMs have been largely driven by the rapid development of pre-training techniques in natural language processing (NLP) and computer vision (CV). Devlin et al. (2019) proposed a language representation model, bidirectional encoder representations from Transformers (BERT), which learns contextualized representations by pre-training on large-scale text data and achieves excellent performance across various downstream tasks. Subsequently, Brown et al. (2020) proposed a large-scale language model pre-trained on massive text data, namely GPT-3, which demonstrated excellent language understanding capabilities. With only a few examples, GPT-3 can adapt to different NLP tasks. Inspired by the success of Transformer-based architectures in NLP, Dosovitskiy et al. (2020) proposed Vision Transformer (ViT), which extended the Transformer architecture to CV tasks. Caron et al. (2021) further introduced the self-supervised DINO framework, enabling visual representation learning

from unlabeled images and demonstrating strong generalization capability in object recognition tasks. Kirillov et al. (2023) proposed the segment anything model (SAM), which was trained on large-scale image datasets with segmentation annotations and achieved zero-shot segmentation across diverse objects.

However, due to the scarcity of large-scale labeled seismic data, the prohibitive computational cost of pre-training, and the inherent complexity of seismic wavefields, research on foundation models (FMs) for seismic exploration remains in its infancy. Only a limited number of studies have explored the potential of FMs for seismic data processing and interpretation tasks. Harsuko and Alkhalifah (2022) proposed the StorSeismic model, which is based on the BERT (Devlin et al. 2019). The model was pre-trained on pre-stack seismic data and subsequently adapted to downstream tasks by using fine-tuning. Sheng et al. (2025) developed seismic FM based on the masked autoencoder architecture (He et al. 2022) and transferred the pre-trained model to various seismic downstream tasks. Dong et al. (2026) proposed a lightweight seismic processing FM (SPFM) based on the Mamba model (Gu and Dao 2024). SPFM can effectively capture global features of seismic data while alleviating the quadratic computational complexity inherent in Transformer-based models. Cheng et al. (2025b) introduced a generative seismic FM (GSFM) based on diffusion models. Through probabilistic modeling, GSFM effectively addresses multi-task seismic processing problems and achieves uncertainty quantification.

In this paper, we propose for the first time a new seismic denoising paradigm based on the seismic denoising foundation model (SeisDeFM). This paradigm aims to train the denoising model with massive clean seismic data to learn the intrinsic representation of effective seismic signals, and enable rapid transfer and adaptation to downstream denoising tasks targeting diverse noise types using only a small amount of paired data, thereby meeting the demand for efficient, high-precision processing of industrial-scale massive seismic data. Furthermore, through a case study of SeisDeFM, we comprehensively present and demonstrate its viable architecture, pre-training strategy, downstream task adaptation methods, and dataset construction workflow. The experimental results in the case study verify that the SeisDeFM outperforms task-specific denoising models (TSDM) in training cost, denoising performance, and generalization capability, offering a new paradigm worthy of industry reference and promotion for industrial-scale seismic data denoising.

# 2 Principles of SeisDeFM

In this section, we first introduce the basic theory of SeisDeFM. Subsequently, we elaborate on the framework of SPFM, including training data preparation, viable model architecture, pretraining strategies, downstream adaptation methods, and evaluation metrics for denoising performance.

## 2.1 The basic theory of SeisDeFM

The observed pre-stack seismic data $y \in \mathbb{R}$ can be expressed as the sum of clean signals $x \in \mathbb{R}$ and noises $n \in \mathbb{R}$:

$$y(h) = x + n(h), \quad \forall h \in \{1, 2, \ldots, H\}, \tag{1}$$

where $h$ represents the type of noise. Each type of noise obeys its unique probability distribution:

$$n(h) \sim P_h(\phi_h), \tag{2}$$

where$P_h$represents the specific probability distribution followed by the $h^{\text{th}}$ type of noises, and $\phi_h$ stands for the set of statistical parameters governing the distribution of noises, including the mean value, variance and higher-order statistics, etc. TSDM is optimized under the empirical risk minimization (ERM) framework, where the optimal model parameters are estimated through neural network training:

$$\hat{\theta} = \arg\min_{\theta} \mathbb{E}_{(y,x)\sim P_{TS}} \left[ \mathcal{L} \left( \mathcal{F}_\theta \left( y(h) \right), x \right) \right], \tag{3}$$

where $\theta$ and $\hat{\theta}$ represents the trainable and optimal parameters of TSDM, respectively, $P_{TS}$ represents the data distribution of the observed pre-stack seismic data and clean signals used for training TSDM, $\mathcal{L}$ represents the loss function, and $\mathcal{F}_\theta(\cdot)$ represents the nonlinear mapping of TSDM. Sufficient quantity, feature-complete and labeled training datasets are required to support the training of TSDM for searching the $\hat{\theta}$:

$$\mathcal{D}_{TS} = \{(y_i(h), x_i)\}_{i=1}^{N}, \tag{4}$$

where $\mathcal{D}_{TS}$ represents the dataset used for training TSDM, and $N$ denotes the total number of paired training samples. The intrinsic limitation of TSDM lies in its implicit assumption that $\mathcal{D}_{TS}$ and the test data for field applications are independent and identically-distributed. Given the massive volume of modern seismic data, the diversity of clean data, the complexity of noise, and the scarcity of paired noisy-clean seismic data, the distributions of field data may deviate from pre-defined training datasets, which causes TSDM to underfit field data and to exhibit severe degradation of generalization performance across different surveys and targets.

By employing massive parameterization, large-scale pre-training and small-scale transfer learning, SeisDeFM can break through the capacity bottlenecks of TSDM. It learns physically consistent wavefield representations and rapidly adapts to diverse noise, thereby enabling robust generalization across highly complex and unseen field conditions. Accordingly, SeisDeFM requires abundant training data to learn rich features from diverse field data. The distribution of datasets for training SeisDeFM is denoted as $P_{FM}$, which can be expressed as:

$$P_{FM} \equiv \sum_{j=1}^{J} \alpha_j P_j, \quad \text{with} \quad \sum_{j=1}^{J} \alpha_j \equiv 1, \tag{5}$$

where $P_j$ represents the data distribution of the $j^{\text{th}}$ type of dataset and $\alpha_j$ is the weighting coefficient. Besides datasets that contain various types of noise, SeisDeFM requires data categories covering diverse features of clean signals, which yields $J > H$. Subsequently, SeisDeFM is trained by solving the following optimization problem:

$$\hat{\Theta} = \arg\min_{\Theta} \mathbb{E}_{(y,x)\sim P_{FM}} \left[ \mathcal{L}\left( \mathcal{F}_{\Theta}(y), x \right) \right], \tag{6}$$

where$\Theta$ and$\hat{\Theta}$ represents the trainable and optimal parameters of SeisDeFM, respectively, and $\mathcal{F}_{\Theta}(\cdot)$ represents the the nonlinear mapping of SeisDeFM. Compared with Equation (3), Equation (6) enables SeisDeFM to capture seismic representations that encode local correlations and global dependencies across different data distributions, providing stronger generalization and transfer ability.

The construction and application of SeisDeFM generally follow a two-stage paradigm of pre-training and downstream task adaptation (Liu et al. 2025) . This paradigm aims to realize smooth transfer from universal feature representations to task-specific seismic denoising for seismic exploration. To build robust wavefield representations, SeisDeFM is first pre-trained on massive clean seismic data. Such data are either generated by wave-equation-based forward modeling, or obtained from field data, including nearly clean raw data and denoised data processed by conventional methods. In the pre-training stage, SeisDeFM is optimized with a self-supervised loss function:

$$\mathcal{L}_{pre} = \mathbb{E}_{x\sim P_{FM}} \left[ \left\| \mathcal{F}_{\Theta}\big(S(x)\big) - x \right\|_2^2 \right], \tag{7}$$

where $\mathcal{L}_{pre}$ represents the loss function for SeisDeFM pre-training, $\|\cdot\|_2$ represents the L2 norm, and $S(\cdot)$ denotes the operation to encourage SeisDeFM to learn robust wavefield representations, such as masked reconstruction and contrastive learning, etc. During pre-training, SeisDeFM captures seismic features such as event coherency, reflection patterns, and time-frequency correlations. These features are independent of specific noise types. By encoding physical priors of clean wavefields through self-supervised pre-training, SeisDeFM bypasses the need to learn seismic characteristics from scratch, requiring only small-scale paired noisy-clean data to learn specific noise distributions for rapid downstream adaptation. The paired dataset adopted to adapt SeisDeFM to specific downstream denoising tasks for transfer learning can be denoted as:

$$\mathcal{D}_{TS} = \left\{ \big(y_i(h), x_i\big) \right\}_{i=1}^{M}, \quad M \ll N, \tag{8}$$

where $\mathcal{D}_{TS} \sim P_{FM}$. The transfer learning of SeisDeFM is optimized by a supervised loss function:

$$\mathcal{L}_{tl} \equiv \frac{1}{H} \sum_{h=1}^{H} \left\{ \mathbb{E}_{(y,x)\sim P_{TS}} \left[ \left\| \mathcal{F}_{\Theta}\big(y(h)\big) - x \right\|_2^2 + \lambda \mathcal{L}_{reg} \right] \right\}, \tag{9}$$

where $\mathcal{L}_{reg}$ and $\lambda$ represent the regularization term adapted to specific signal-to-noise characteristics and its corresponding weighting coefficient, respectively. The pre-trained parameters of SeisDeFM serve as a strong initialization, allowing SeisDeFM to rapidly adapt to specific noise distributions with limited labeled data. Through pre-training and transfer learning, SeisDeFM approximates a universal conditional distribution $P_{\text{denoising}}$ rather than learns a mapping specific to a single noise type:

$$P_{\text{denoising}}\big(x \sim P_{FM} \,|\, y \sim P_{FM}\big). \tag{10}$$

Consequently, SeisDeFM exhibits multi-noise adaptability, cross-survey generalization, and efficient transfer learning on unseen datasets. This paradigm shifts seismic denoising from the isolated learning of noise-specific mappings toward universal seismic representations and efficient noise adaptation, forming the theoretical basis for large-scale pre-stack seismic data denoising.

The main workflow of SeisDeFM is shown in Fig. 1. In this paper, we present a case study for SeisDeFM that details the preparation of training datasets as well as the strategies for pre-training and downstream adaptation. Furthermore, we demonstrate SeisDeFM's performance by evaluating its denoising effectiveness against four specific types of noise, including erratic noise, external noise, surface waves, and random noise.

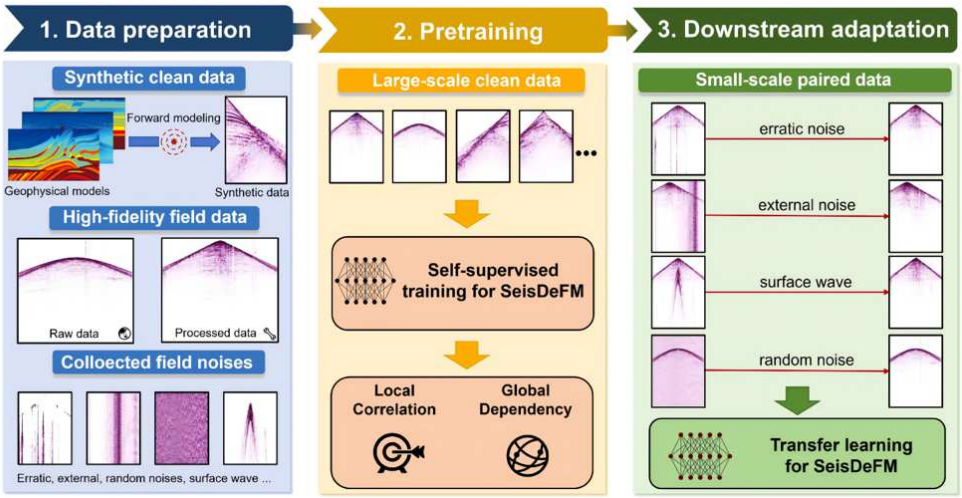



**Fig. 1** The workflow of SeisDeFM.

## 2.2 Data preparation for training SeisDeFM

Building SeisDeFM with strong generalization capabilities depends fundamentally on constructing high-quality, large-scale, representative datasets. In this section, we detail the data preparation strategies for the self-supervised pre-training and downstream adaptation stages.

