High-Fidelity Seismic Super-Resolution Using Prior-Informed Deep Learning With 3D Awareness

Jintao Li, Xinming Wu, *Associate Member, IEEE*, Xianwen Zhang, Xin Du, Xiaoming Sun, Bao Deng, and Guangyu Wang

***Abstract*—The limitations of seismic vertical resolution pose significant challenges for the identification of thin beds. Improving the vertical resolution of seismic data using deep learning methods often encounters challenges related to unrealistic outputs and limited generalization. To address these challenges, we propose a novel framework that improves the fidelity and generalization of seismic super-resolution. Our approach begins with the generation of realistic synthetic training data that aligns with the structural and amplitude characteristics of field surveys. We then introduce an enhanced 2D network with 3D awareness, which builds on the 2D Swin-Transformer and 3D convolution blocks to effectively capture 3D spatial features while maintaining computational efficiency. This network addresses the limitations of traditional 2D approaches by reducing stitching artifacts and improving spatial consistency. Finally, we develop a prior-informed fine-tuning strategy using field data without the need for labels, which incorporates a self-supervised data consistency loss and a spectral matching loss based on prior knowledge. This strategy ensures that the super-resolution results preserve the original low frequency information while yielding a spectral distribution as expected. Experiments on multiple field datasets demonstrate the robustness and generalization capability of our method, making it a practical solution for seismic resolution enhancement in diverse field datasets.**

***Index Terms*—Geophysical image processing, superresolution, selfsupervised learning, image enhancement.**

I. Introduction

IN MANY mature hydrocarbon-producing fields, the progression of exploration and development has shifted the focus from conventional structural traps to more intricate lithologic or stratigraphic traps [1]. This shift highlights the need for accurately identifying thin beds, as most thicker reservoirs have already been developed. Thin beds, often smaller in size, require higher seismic resolution for precise characterization.

The limitations of seismic vertical resolution, commonly considered to be between one-eighth and one-quarter of the wavelength, pose substantial challenges for the identification of thin beds [2]. These subtle seismic signatures are frequently masked by the tuning effect or lost due to insufficient resolution. A popular approach to addressing this issue is to improve the seismic resolution. These resolution enhancement methods primarily focus on widening the seismic data frequency bandwidth and increasing the dominant frequency.

Driven by recent advances in deep learning, seismic super-resolution has attracted increasing attention and demonstrated promising results. However, unlike natural image applications, acquiring large-scale, high-quality labels in seismic exploration is extremely difficult. Consequently, most methods rely on synthetic data for training and then apply the models to field data, achieving reasonable performance in some surveys.

Most existing approaches use synthetic datasets generated following Wu et al. [3], which were originally developed for high-level interpretation tasks. These datasets emphasize structural patterns but overlook the requirements of low-level tasks such as super-resolution, which depend heavily on local details and amplitude variations. Because such synthetic data do not capture the amplitude anomalies commonly observed in field seismic data, models trained on them often fail to generalize well, producing unrealistic results that resemble synthetic patterns rather than the complex, heterogeneous characteristics of field data.

Another challenge arises from the network architecture itself. Existing seismic super-resolution methods based on 2D networks often suffer from stitching artifacts when extended to 3D volumes, suggesting that full 3D spatial context is necessary for stable prediction. Swin-Transformer [4], [5] has recently demonstrated strong modeling capability in image super-resolution, and it is natural to consider using its expressive power for seismic applications. However, directly adopting its 3D extension is impractical because the attention mechanism requires memory that grows quadratically with the number of tokens, making it infeasible for large 3D seismic data.

Even with improved synthetic data and a more advanced network, a clear domain gap remains when models trained on

Received 23 January 2025; revised 28 June 2025, 13 September 2025, and 10 December 2025; accepted 9 February 2026. Date of publication 18 February 2026; date of current version 23 February 2026. This work was supported in part by the National Science Foundation of China under Grant 42374127 and in part by DeepEarth Probe and Mineral Resources Exploration-National Science and Technology Major Project under Grant 2024ZD1002100. The associate editor coordinating the review of this article and approving it for publication was Prof. Ying Fu. *(Corresponding author: Xinming Wu.)*

Jintao Li is with the State Key Laboratory of Precision Geodesy, School of Earth and Space Sciences, University of Science and Technology of China, Hefei 230026, China, also with the Mengcheng National Geophysical Observatory, University of Science and Technology of China, Mengcheng 233500, China, and also with the State Key Laboratory of Ocean Sensing & Ocean College, Zhejiang University, Zhoushan 316021, China (e-mail: lijintao@mail.ustc.edu.cn).