The objective of the pre-training stage is to enable SeisDeFM to comprehensively learn the universal geophysical laws of seismic wavefields and the spatial correlations of seismic events. Therefore, sufficient clean seismic data are required for pre-training. Synthetic clean data can be obtained via wave-equation-based forward modeling, whereas clean field data are derived from the nearly clean raw data or the data processed with conventional processing approaches. The detailed methods for constructing clean seismic datasets are as follows:

(1) Forward modeling based on wave equations. Forward modeling involves establishing geophysical models for the subsurface medium (velocity, density, Q factor and

Thomsen parameters etc.), setting an appropriate observation system (towed marine streamer, ocean-bottom cable (OBC) and vertical seismic profile (VSP) etc.), and performing numerical simulation with corresponding wave equations (acoustic, elastic, viscoelastic and anisotropic etc.) to produce clean seismic records. Taking the isotropic acoustic wave equation as an example, we implement the forward modeling process as follows:

$$\frac{1}{v^2(\boldsymbol{r})}\frac{\partial^2 u(\boldsymbol{r},t)}{\partial t^2} = \nabla^2 u(\boldsymbol{r},t) + f(\boldsymbol{r},t). \tag{11}$$

where u, v and f represent the wavefield, velocity model and source function, respectively, r denotes the coordinate vector in 3-dimensional (3D) space. The primary advantage of this method lies in its ability to acquire absolutely clean data with known ground-truth labels. However, numerical simulations are often overly idealized and struggle to perfectly replicate the complex medium characteristics and true wavefield responses inherent in field data. Consequently, models pre-trained solely on synthetic data often suffer from generalization limitations when applied to field data.

(2) Field data with high-fidelity. This approach directly utilizes massive volumes of field seismic data for training. For seismic acquisition carried out under favorable environmental conditions with simple subsurface structures and no complex near-surface interferences, the raw shot gathers or data in partial spatio-temporal regions exhibit relatively high SNR and can be adopted as clean data. However, the raw clean data requires stringent conditions, and such data are therefore scarce in quantity. For raw data containing a small amount of noise, the data denoised by conventional methods can be used as the clean field data for pre-training. The clean field data preserve to the greatest extent the complex kinematic and dynamic characteristics of field seismic wavefields, ensuring that the features extracted by SeisDeFM adhere strictly to geophysical laws. However, refining raw field data into high-fidelity clean data demands substantial manual effort from domain experts as well as rigorous validations. This procedure renders such high-quality field datasets exceedingly scarce and costly to acquire.

During the downstream adaptation stage for specific denoising tasks, it is necessary to construct paired noisy-clean datasets for supervised transfer learning. To maximize SeisDeFM's adaptability to complex field conditions, target noises must be added into the clean data. The mainstream practice involves extracting real noise directly from field seismic data. To enhance the robustness of the downstream tasks against highly variable noise distributions, an augmentation strategy for different kinds of noises by generative models can be adopted (Dong et al. 2022).

In this paper, to achieve superior denoising performance, we adopt the clean field data as the pre-training dataset for SeisDeFM to mitigate domain-shift issues arising from synthetic clean data. Meanwhile, real noise is added to clean field data to build paired datasets for transfer learning in downstream tasks. As illustrated in Fig. 2, the clean data contain abundant local event features, and the real noise covers a highly diverse set of complex field interference patterns, guaranteeing sufficient diversity in both clean-signal and noise distributions. During training, using a complete shot gather as a single input channel greatly helps SeisDeFM capture global wavefield dependencies, whereas cropping a shot gather into smaller patches mainly serves to

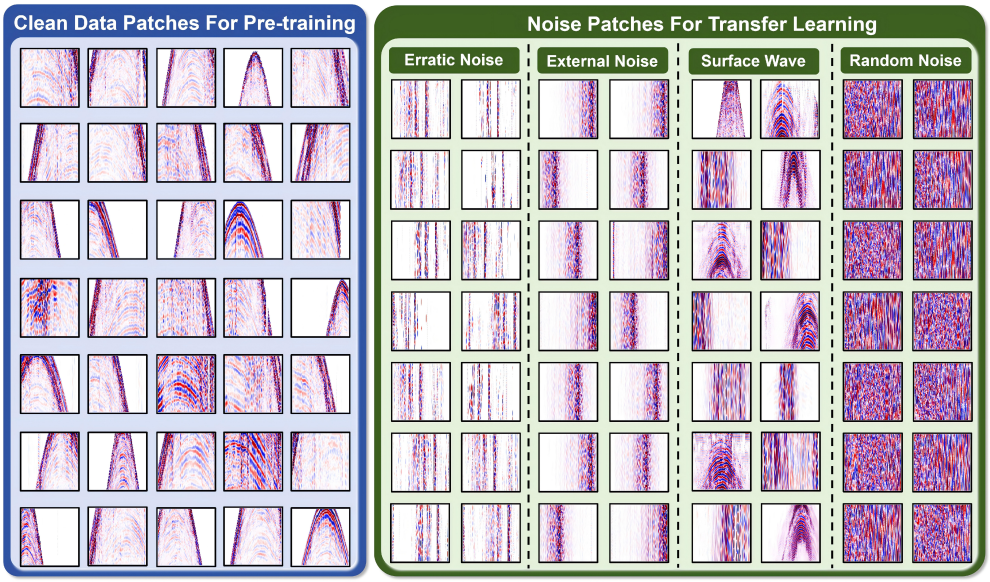



**Fig. 2** Features of clean data patches for SeisDeFM pre-training and features of noise patches for SeisDeFM transfer learning.

reduce computational costs. To obtain better denoising performance, the case study presented in this paper uses the strategy of taking the complete shot gather as a single-channel input.

## 2.3 The architecture of SeisDeFM

In recent years, foundation-model architectures including Transformer, Mamba, and diffusion models have exhibited revolutionary performance in diverse fields, benefiting from their powerful global context modeling and capacity for generating complex data distributions (Kumar 2024; Cheng et al. 2025b; Dong et al. 2026). Nevertheless, naive deployment of these generic architectures on seismic data cannot fully characterize the unique spatio-temporal relationships and strong inter-trace coherence embedded in seismic wavefields. To tackle this limitation, we present a SeisDeFM architecture tailored for seismic denoising in our case study (Fig. 3).

The network adopts a residual U-Net backbone to extract local multi-scale features, alongside a bottleneck equipped with a trace-wise Transformer for global information fusion. Such a design trades off the U-Net's strength in retaining local seismic events and the Transformer's superiority in capturing long-range trace dependencies, thereby optimizing SeisDeFM for space-time-domain signal-noise separation. The internal data flow of SeisDeFM consists of four key stages:

(1) Local multi-scale feature encoding. The input seismic data for SeisDeFM can be denoted as $Y \in \mathbb{R}^{B\times C\times T\times X}$, where B, C, T and X represent the batch size, number of channels, number of time samples, and number of traces, respectively. The U-Net encoder consists of five successive residual convolutional blocks and max-pooling layers. Each residual block includes $3 \times 3$ convolutions, group normalization,

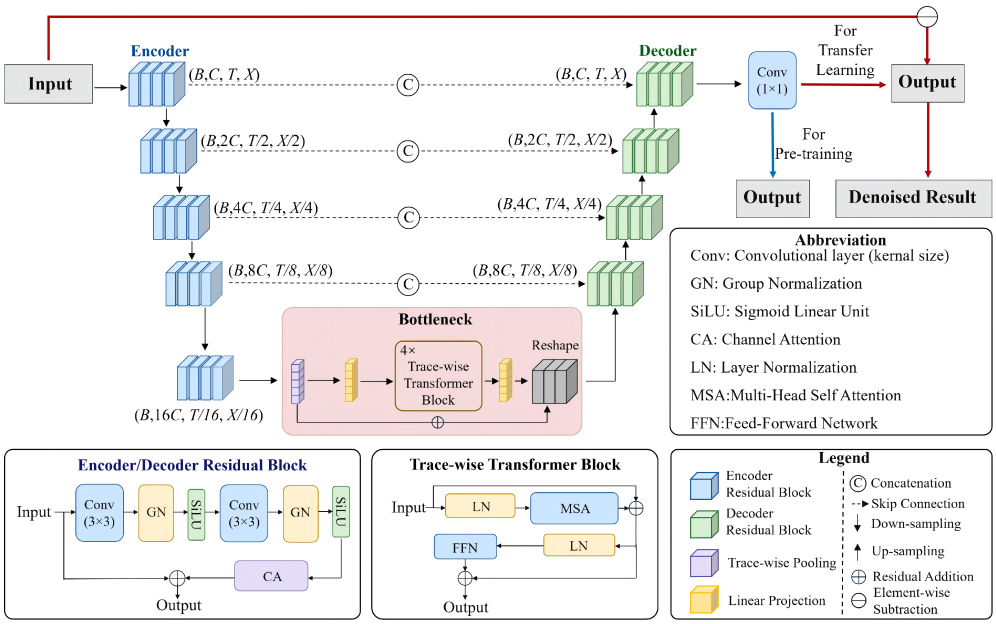



**Fig. 3** The detailed architecture of SeisDeFM.

Sigmoid-Linear Unit (SiLU) activation functions, and a channel attention mechanism. Skip connections are adopted within each block to mitigate the vanishing-gradient problem in deep networks. By progressively reducing spatial resolution and enlarging the receptive field, the encoder extracts multi-scale local features of seismic waveforms.

(2) Trace-wise global feature extraction at the bottleneck. To mitigate the limited receptive field inherent to the convolution-based U-Net, SeisDeFM integrates a bottleneck module equipped with Transformer blocks to extract global features. The bottleneck's input feature maps undergo temporal pooling and are split into sequence tokens along the seismic-trace dimension. Inside the trace-wise Transformer blocks, each token generates query (Q), key (K), and value (V) through linear projections. Multi-head self-attention (MSA) is subsequently applied to model dependencies among distinct seismic traces:

$$\mathrm{MSA}(Q, K, V) \equiv \mathrm{softmax}\left(\frac{QK^{\mathrm{TP}}}{\sqrt{d_k}}\right) V, \tag{12}$$

where $d_k$ denotes the dimension of the keys, and TP represents the transposition operation. This mechanism allows the network to adaptively exploit global context from distant seismic traces for learning both noise patterns and signal features. The resulting global information is then fused with local CNN-derived features through residual connections, yielding robust seismic representations.

(3) The architecture of the decoder. The decoder gradually recovers the spatio-temporal resolution of bottleneck feature maps using transposed convolutions. In every upsampling step, skip connections concatenate encoder-derived shallow features of the

matching scale with the current high-dimensional features along the channel dimension. Such a design offsets the irreversible spatial-detail loss caused by downsampling and enables effective reconstruction of input data within the space-time domain.

(4) The output data of SeisDeFM. The feature maps produced by the decoder are mapped to a single-channel feature map via a $1 \times 1$ convolutional layer. During the self-supervised pre-training stage, the training objective of SeisDeFM typically involves full-wavefield reconstruction of masked incomplete input data and the maximization of the distance between positive and negative sample pairs for contrastive learning. Accordingly, the output of the $1 \times 1$ convolutional layer can directly serve as the final prediction in the pre-training stage, as indicated by the blue arrow. In the transfer-learning stage, requiring the pre-trained model to directly predict clean seismic signals for downstream denoising tasks frequently leads to convergence difficulties owing to the complex characteristics of seismic wavefields. To tackle this issue, the training objective is switched to predicting the noise component to be removed. A global residual connection is built between the original noisy input and the predicted noise, as indicated by the red arrows. The final denoised seismic data are obtained by subtracting the predicted noise from the original noisy input:

$$\hat{x} = y - \hat{n}, \tag{13}$$

where $\hat{n}$ and $\hat{x}$ represent the predicted noise and denoised result, respectively.

## 2.4 The pre-training of SeisDeFM

At present, self-supervised pre-training schemes for foundation models mainly fall into two categories: generative approaches (including generative models, reconstructive methods, and autoregressive methods) and contrastive approaches (including context-instance and instance-instance contrasts) (Liu et al. 2025). Inspired by these existing paradigms, our case study implements a joint self-supervised learning strategy for SeisDeFM that combines masked wavefield reconstruction with contrastive learning (Fig. 4). This scheme targets the simultaneous capture of local seismic-event continuity and global wavefield correlations. In this pre-training stage, the masked reconstruction and contrastive learning branches receive identical copies from the input batch while sharing all network parameters.

To enable SeisDeFM to learn the principles of seismic wavefields, especially the inter-trace dependencies, we employ a trace-wise masking strategy that better conforms to field seismic scenarios with missing traces or sparse sampling. Specifically, 50% of the traces in the input seismic data are randomly masked to generate the incomplete data. The model is tasked with reconstructing the complete data from this incomplete input. The reconstruction loss function for clean seismic data is defined as the mean squared error (MSE) computed only over the masked traces:

$$\mathcal{L}_{mr} \equiv \frac{1}{L} \sum_{l \in L} \| \mathcal{F}_{\Theta}(x_{\text{mask}})_l - x_l \|_2^2 , \tag{14}$$

where $\mathcal{L}_{mr}$ represents the mask-reconstruction loss function, $x_{\text{mask}}$ denotes the masked seismic data, and $a$ denotes the index set of the masked traces with the maximal number of $A$.

To further enhance the ability of SeisDeFM to represent the global characteristics of clean seismic data, we introduce a contrastive learning mechanism. Given the input clean seismic data $x$, we generate two distinct augmented views $x_{v1}$ and $x_{v2}$ by randomly selecting and combining two out of the following four data augmentation strategies:

(1) All seismic traces are shifted by 5 samples along the time axis, either in the forward or backward direction.

(2) Randomly setting 5% of the seismic traces to zero which is similar to the operation of trace-wise masking.

(3) Scaling amplitudes by multiplying the seismic traces with a random coefficient uniformly sampled between 0.8 1.2.

(4) Adding slight Gaussian noise with a standard deviation:

$$\sigma_G \equiv 0.02 \times \sigma_x, \tag{15}$$

where $\sigma_G$ and $\sigma_x$ represent the standard deviation of the Gaussian noise and the seismic data to be added noise. This process adaptively adds noise based on the statistical characteristics of the corresponding input seismic data, preventing excessive noise from covering effective signals.

The two augmented views are processed by SeisDeFM with shared network parameters. To compute the contrastive loss, a linear projection head is appended to the output of the $1\times$ 1 convolutional layer to flatten and project the 2-dimensional (2D) feature maps into 1-dimensional (1D) latent vectors:

$$z_1 = H\big[\mathcal{F}_\Theta(x_{v1})\big] \text{ and } z_2 = H\big[\mathcal{F}_\Theta(x_{v2})\big], \tag{16}$$

where $z_1$ and $z_2$ represent the 1D latent vectors after linearly projecting, and $G(\bullet)$ represents the operator of linear projection. Specifically, the linear projection head is implemented as a two-layer multi-layer perceptron (MLP), with a rectified linear unit (ReLU) activation sandwiched between the two-layer of MLP.

Contrastive learning employs an information noise contrastive estimation (InfoNCE) loss function. InfoNCE first calculates the pairwise feature similarity matrix between the two augmented views. Then, it treats the two augmented views derived from the same sample as a positive pair, and views from all other samples within the batch as negative samples. Finally, it optimizes the relative similarity between positive and negative samples via a cross-entropy loss formulation:

$$\mathcal{L}_{ct} = -\frac{1}{B}\sum_{p=1}^{B}\log\frac{\exp\big[sim(z_1^p, z_2^p)/\tau\big]}{\sum_{q=1}^{B}\exp\big[sim(z_1^p, z_2^q)/\tau\big]}, \tag{17}$$

where $\mathcal{L}_{ct}$ represents the InfoNCE loss function, p and q denote the indices of the query sample and the comparison sample within the batch, respectively, $z_1^p$and $z_2^p$represent the 1D latent vectors of the two augmented views generated from the $p^{\text{th}}$ seismic data in a batch which composes a positive pair,$z_2^q$represents the 1D latent vector of the augmented view from the $q^{\text{th}}$ sample which is used to compose a negative pair with when p $\neq$ q, $\tau$ is a temperature parameter used to scale the similarity scores and regulate the model's penalty for hard negative samples which is set to 0.1 in this study, and $sim(\cdot)$ denotes the cosine similarity function.

The overall pre-training loss function of SeisDeFM is the weighted sum of the masked reconstruction loss and the contrastive learning loss:

$$\mathcal{L}_{pt} \equiv \mathcal{L}_{mr} + \gamma \mathcal{L}_{ct}, \tag{18}$$

where $\mathcal{L}_{pt}$ represents the complete pre-training loss function, and $\gamma$ is a weighting factor that adjusts the relative importance of the two loss terms. Within this joint optimization framework, the masked reconstruction primarily focuses on capturing the local continuity and inter-trace features of seismic events, and the contrastive learning emphasizes learning the global characteristics and long-range contextual representations of the entire seismic data. Consequently, this pre-training strategy efficiently equips SeisDeFM with a robust, generalizable feature representation that serves as a powerful initialization for downstream denoising tasks.

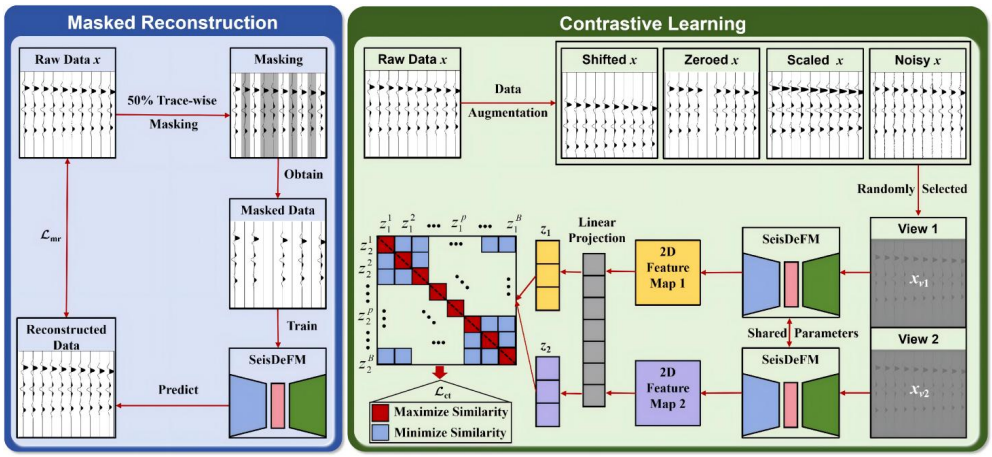



**Fig. 4** The joint pre-training strategy of SeisDeFM.

## 2.5 Transfer learning of SeisDeFM for downstream tasks

Transfer learning of foundation models for downstream tasks fundamentally aims to adapt pre-trained parameters to specific task distributions at minimal computational cost while preserving the intrinsic learned capabilities. Transfer-learning approaches can be broadly categorized into full-parameter fine-tuning, parameter-efficient fine-tuning (PEFT) techniques (including additive and selective fine-tuning, reparameterized fine-tuning, and hybrid fine-tuning), instruction tuning, multi-task joint fine-tuning, unsupervised domain adaptation, and knowledge-distillation-based transfer learning. Considering the strong non-stationarity of seismic wavefields across heterogeneous subsurface media and the complexity of field noise, we present a case of downstream-task adaptation for SeisDeFM, in which a full-parameter fine-tuning strategy is adopted.

During the transfer-learning stage, the network parameters obtained from pre-training serve as a robust initialization for SeisDeFM. To adapt the model to the

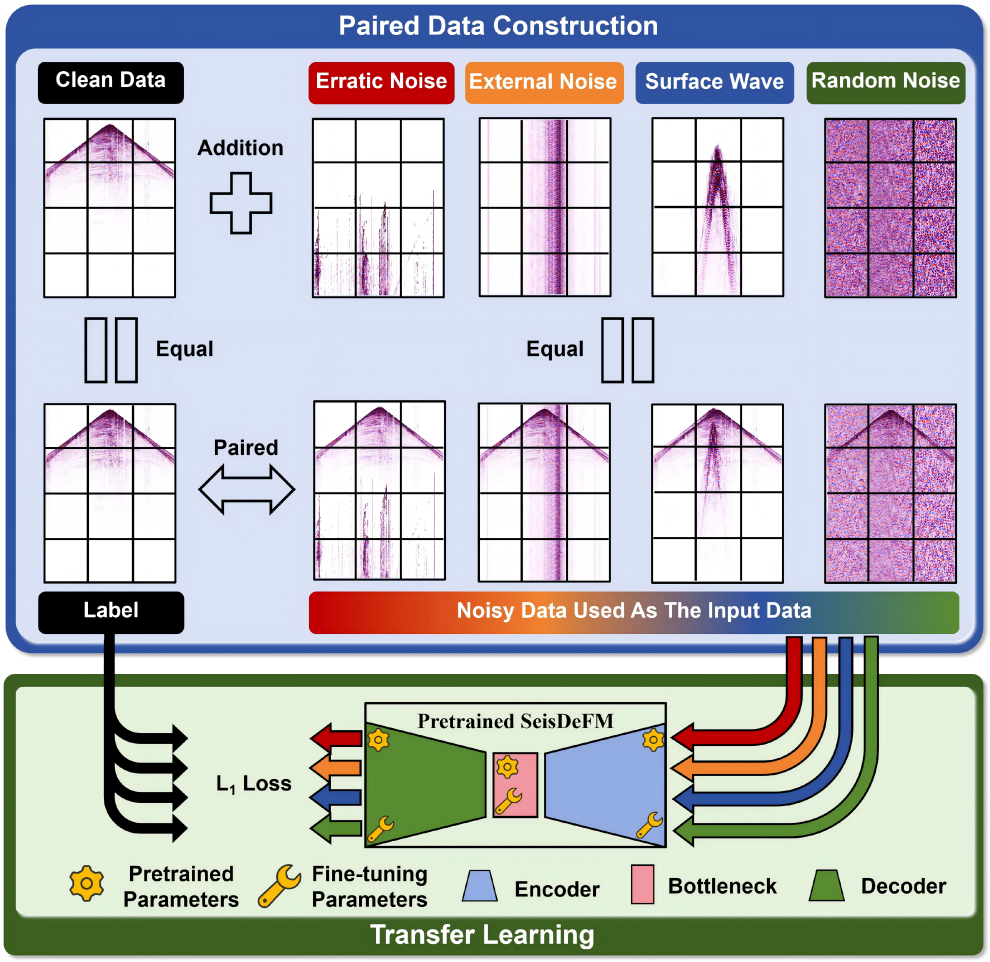



**Fig. 5** The downstream adaptation strategy of SeisDeFM.

specific denoising task, we update all model parameters using a small-scale paired dataset, as formulated in Equation (8). The denoised results are obtained via element-wise wavefield subtraction, as illustrated in Equation (13). The optimal parameters of SeisDeFM are solved for under the L1-norm constraint for downstream denoising tasks:

$$\widehat{\Theta} = \arg\min_{\Theta} \frac{1}{M} \sum_{i=1}^{M} \left\| F_{\Theta}\left[y_i(h)\right] - x_i \right\|_1, \tag{19}$$

By leveraging the inherent wavefield characteristics captured during pre-training, this full-parameter fine-tuning strategy enables rapid alignment with the target noise distribution while effectively mitigating the risk of overfitting on small-scale downstream datasets. The downstream adaptation strategy adopted in our SeisDeFM case study is illustrated in Fig. 5. To comprehensively demonstrate the efficient downstream

adaptation capabilities, we select erratic noise, external noise, surface waves and random noise as downstream evaluation cases. These interference components represent the most typical seismic interferences, ranging from ubiquitous incoherent energies and localized high-amplitude anomalies to strongly coherent dispersive wavefields, thereby rigorously validating the universal multi-noise adaptability of SeisDeFM.

## 2.6 The evaluation metrics for SeisDeFM

To comprehensively evaluate the performance of SeisDeFM, multiple evaluation metrics should be employed, which can be broadly categorized into computational efficiency metrics and quantitative denoising quality metrics.

Evaluating SeisDeFM requires a detailed accounting of both data requirements and computational cost. The following metrics are usually used to quantify the efficiency of SeisDeFM:

(1) Data volume and training time. We record the volume of training data, including the clean data used for pre-training and the paired clean–noisy data used for transfer learning, as well as the corresponding training time for both the self-supervised pre-training stage and the downstream adaptation stage. A robust SeisDeFM should demonstrate that pre-training with massive data significantly shortens convergence time and reduces dependency on data volume during transfer learning.

(2) Trainable parameters within SeisDeFM. This metric counts the total number of trainable weights and biases in the SeisDeFM architecture. It reflects the memory footprint and the feasibility of deploying SeisDeFM on computationally constrained platforms.

(3) Floating-point operations. Floating-point operations (FLOPs) measure the total number of mathematical operations required to process input seismic data. This metric serves as a direct indicator of SeisDeFM's computational complexity and inference speed, which is vital for large-scale industrial seismic data processing.

To evaluate the fidelity of signal–noise separation achieved by SeisDeFM denoising, the following four core quantitative metrics can be adopted for this purpose:

(1) Mean absolute error (MAE) calculates the average absolute difference between the denoised and the clean seismic data:

$$\mathrm{MAE} = \frac{1}{M}\sum_{i=1}^{M}\left|\widehat{x}_i - x_i\right|, \tag{20}$$

(2) Root-mean-square error (RMSE) measures the standard deviation of residual errors:

$$\mathrm{RMSE} = \sqrt{\frac{1}{M}\sum_{i=1}^{M}\left(\widehat{x}_i - x_i\right)^2}. \tag{21}$$

By squaring the amplitude differences, RMSE heavily penalizes larger local errors, effectively reflecting the residual noise energy remaining in the denoising result.

(3) Peak signal-to-noise ratio (PSNR) evaluates the logarithmic ratio of the maximum possible signal power to the mean squared error of the signal distortion:

$$\mathrm{PSNR} = 10\log_{10}\left[\frac{\max(x)^2}{\mathrm{MSE}}\right], \tag{22}$$

where $\mathrm{MSE} = \mathrm{RMSE}^2$, and $\max(\cdot)$ denotes the operator that calculates the peak amplitude of the clean seismic data. A higher PSNR value indicates a higher degree of overall signal fidelity and superior noise suppression.

(4) Structural similarity index measure (SSIM) measures the structural coherence between the denoised and clean data by comparing local patterns of pixel intensities:

$$\mathrm{SSIM}(x, \widehat{x}) = \frac{\left(2\mu_x\mu_{\widehat{x}} + c_1\right)\left(2\sigma_{x\widehat{x}} + c_2\right)}{\left(\mu_x^2 + \mu_{\widehat{x}}^2 + c_1\right)\left(\sigma_x^2 + \sigma_{\widehat{x}}^2 + c_2\right)}, \tag{23}$$

where $\mu_x$ and $\mu_{\widehat{x}}$ are the mean values of clean data and denoised result, respectively, $\sigma_x^2$ and $\sigma_{\widehat{x}}^2$ are the variances of clean data and denoised result, respectively, $\sigma_{x\widehat{x}}$ denotes the covariance between the clean data and the denoised result, and $c_1$, $c_2$ are small constants to stabilize the division. A higher SSIM directly corresponds to the better preservation of local features of seismic events.

# 3 A case study of SeisDeFM

To validate the effectiveness of SeisDeFM, a series of experiments were conducted under different noise conditions. The noisy seismic data used in this study were collected from a field seismic survey in China. The complete dataset contains 12,000 samples, including 10,000 clean-noisy paired training samples, 1,000 validation samples, and 1,000 testing samples. All data used in this case study have a sampling interval of 0.002 s and a trace spacing of 25 m. In our setup, SeisDeFM was first pre-trained on all clean data from the full training dataset for 30 epochs. The pre-trained weights were then adopted to initialize SeisDeFM for the downstream denoising task. During the subsequent supervised transfer training stage, only 1,000 paired samples were randomly selected from the training set for another 30-epoch training. U-Net and Transformer baselines, which belong to TSDM, were adopted as competing models and trained on the complete training dataset for 30 epochs. The adopted Transformer baseline uses four attention heads in the multi-head attention module, and four Transformer blocks. The adopted U-Net baseline consists of a four-layer encoder and a four-layer decoder, where each layer contains two 3× 3 convolution layers and two 1× 1 channel attention convolution layers. All models in this study were trained on the same NVIDIA A10 GPU with a batch size of 8. The AdamW optimizer was employed for training with an initial learning rate of $1 \times 10^{-4}$, and the learning rate remained fixed throughout training, with no additional decay schedule applied.

## 3.1 The analysis of training efficiency

Since the dataset size and experimental environment are identical for all experiments in this case, the computational complexity of each model does not vary across different downstream denoising tasks. Accordingly, this section compares and analyzes the computational complexity and training efficiency of SeisDeFM and TSDM. Table 1 shows that the U-Net baseline contains 10.67 M trainable parameters and

requires 341.14 GFLOPs. Its convolution-dominated encoder-decoder architecture enables repeated multiscale feature extraction and dense spatial aggregation, resulting in the longest supervised training time of 1,538 min. The Transformer baseline adopted in this study contains 14.52 M trainable parameters and requires 38.40 GFLOPs. Although the Transformer and U-Net baselines have comparable numbers of trainable parameters, Transformer requires substantially fewer FLOPs due to its more efficient feature representation, while its training time remains comparable because self-attention involves computationally intensive query–key similarity calculations and attention-weighted feature aggregation, which introduce substantial memory and computational overhead. SeisDeFM has 12.84 M trainable parameters and consumes 604 GFLOPs in the pre-training stage. After pre-training, SeisDeFM in the transfer-learning stage has 12.51 M trainable parameters and exhibits a computational cost comparable to the U-Net baseline with 341.20 GFLOPs. These results demonstrate that the additional Transformer-based bottleneck introduce only a modest increase in model size and computational overhead. More importantly, SeisDeFM requires merely 1,000 paired samples for transfer learning, shortening its supervised training time to 108 min. Therefore, although SeisDeFM is not the lightest architecture, its pre-training computational cost is comparable to those of TSDM models. It further supports fast and lightweight transfer learning for noise-specific denoising tasks using a small number of paired samples, achieving accurate denoising under multiple noise types.

**Table 1** Quantitative comparison of training efficiency.

| Method | Training Data Volume | Training Time (min) | Trainable Parameters (M) | GFLOPs |
|---|---|---|---|---|
| U-Net | 10,000 | 1,538 | 10.67 | 341.14 |
| Transformer | 10,000 | 1,204 | 14.52 | 38.40 |
| SeisDeFM (pre-training) | 10,000 | 1,396 | 12.84 | 604.00 |
| SeisDeFM (transfer learning) | 1,000 | 108 | 12.51 | 341.20 |

## 3.2 Erratic Noise Attenuation

Erratic noise corresponds to sparse, non-Gaussian, impulsive random anomalies during field seismic data acquisition. It generally manifests as locally isolated high-amplitude perturbations that can mask weak reflections. To evaluate the denoising capability of SeisDeFM for noisy seismic data contaminated with erratic noise, we compared the denoised results obtained by Transformer, U-Net and SeisDeFM under identical training conditions.

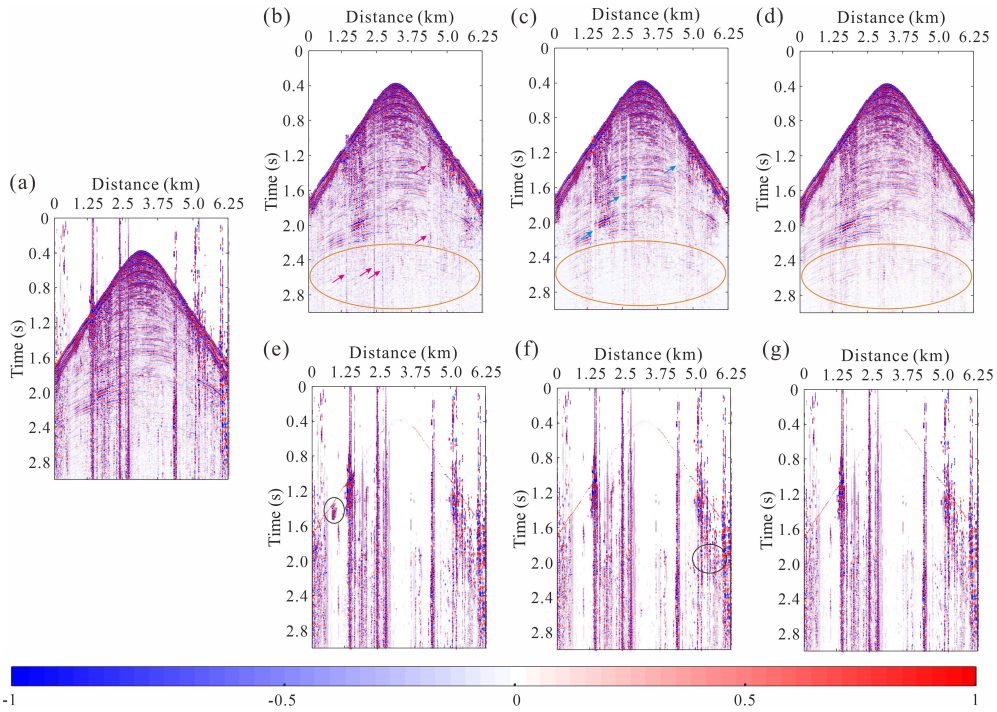



**Fig. 6** Comparisons of erratic-noise attenuation. (a) Raw noisy input data, (b)-(d) denoised results of Transformer, U-Net, and SeisDeFM, (e)-(g) the residual profiles between the input data shown in (a) and the denoised results shown in (b)-(d).

A representative sample of raw noisy data was selected for comparison, as shown in Fig. 6(a). Fig. 6(b) shows the denoised result obtained by Transformer, and Fig. 6(e) shows its residual profile. As highlighted by the red arrows in Fig. 6(b), Transformer fails to completely eliminate all the high-amplitude erratic noise, with residual noise remaining at different time instants. This phenomenon is mainly caused by the lack of local waveform constraints in the global representation of self-attention, where effective reflections and sparse noise perturbations may be mixed in the global attention space. The corresponding residual map shown in Fig. 6(e) further reveals that, in addition to incomplete noise suppression, Transformer suffers from locally continuous signal leakage as indicated by the black circle. Fig. 6(c) shows the denoised result of U-Net. Compared with Transformer, U-Net achieves better removal of the high-amplitude erratic noise. Its convolution-based local feature extraction performs well in suppressing the erratic noise characterized by abnormal amplitude variations. However, as indicated by the blue arrows in Fig. 6(c), some of the seismic signals are also removed during denoising. In addition, the black circle in Fig. 6(f) indicates that the signal leakage exists in the residual profile. This is because the limited receptive field of convolutional operations prevents U-Net from fully capturing long-range dependencies and the global continuity of seismic events, causing weak reflections to be misidentified as noise. In contrast, SeisDeFM produces the cleanest denoised result (Figs. 6(d)

and 6(g)). In particular, as highlighted by the yellow circles in Figs. 6(b)-6(d), both the Transformer and U-Net baselines exhibit obvious residual noise and signal loss in the denoised results of late arrivals. In contrast, SeisDeFM preserves the continuity and amplitude of events for deep weak signals while removing noise effectively. By combining the local feature extraction capability of U-Net with the long-range dependency modeling of Transformer, and by introducing masked reconstruction and contrastive learning during pre-training, SeisDeFM better distinguishes erratic noises from physically meaningful seismic events. As a result, it suppresses strong noise while maintaining event continuity. Notably, SeisDeFM uses only 1,000 transfer-learning samples and requires the shortest training time (Table 1), while yielding the best denoising performance.

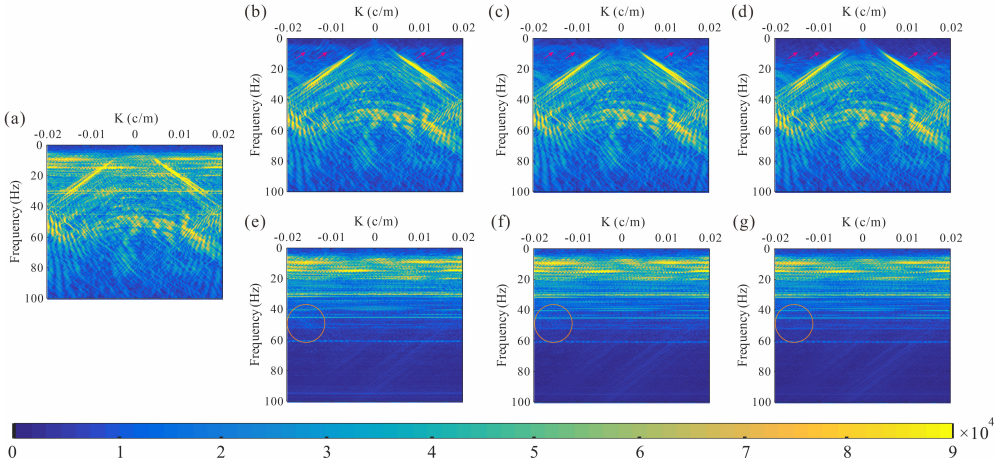



**Fig. 7** F-K spectra comparisons of erratic noise attenuation. (a)-(g) F-K spectra of the data shown in Figs. 6(a)-6(g).

We further conducted F-K spectral analysis on the data shown in Fig. 6, as presented in Fig. 7. As indicated by the red arrows in Figs. 7(b)-7(d), obvious low-frequency residual noise exists in the denoised results of Transformer and U-Net, whereas SeisDeFM achieves favorable separation of low-frequency signals and noise. Meanwhile, as indicated by the yellow circles in Figs. 7(e) and 7(f), signal leakage can be observed in the denoised results of Transformer and U-Net. However, this limitation is effectively alleviated by SeisDeFM as illustrated by Fig 7(g). These F-K spectra demonstrate that SeisDeFM further improves erratic-noise suppression and maintains the spectral integrity of seismic signals.

## 3.3 External Noise Attenuation

External noise refers to interference introduced by environmental or human activities during seismic acquisition, such as vehicle movement, mechanical vibration, electrical interference, and other nearby operations. This type of noise may exhibit coherent characteristics and can substantially degrade the quality of field seismic records.

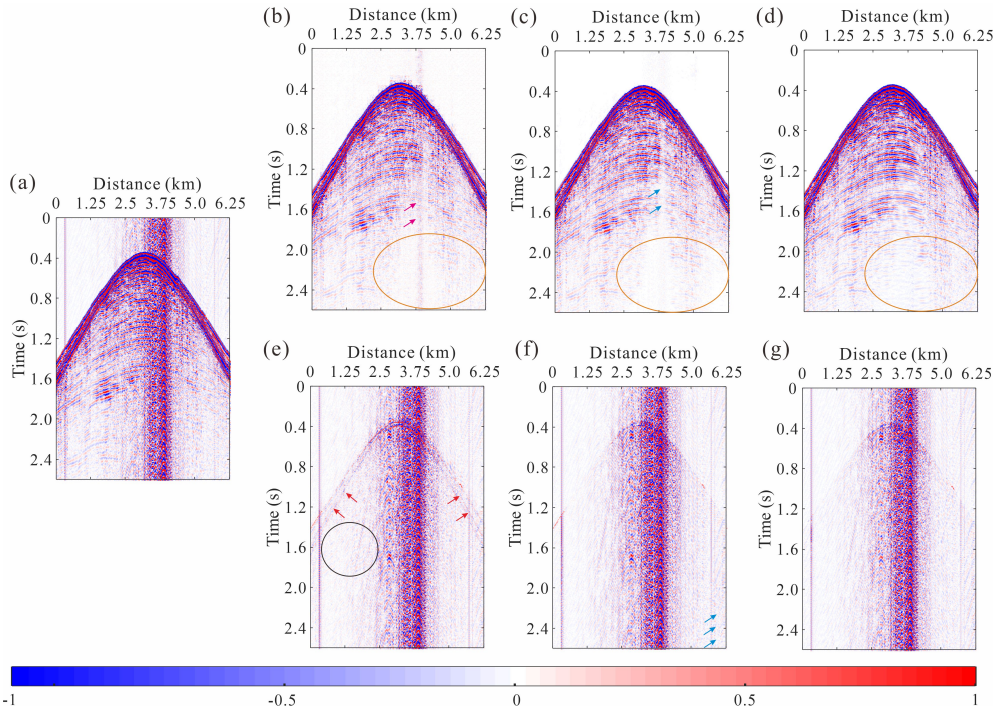



**Fig. 8** Comparisons of external-noise attenuation. (a) Raw noisy input data, (b)-(d) denoised results of Transformer, U-Net, and SeisDeFM, (e)-(g) the residual profiles between the input data shown in (a) and the denoised results shown in (b)-(d).

The denoised results for a raw noisy data are shown in Fig. 8. As can be seen in Figs. 8(b), strong residual external-noise can be observed in the denoised results of Transformer, as indicated by the red arrows. Meanwhile, the residual profile of Transformer shows strong leakage of early arrivals (as indicated by the red arrows) and continuous reflections at later time (as indicated by the black circle). This reveals that the global representation of the Transformer baseline is insufficient to distinguish external noise from weak but continuous seismic reflections. Fig. 8(c) shows that U-Net removes most of the external noise, but signal leakage still exists in the region contaminated by strong noise, as indicated by the blue arrows in Fig. 8(c). The residual profile of the U-Net denoised result shows incomplete denoising for late arrivals, as indicated by the blue arrows in Fig. 8(f). In contrast, SeisDeFM produces the cleanest denoised result and suppresses most external noise (Figs. 8(d) and 8(g)). In particular, as highlighted by the yellow circles in Figs. 8(b)-8(d), both the Transformer and U-Net baselines exhibit severe signal leakage in the denoised results of late arrivals, which indicates that effective reflection energies are incorrectly removed together with the external noise. However, SeisDeFM preserves the continuity and amplitude of events for deep weak signals while removing noise effectively. These results demonstrate that SeisDeFM can better suppress external noise while preserving effective seismic signals.

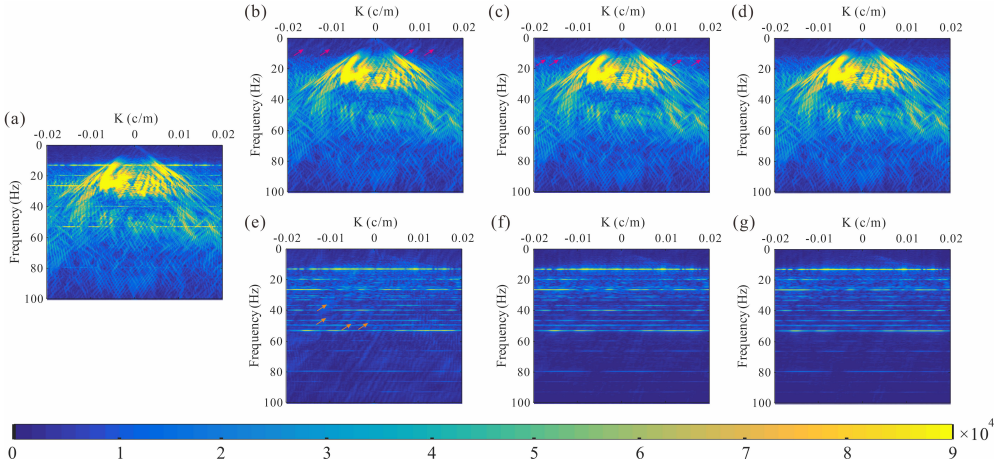



**Fig. 9** F-K spectra comparisons of external-noise attenuation. (a)-(g) F-K spectra of the data shown in Figs. 8(a)-8(g).

To further evaluate the F-K-spectral behavior of different methods, we performed F-K domain analysis on the data shown in Fig. 8, with the results shown in Fig. 9. The denoised result obtained by Transformer introduces low-frequency noise that does not exist in the original noisy data, as indicated by the red arrows in Fig. 9(b). Furthermore, the denoised residual of Transformer exhibits obvious signal energy leakage, as indicated by the yellow arrows in Fig. 9(e). Compared with Transformer, the F-K spectrum of the U-Net denoised result shown in Fig. 9(c) does not introduce extra low-frequency noise. Nevertheless, residual low-frequency external noise from the original noisy data remains, as indicated by the red arrows in Fig. 9(c). In contrast, SeisDeFM yields the cleanest F-K spectrum for the denoised result, demonstrating its stronger ability to separate external noise from effective seismic signals.

## 3.4 Surface Waves Attenuation

Surface waves are coherent noise characterized by large amplitudes, low frequencies, and low apparent velocities. Surface waves often mask near-offset reflections and degrade subsequent seismic processing.

The denoised results of the raw noisy data are shown in Fig. 10. Prominent surface-wave residuals can be observed in the denoised result of Transformer, as indicated by the blue arrows and yellow circle in Fig. 10(b). Furthermore, the residual profile of the Transformer denoised result exhibits direct-wave damage (indicated by the red arrows in Fig. 10(e)) and leakage of continuous reflections at early and later time (indicated by the black circles in Fig. 10(e)). Although the denoised result of U-Net yields better surface-wave suppression compared with that of the Transformer, surface-wave residuals exist in regions with deep weak reflections, and the amplitude recovery of effective events is insufficient, as indicated by the yellow circle in Fig. 10(c). In contrast, SeisDeFM suppresses surface waves more completely while preserving the continuity and amplitude fidelity of effective reflections (Figs. 10(d) and 10(g)), showing a better

balance between noise attenuation and signal preservation. This test demonstrates that SeisDeFM achieves the best performance while requiring only 1,000 transfer-learning samples, indicating that the pre-trained seismic feature representations are also effective for suppressing coherent surface waves.

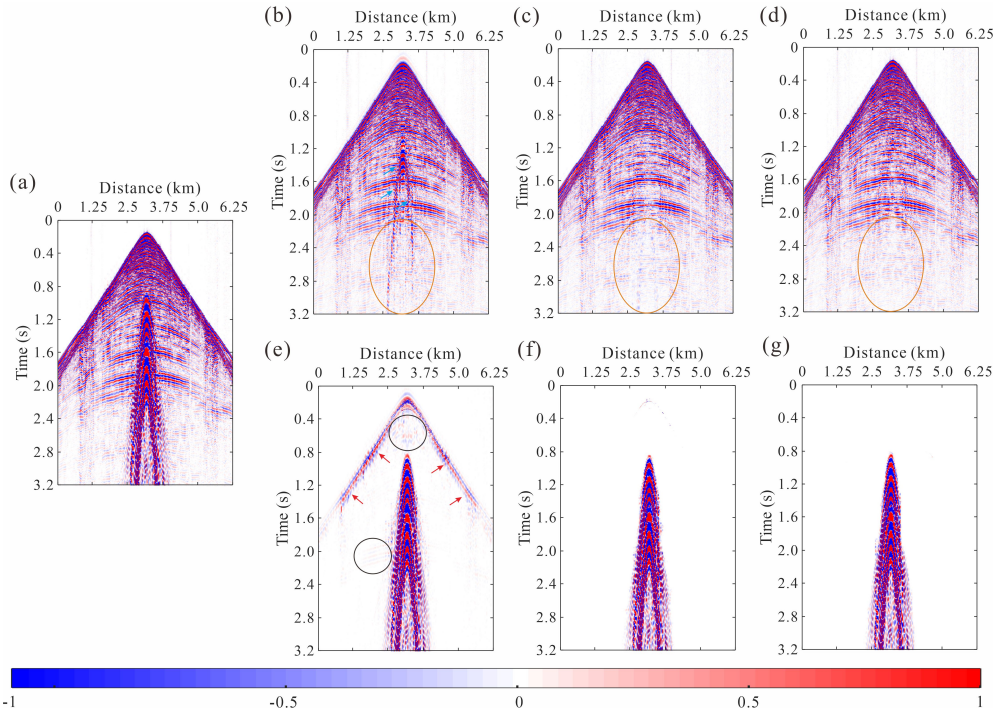



**Fig. 10** Comparisons of surface-wave attenuation. (a) Raw noisy input data, (b)-(d) denoised results of Transformer, U-Net, and SeisDeFM, (e)-(g) the residual profiles between the input data shown in (a) and the denoised results shown in (b)-(d).

We further compared the corresponding F-K spectra for the data in Fig. 10, with the results presented in Fig. 11. The F-K spectrum of the Transformer's denoised result shows strong residual surface-wave energy in the low-frequency band, as indicated by the red circle in Fig. 11(b). Moreover, in the F-K spectrum of the Transformer's residual profile shown in Fig. 11(e), apart from the suppressed low-frequency surface-wave energy, the energy of leaked signals is widely distributed across the full frequency band. The F-K spectrum of U-Net's denoised result achieves considerable improvement over that of Transformer. The surface-wave energy in the low-frequency band is effectively eliminated, yet leakage of low-frequency signals remains, as indicated by the red arrow in Fig. 11(c). In contrast, as shown in Figs. 11(d) and 11(g), SeisDeFM produces the F-K spectra of the denoised result and residual profile with weaker surface-wave energy and less signal leakage in the low-frequency band, demonstrating better performance for SeisDeFM in suppressing surface waves in the F-K domain.

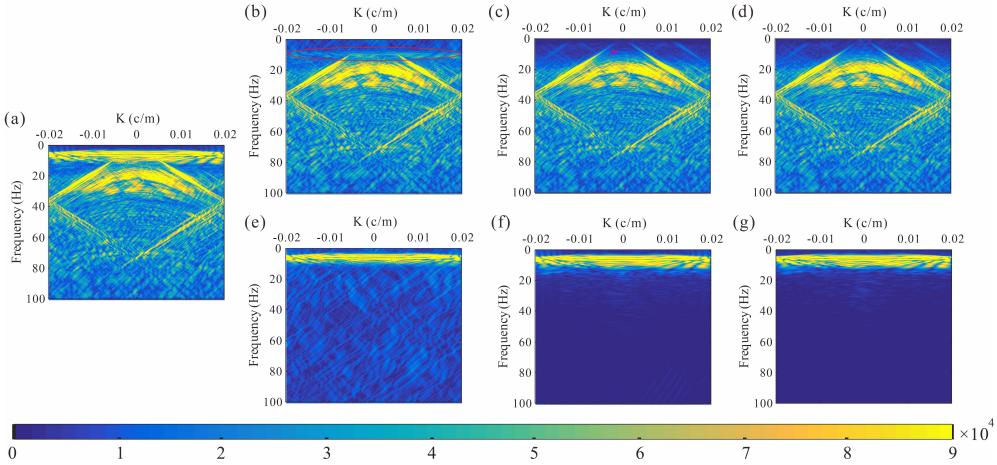



**Fig. 11** F-K spectra comparisons of surface-wave attenuation. (a)-(g) F-K spectra of the data shown in Figs. 10(a)-10(g).

## 3.5 Random Noise Attenuation

We further performed a denoising test on field seismic data contaminated by random noise. Random noise does not follow a stable spatial pattern and may be distributed over a broad frequency range, making it difficult to attenuate without damaging weak reflection events.

Comparison of denoised results for the field data under random-noise contamination is shown in Fig. 12. As indicated by the yellow circles in Figs. 12(b)–12(d), the denoised results of the Transformer and U-Net baselines contain noise residuals in regions with deep weak signals and exhibit poor recovery of effective reflections. By contrast, SeisDeFM can effectively remove random noise while restoring the continuity and amplitude of deep reflections. As indicated by the red arrows in Fig. 12(e), the absolute values at the region of early arrivals in the Transformer's denoising residual are small. This indicates that Transformer has poor denoising performance for the region where high-amplitude early arrivals overlap with random noise, making effective signal-noise separation difficult. As indicated by the black arrows in Fig. 12(f), the absolute values at the region of early arrivals in U-Net's denoising residual are large. This indicates that U-Net misidentifies early arrivals as noise and attenuates them while removing random noise, resulting in signal leakage in this region. SeisDeFM performs the best in both noise suppression and signal preservation (Figs. 12(d) and 12(g)).

We further compared the corresponding F-K spectra for the data in Fig. 12, with the results presented in Fig. 13. As indicated by the red circles in Figs. 13(b)–13(d), the residual energy of low-frequency noise exists in the F-K spectra of the denoised result from the Transformer and U-Net baselines. However, SeisDeFM can suppress low-frequency noise while preserving the effective signal energy in the low-frequency band. As indicated by the red arrows in Figs. 13(e) and 13(f), leaked signal energy exists in the F-K spectra of the denoising residuals from the Transformer and U-Net

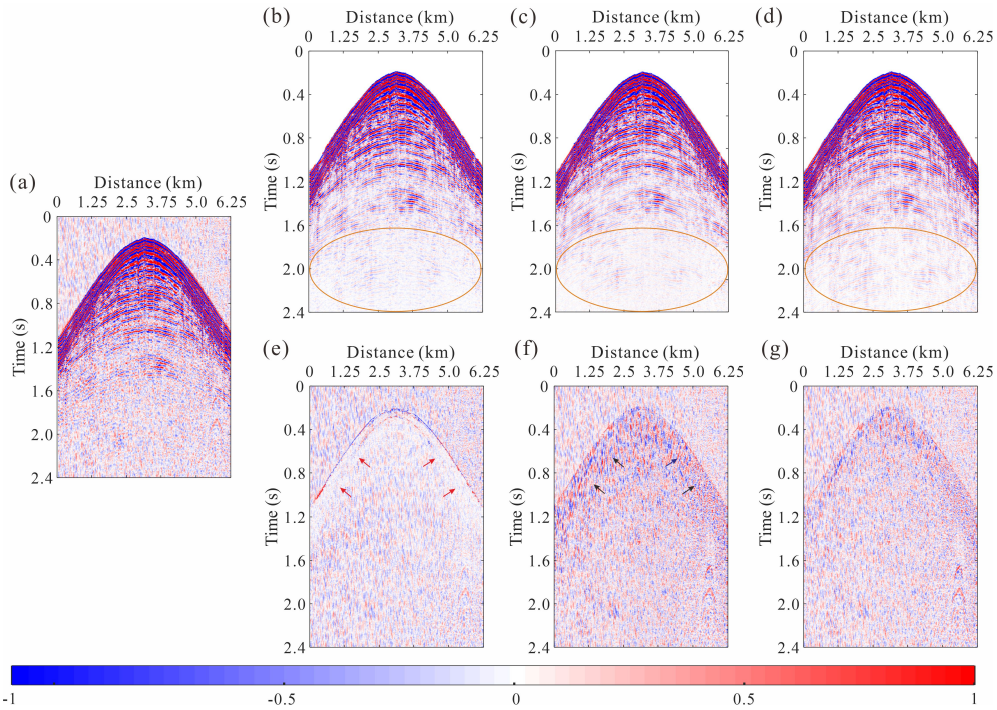



**Fig. 12** Comparisons of random-noise attenuation. (a) Raw noisy input data, (b)-(d) denoised results of Transformer, U-Net, and SeisDeFM, (e)-(g) the residual profiles between the input data shown in (a) and the denoised results shown in (b)-(d).

baselines. In contrast, SeisDeFM shows less random-noise energy in the denoised F-K spectrum (Fig. 13(d)) and weaker signal energy in the residual F-K spectrum (Fig. 13(g)). This test demonstrates that SeisDeFM not only suppresses random noise in the time-space domain but also better preserves the F-K spectral distribution of signals.

# 4 Discussion

## 4.1 Ablation study of the pre-training strategy

To analyze the contribution of different pre-training strategies, we conducted an ablation study on the external-noise attenuation task under an extremely low SNR condition. We compared four experimental settings: training from scratch without pre-training, pre-training with masked reconstruction only, pre-training with contrastive learning only, and the proposed joint pre-training strategy that combines masked reconstruction with contrastive learning. All models were trained via transfer learning for 30 epochs. The denoised result of a test sample is shown in Fig. 14. Without pre-training, obvious noise remains in the denoised result, as indicated by the red arrows in Fig. 14(b). Moreover, severe signal leakage is exits in the deep reflections, as indicated by the yellow circle in Fig. 14(b). As shown in Fig. 14(c), the denoised result

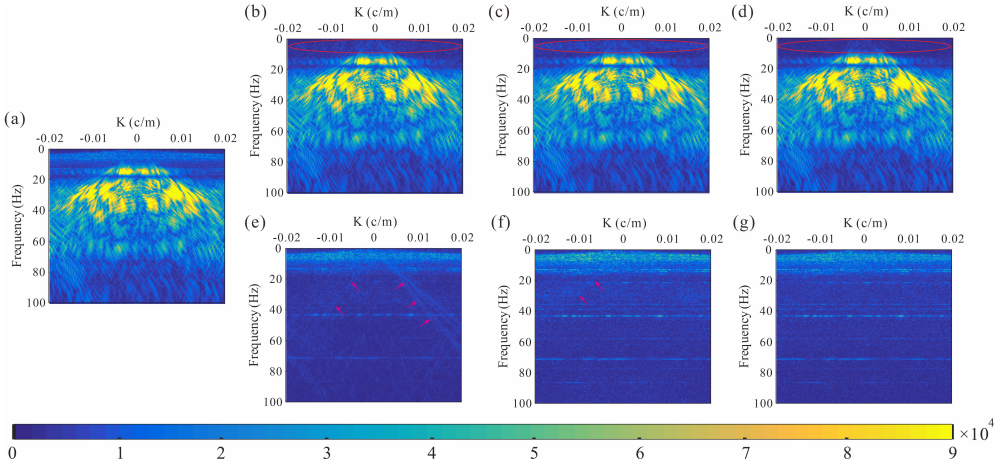



**Fig. 13** F-K spectra comparisons of random noise attenuation. (a)-(g) F-K spectra of the data shown in Figs. 12(a)-12(g).

obtained under the mask-reconstruction-only pre-training strategy reduces external-noise residuals. This occurs because SeisDeFM acquires the capability to recover effective events during the self-supervised pre-training stage of mask reconstruction, enabling better separation of signals and noise after transfer learning for the downstream task. However, signal leakage still exists in deep weak reflection regions after denoising, as indicated by the yellow circle in Fig. 14(c). As shown in Fig. 14(d), the denoised result under the contrastive-learning-only pre-training strategy also reduces external-noise residuals. This is because SeisDeFM acquires the capability to distinguish positive and negative sample pairs during contrastive learning, which enables better separation of noise from noisy data. However, severe signal leakage appears in regions covered by strong noise (as indicated by the blue arrows in Fig. 14(d)) and in deep weak reflection regions (as marked by the yellow circle in Fig. 14(d)). In contrast, the joint pre-training strategy yields cleaner denoised results (Fig. 14(e)) and mitigates signal leakage in the residual profiles (Fig. 14(i)). As indicated by the black circles in Figs. 14(f)–14(i), leakage of continuous events exists in the denoising residuals obtained from the no pre-training scheme, the mask-reconstruction-only pre-training strategy, and the contrastive-learning-only pre-training strategy. This ablation study demonstrates that context recovery and representation alignment in the pre-training stage are complementary for separating noises from seismic events. Although the joint pre-training strategy incurs higher training costs, the additional training time is little compared with the two standalone pre-training approaches, and it yields significant improvements in denoising performance.

## 4.2 Comparisons of different transfer learning methods

We also investigated the effects of different transfer learning methods on noise attenuation after pre-training, using test data containing random noise. Three transfer learning methods were compared: updating only the decoder of the SeisDeFM

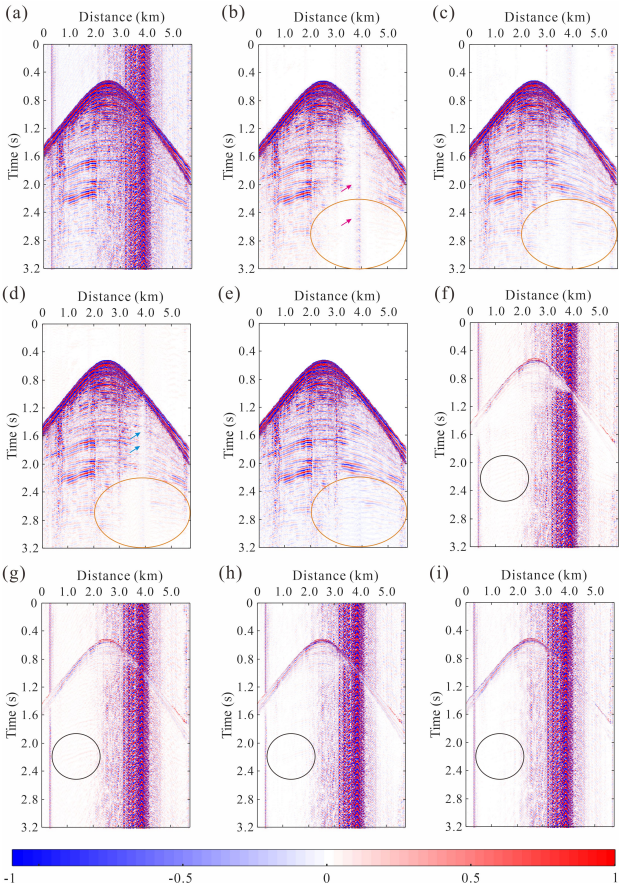



**Fig. 14** Comparisons of external-noise attenuation via different pre-training strategies. (a) Raw noisy input data, (b)-(e) denoised results based on no pre-training, masked reconstruction, contrastive learning, and the joint strategy, (f)-(i) the residual profiles between the input data shown in (a) and the denoised results shown in (b)-(e).

architecture, adding and updating lightweight adapter modules, and fine-tuning all model parameters. All methods were trained for 30 epochs with a learning rate of $1 \times 10^{-4}$. In the decoder-only setting, the encoder was frozen and only the decoder and output layer were updated. In the adapter setting, lightweight trainable adapter modules were inserted into the original architecture, while the main encoder and decoder weights were frozen. The architecture of an adapter consists of normalization, dimension-reduction projection, activation, dropout, dimension-expansion projection, and residual addition. One adapter is embedded after each encoder layer and each decoder layer. In our case study of SeisDeFM, full-parameter fine-tuning was adopted, wherein all model parameters were fine-tuned so that pre-trained representations could fully adapt to the downstream denoising task.

Table 2 lists the training cost and parameter statistics for different transfer learning methods. The adapter strategy has a slightly larger total parameter count due to the

introduction of extra adapter modules, increasing the model size from 12.51 M to 12.64 M. However, only 142,562 parameters are trainable, corresponding to 1.13% of the total parameters, and the training time is reduced to 80 min. The decoder-only method uses the same network structure as the full-parameter fine-tuning method, so the two strategies have identical total parameters and FLOPs. Because the encoder is frozen, decoder-only tuning updates 3.25 M parameters, accounting for 26.02% of the total parameters, and requires 97 min for training. Full-parameter fine-tuning updates all 12.51 M parameters and takes 108 min for training. Although it introduces a slightly higher computational cost, it provides the network with greater flexibility to adapt to different levels of feature representations.

**Table 2** Quantitative comparisons of model complexity under different transfer learning methods.

| **Method** | **Training Time (min)** | **Total Parameters** | **Trainable Parameters** | **Trainable Ratio** | **GFLOPs** |
|---|---|---|---|---|---|
| Decoder-only | 97 | 12.51 M | 3,254,461 | 26.02% | 341.20 |
| Adapter | 80 | 12.64 M | 142,562 | 1.13% | 353.22 |
| Full-parameter fine-tuning | 108 | 12.51 M | 12,508,666 | 100.00% | 341.20 |

Comparisons of denoised results for the raw noisy data are presented in Fig. 15. The decoder-only and adapter-based tuning methods exhibit poor signal preservation in the deep regions contain weak reflections indicated by the yellow circles in Figs. 15(b) and 15(c). This indicates that partial-parameter updating may not be sufficient to correct the mismatch between the pre-trained representation and the target random-noise distribution. Furthermore, since the adapter-based method has the fewest trainable parameters, most signal leakage appears in its denoised result, as indicated by the yellow circles in Fig. 15(c). As indicated by the red arrows in Figs. 15(e)–15(f), the decoder-only tuning and adapter-based tuning fail to effectively separate signals from noise in early-arrival regions. This leads to small absolute values of the denoising residuals in these regions, and the adapter-based tuning strategy yields even worse performance. In contrast, full-parameter fine-tuning produces a cleaner denoised result and a residual profile with less signal leakage (Figs. 15(d) and 15(g)). This test demonstrates that updating the complete network is beneficial when the target noise pattern differs from the information learned during pre-training.

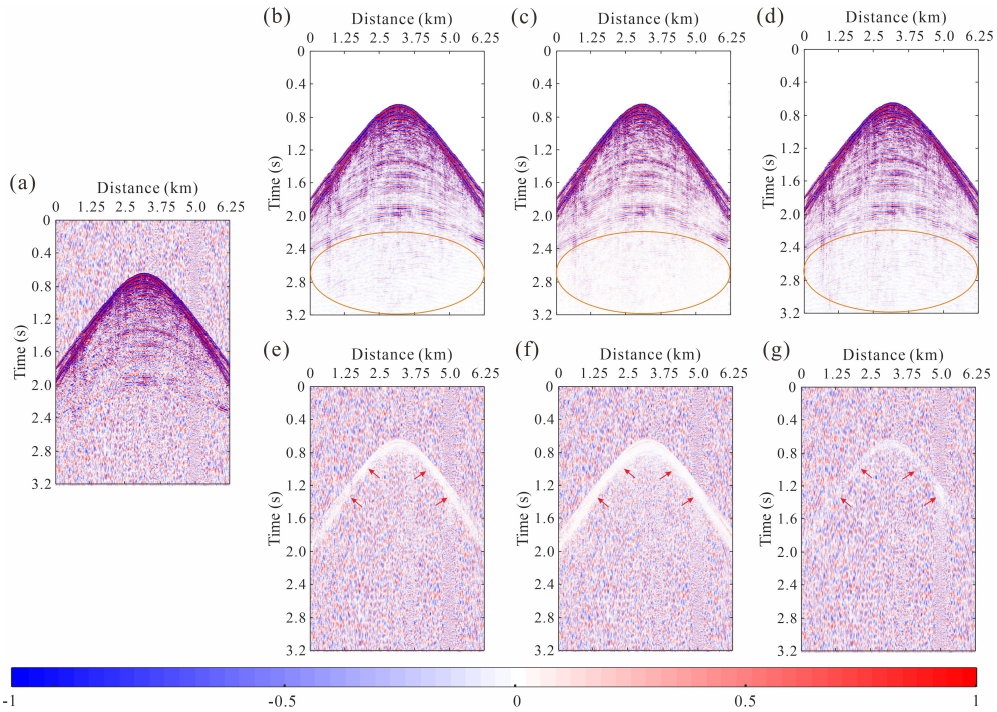



**Fig. 15** Comparisons of random-noise attenuation via different transfer learning methods. (a) Raw noisy input data, (b)-(d) denoised results of decoder-only tuning, adapter-based tuning and full fine-tuning, (e)-(g) the residual profiles between the input data shown in (a) and the denoised results shown in (b)-(d).

# 5 Conclusions

This paper presents a dedicated seismic denoising foundation model termed SeisDeFM to address the critical limitation of poor generalization among existing task-specific deep-learning denoising methods. SeisDeFM achieves stronger generalizability across diverse noise types and complex survey conditions. We present a case study of SeisDeFM, which adopts a hybrid architecture integrating a residual U-Net backbone for local multi-scale feature extraction and a trace-wise Transformer bottleneck for global wavefield dependency learning. It employs a joint self-supervised pre-training strategy combining masked wavefield reconstruction and contrastive learning, enabling the model to learn universal geophysical priors for seismic wavefields from large-scale clean data without paired noisy-clean ground-truth data. Extensive experiments on field seismic data contaminated by erratic noise, external noise, surface waves, and random noise demonstrate that SeisDeFM consistently outperforms the U-Net and Transformer baselines using only one-tenth of the labeled training data. This work validates that the foundation-model paradigm breaks the generalization ceiling of conventional task-specific denoising approaches, shifting seismic denoising from isolated task-specific mapping toward universal wavefield representation and efficient noise

adaptation, and provides a promising pathway for large-scale industrial seismic data processing.

## Acknowledgements

This study was jointly supported by the National Natural Science Foundation of China (Grant No. 42574169), the Deep Earth Probe and Mineral Resources Exploration—National Science and Technology Major Project (Grant No. 2025ZD1007600), and the Jilin Provincial Natural Science Foundation-Free Exploration Project (Grant No. YDZJ202501ZYTS530).

## Declarations

**Conflict of interest.** The submission of this manuscript involves no conflicts of interest.

## References

Abma R, Claerbout J (1995) Lateral prediction for noise attenuation by tx and fx techniques. Geophysics 60(6):1887–1896

Awais M, Naseer M, Khan S, et al (2025) Foundation models defining a new era in vision: A survey and outlook. IEEE Transactions on Pattern Analysis and Machine Intelligence 47(4):2245–2264

Bahdanau D, Cho K, Bengio Y (2014) Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:14090473

Bao Q, Xu W, Shi W, et al (2025) Unsupervised seismic erratic noise suppression using implicit neural representation. IEEE Geoscience and Remote Sensing Letters 22:1–5

Beckouche S, Ma J (2014) Simultaneous dictionary learning and denoising for seismic data. Geophysics 79(3):A27–A31

Bekara M, Van der Baan M (2007) Local singular value decomposition for signal enhancement of seismic data. Geophysics 72(2):V59–V65

Bekara M, Van der Baan M (2009) Random and coherent noise attenuation by empirical mode decomposition. Geophysics 74(5):V89–V98

Beresford-Smith G, Rango R (1989) Suppression of ground roll by windowing in two domains. First Break 7(2)

Birnie C, Ravasi M, Liu S, et al (2021) The potential of self-supervised networks for random noise suppression in seismic data. Artificial Intelligence in Geosciences 2:47–59

Brown T, Mann B, Ryder N, et al (2020) Language models are few-shot learners. Advances in Neural Information Processing Systems 33:1877–1901

Canales L (1984) Random noise reduction. In: SEG Technical Program Expanded Abstracts 1984. Society of Exploration Geophysicists, 525–527. Society of Exploration Geophysicists

Caron M, Touvron H, Misra I, et al (2021) Emerging properties in self-supervised vision transformers. Proceedings of the IEEE/CVF International Conference on Computer Vision, 9650–9660

Chen H, Wang J, Liu B (2025) Dual-domain mamba for seismic random noise suppression. Journal of Applied Geophysics 243:105951

Chen Y, Ma J (2014) Random noise attenuation by fx empirical-mode decomposition predictive filtering. Geophysics 79(3):V81–V91

Cheng M, Lin J, Dong X, et al (2025a) A multitask deep-learning model for the denoising, interpolation, and wavefield separation of distributed acoustic sensing-vertical seismic profiling data. Geophysics 90(6):V559–V568

Cheng S, Cheng Z, Jiang C, et al (2024) An effective self-supervised learning method for attenuating various types of seismic noise. Geophysics 89(6):V589–V604

Cheng S, Harsuko R, Alkhalifah T (2025b) A generative foundation model for an all-in-one seismic processing framework. Surveys in Geophysics 46(6):1173–1215

Devlin J, Chang MW, Lee K, et al (2019) BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 4171–4186

Dong X, Li Y (2020) Denoising the optical fiber seismic data by using convolutional adversarial network based on loss balance. IEEE Transactions on Geoscience and Remote Sensing 59(12):10544–10554

Dong X, Jiang H, Zheng S, et al (2019a) Signal-to-noise ratio enhancement for 3C downhole microseismic data based on the 3D shearlet transform and improved back-propagation neural networks. Geophysics 84 (4): V245–V254

Dong X, Li Y, Yang B (2019b) Desert low-frequency noise suppression by using adaptive dncnns based on the determination of high-order statistic. Geophysical Journal International 219(2):1281–1299

Dong X, Li Y, Zhong T, et al (2020) Random and coherent noise suppression in das-vsp data by using a supervised deep learning method. IEEE Geoscience and Remote Sensing Letters 19:1–5

Dong X, Lin J, Lu S, et al (2022) Seismic Shot Gather Denoising by Using a Supervised-Deep-Learning Method with Weak Dependence on Real Noise Data: A Solution to the Lack of Real Noise Data. Surveys in Geophysics 43: 1363–1394

Dong X, Yu W, Lin J, et al (2026) Light-weighted foundation model for seismic data processing based on representative and non-redundant pre-training dataset. Surveys in Geophysics, https://doi.org/10.1007/s10712-026-09957-4

Dosovitskiy A, Beyer L, Kolesnikov A, et al (2020) An image is worth 16×16 words: Transformers for image recognition at scale. arXiv preprint arXiv:201011929

Durall R, Ghanim A, Ettrich N, et al (2024) An in-depth study of u-net for seismic data conditioning: Multiple removal by moveout discrimination. Geophysics 89(1):WA233–WA246

Gao J, Mao J, Man W, et al (2006) On the denoising method of prestack seismic data in wavelet domain. Chinese Journal of Geophysics (in Chinese) 49(4): 1155-1163

Gomez J, Velis D (2016) A simple method inspired by empirical mode decomposition for denoising seismic data. Geophysics 81(6):V403–V413

Gu A, Dao T (2024) Mamba: Linear-time sequence modeling with selective state spaces. arXiv preprint arXiv:231200752

Gulunay N (1986) Fxdecon and complex wiener prediction filter. In: SEG technical program expanded abstracts 1986. Society of Exploration Geophysicists, 279–281. Society of Exploration Geophysicists

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

Harsuko R, Alkhalifah T (2022) StorSeismic: A new paradigm in deep learning for seismic processing. IEEE Transactions on Geoscience and Remote Sensing 60:1–15

He K, Chen X, Xie S, et al (2022) Masked autoencoders are scalable vision learners. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 16000–16009

Hennenfent G, Herrmann F (2006) Seismic denoising with nonuniformly sampled curvelets. Computing in Science & Engineering 8(3):16–25

Hlebnikov V, Elboth T, Vinje V, et al (2021) Noise types and their attenuation in towed marine seismic: A tutorial. Geophysics 86(2):W1–W19

Huang N, Shen Z, Long S, et al (1998) The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society of London Series A: mathematical, physical and engineering sciences 454(1971):903–995

Huang Y, Xu J, Lai J, et al (2023) Advancing transformer architecture in long-context large language models: A comprehensive survey. arXiv preprint arXiv:231112351

Kaur H, Fomel S, Pham N (2020) Seismic ground-roll noise attenuation using deep learning. Geophysical Prospecting 68(7):2064–2077

Kirillov A, Mintun E, Ravi N, et al (2023) Segment anything. Proceedings of the IEEE/CVF International Conference on Computer Vision, 4015–4026

Kumar A (2024) Vision transformer based foundation model for oil reservoir forecasting. 85th Annual International Conference and Exhibition, EAGE, Extended Abstracts

Lellouch A, Lindsey N, Ellsworth W, et al (2020) Comparison between distributed acoustic sensing and geophones: Downhole microseismic monitoring of the forge geothermal experiment. Seismological Society of America 91(6):3256–3268

Li F, Liu H, Wang W, et al (2024a) Swin transformer for seismic denoising. IEEE Geoscience and Remote Sensing Letters 21:1–5

Li G, Li Y, Yang B (2017) Seismic exploration random noise on land: Modeling and application to noise suppression. IEEE Transactions on Geoscience and Remote Sensing 55(8):4668–4681

Li Y, Zhang M, Zhao Y, et al (2022) Distributed acoustic sensing vertical seismic profile data denoising based on multistage denoising network. IEEE Transactions on Geoscience and Remote Sensing 60:1–17

Li Y, Zhang H, Huang J, et al (2024b) Conditional denoising diffusion probabilistic model for ground-roll attenuation. IEEE Transactions on Geoscience and Remote Sensing 62:1–12

Li Z, Sun N, Gao H, et al (2021) Adaptive subtraction based on u-net for removing seismic multiples. IEEE Transactions on Geoscience and Remote Sensing 59(11):9796–9812

Li Z, Sun K, Zeng T, et al (2023) Unsupervised fista-net-based adaptive subtraction for seismic multiple removal. IEEE Transactions on Geoscience and Remote Sensing 61:1–9

Liang X, Li Y, Zhang C (2018) Noise suppression for microseismic data by non-subsampled shearlet transform based on singular value decomposition. Geophysical Prospecting 66(5): 894 - 903

Liu Q, Chen Z, Ma J (2025) Foundation models for exploration geophysics. Geophysics 90(6): H29-H52

Liu X, Hu T, Wang S, et al (2022) Seismic internal multiple suppression based on convolutional neural network. IEEE Geoscience and Remote Sensing Letters 19:1–5

Liu Z, Lin Y, Cao Y, et al (2021) Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF international conference on computer vision, 10012–10022

Lu S (2021) Migration using sea surface-related multiples: Challenges and opportunities. Geophysics 86(5):WC11–WC19

Naghizadeh M, Sacchi M (2012) Multicomponent f-x seismic random noise attenuation via vector autoregressive operators. Geophysics 77(2):V91–V99

Nesterova O, Barone I, Cassiani G, et al (2026) Fully integrated optical distributed acoustic sensing and seismic node acquisition for near-surface wave characterization in a complex urban archaeological site. Geophysics 91(4):B99–B112

Pham N, Li W (2022) Physics-constrained deep learning for ground roll attenuation. Geophysics 87(1):V15–V27

Qi Z, Li Z, Sun N, et al (2025) Semi-supervised interpretable fista-net for adaptive subtraction and removal of seismic multiples. IEEE Transactions on Geoscience and Remote Sensing 63: 1–14

Qian F, Guo W, Liu Z, et al (2022) Unsupervised erratic seismic noise attenuation with robust deep convolutional autoencoders. IEEE Transactions on Geoscience and Remote Sensing 60:1–16

Qian F, Hua H, Wen Y, et al (2024) Unsupervised 3-d seismic erratic noise attenuation with robust tensor deep learning. IEEE Transactions on Geoscience and Remote Sensing 62:1–16

Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention, Springer, pp 234–241. Springer

Saad O, Alkhalifah T (2026) Distributed acoustic sensing denoising using a self-supervised conditional diffusion model. Geophysical Prospecting 74(3):e70154

Saad O, Ravasi M, Alkhalifah T (2024) Noise attenuation in distributed acoustic sensing data using a guided unsupervised deep learning network. Geophysics 89(6):V573–V587

Sheng H, Wu X, Si X, et al (2025) Seismic foundation model: A next generation deep-learning model in geophysics. Geophysics 90(2):IM59–IM79

Sun H, Yang F, Ma J (2022) Seismic random noise attenuation via self-supervised transfer learning. IEEE geoscience and remote sensing letters 19:1–5

Tao L, Ren H, Ye Y, et al (2022) Seismic surface-related multiples suppression based on sagan. IEEE Geoscience and Remote Sensing Letters 19:1–5

Treitel S (1974) The complex wiener filter. Geophysics 39(2):169–173

Trickett S, Burroughs L, Milton A (2012) Robust rank-reduction filtering for erratic noise. In: SEG International Exposition and Annual Meeting, SEG, pp SEG–2012. Society of Exploration Geophysicists

Vaswani A, Shazeer N, Parmar N, et al (2017) Attention is all you need. Advances in neural information processing systems 30

Verschuur D (2013) Seismic multiple removal techniques: past, present and future. EAGE

Verschuur D, Staal X, Berkhout A (2016) Joint migration inversion: Simultaneous determination of velocity fields and depth images using all orders of scattering. The Leading Edge 35(12):1037–1046

Wang K, Hu T, Wang S, et al (2022a) Seismic multiple suppression based on a deep neural network method for marine data. Geophysics 87(4):V341–V365

Wang K, Hu T, Zhao B (2022b) An unsupervised deep neural network approach based on ensemble learning to suppress seismic surface-related multiples. IEEE Transactions on Geoscience and Remote Sensing 60:1–14

Wang K, Hu T, Zhao B (2023) Seismic internal multiple attenuation based on unsupervised deep learning with a local orthogonalization constraint. IEEE Transactions on Geoscience and Remote Sensing 61:1–9

Wang S, Song P, Tan J, et al (2022c) Attention-based neural network for erratic noise attenuation from seismic data with a shuffled noise training data generation strategy. IEEE Transactions on Geoscience and Remote Sensing 60:1–16

Wang Y, Lu W, Liu J, et al (2019) Random seismic noise attenuation based on data augmentation and cnn. Chinese Journal of Geophysics 62(1):421–433

Weglein AB (1999) Multiple attenuation: an overview of recent advances and the road ahead (1999). The Leading Edge 18(1):40–44. https://doi.org/10.1190/1.1438150

Xie P, Liu Y, Liu C, et al (2025) Simultaneous suppression of seismic random and erratic noise using pinn with high-frequency preservation. Journal of Geophysics and Engineering 22(6):1796–1808

Yang L, Fomel S, Wang S, et al (2023a) Denoising distributed acoustic sensing data using unsupervised deep learning. Geophysics 88(4):V317–V332

Yang L, Fomel S, Wang S, et al (2023b) Slknet: An attention-based deep-learning framework for downhole distributed acoustic sensing data denoising. Geophysics 88(6):WC69–WC89

Yang L, Wang S, Chen X, et al (2023c) Deep learning with fully convolutional and dense connection framework for ground roll attenuation. Surveys in Geophysics 44(6):1919–1952

You J, Sun R, Huang X, et al (2025) Deepimaging: Deep neural networks driven full waveform migration. IEEE Transactions on Geoscience and Remote Sensing 64:1–19

Yu S, Ma J (2017) Complex variational mode decomposition for slop-preserving denoising. IEEE transactions on geoscience and remote sensing 56(1):586–597

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

Zhang C, Baan M (2019) Strong random noise attenuation by shearlet transform and time-frequency peak filtering. Geophysics 84(6):1-49

Zhang D, Verschuur E (2026) Physics-guided deep learning for adaptive surface-related multiple subtraction. Geophysical Prospecting 74(4):e70180

Zhang H, Zhou Y, Zhou J, et al (2026) Revealing lakebed sedimentary history in yigong lake, tibetan plateau, using distributed acoustic sensing. Geophysics 91(4):B125–B137

Zhang M, Hu T, Duan W, et al (2025) Deep learning for efficient prestack seismic internal multiples suppression with global-local feature fusion and optuna optimization. Journal of Geophysics and Engineering gxaf142

Zhang Y, Wang B (2023) Unsupervised seismic random noise attenuation by a recursive deep image prior. Geophysics 88(6):V473–V485

Zhang Y, Zhang Y, Dong H, et al (2024) Stugan: An integrated swin transformer-based generative adversarial networks for seismic data reconstruction and denoising. IEEE Transactions on Geoscience and Remote Sensing 62:1–15

Zhao H, Zhou Y, Bai T, et al (2023) A u-net based multi-scale deformable convolution network for seismic random noise suppression. Remote Sensing 15(18):4569

Zhong T, Li Y, Wu N, et al (2015) Statistical properties of the random noise in seismic data. Journal of applied Geophysics 118:84–91

Zhong T, Cheng M, Dong X, et al (2022) Seismic random noise suppression by using deep residual u-net. Journal of Petroleum Science and Engineering 209:109901

Zhong T, Cheng M, Wang S, et al (2024) Joint-guided denoising network for erratic noise attenuation. IEEE Transactions on Geoscience and Remote Sensing 62:1–13