Xinming Wu, Xiaoming Sun, Bao Deng, and Guangyu Wang are with the State Key Laboratory of Precision Geodesy, School of Earth and Space Sciences, University of Science and Technology of China, Hefei 230026, China, and also with the Mengcheng National Geophysical Observatory, University of Science and Technology of China, Mengcheng 233500, China (e-mail: xinmwu@ustc.edu.cn; sunxiaoming0305@ustc.edu.cn; dbao@mail.ustc.edu.cn; wangguangyu@mail.ustc.edu.cn).

Xianwen Zhang and Xin Du are with CNOOC Research Institute Ltd., Beijing 100027, China (e-mail: zhangxw4@cnooc.com.cn; duxin7@cnooc.com.cn).

Digital Object Identifier 10.1109/TIP.2026.3663926

synthetic data are applied to field surveys. This gap often leads to outputs that are not fully aligned with the characteristics of real seismic data. Similar adaptation issues have also been observed in other imaging domains, where external–internal learning strategies are used to reduce the mismatch between training data and real observations [6]. These challenges indicate that additional geophysical prior information is required for the model to adapt reliably to field data.

In this work, we address the above limitations through a combination of improved data construction, enhanced network design, and prior-informed adaptation. We first generate synthetic data that more faithfully reflect the structural and amplitude characteristics of the target area, reducing the discrepancy between synthetic and field observations. Building on this, we develop an enhanced Swin-based architecture with 3D awareness that captures spatial context effectively while avoiding the memory cost of full 3D attention and the stitching artifacts common in purely 2D approaches. Most importantly, we introduce a prior-informed fine-tuning strategy that enables the model to adapt to field data without requiring labels. This strategy is guided by two prior-informed losses. The first preserves reliable structural information by enforcing low-frequency consistency between the output and the input. The second encourages the enhanced spectrum to follow a physically motivated broadened profile, leading to more realistic high-frequency details. Together, these losses enable the model to generate results that are structurally faithful and perceptually consistent with field seismic data. This prior-informed losses also improve cross-survey generalization, extending the applicability of the method to a wider range of geological settings. Experiments on multiple field datasets demonstrate that the proposed approach delivers robust, high-fidelity, and well-generalized super-resolution results across diverse geological settings.

In summary, our main contributions are summarized as follows:

- We propose a structure-aligned and amplitude-realistic synthetic data generation strategy tailored to the target survey, producing volumes that better reflect field-like structures and amplitudes, and effectively reducing the domain gap between synthetic and real data.
- We design a 2D Swin-Transformer architecture with 3D awareness, which improves spatial context modeling and effectively reduces stitching artifacts while remaining computationally efficient.
- We introduce a prior-informed, label-free fine-tuning strategy based on self-supervised data consistency and spectral matching losses, enabling the model to adapt to field data without requiring high-resolution labels and enhancing its generalization across surveys.

This paper proposes a deep learning framework for seismic super-resolution. Section II reviews related work. Section III introduces our method, which includes realistic synthetic data generation, an enhanced 2D Swin-Transformer with 3D awareness, and a prior-informed fine-tuning strategy. Section IV presents experiments on field datasets and discusses the effectiveness, generalization, and limitations of the proposed approach.

II. Related Work

A. Traditional Methods

Traditional methods for seismic resolution enhancement can be broadly categorized into deconvolution, absorption compensation, and spectral enhancement. Deconvolution techniques [7], [8], [9], [10], [11], [12] work by compressing the seismic wavelet to extract more detailed signals with wider frequency content from the original data.

Seismic waves propagating through viscoelastic media are subject to absorption effects, which significantly reduce seismic resolution by attenuating high-frequency components. To address the issue, absorption compensation methods, such as inverse Q filtering [13], [14], [15], [16], [17], are employed to restore the attenuated energy and correct phase distortions, and in turn enhance seismic resolution.

Spectral enhancement methods, including spectral balancing [18], [19], [20], [21] and spectral blueing [22], [23], [24], aim to improve the resolution of seismic data by optimizing its frequency characteristics. These methods typically involve transforming the seismic signal into an another domain (e.g., frequency domain via Fourier transform, time-frequency domain via S-Transform or wavelet transform, or other related domains), where the spectrum is adjusted through operations such as balancing, modulation, or selective amplification. After processing, the signal is transformed back to the original domain, resulting in enhanced resolution and improved interpretability of seismic data.

B. Deep-Learning for Seismic Super-Resolution

Existing deep leanrning methods for seismic super-resolution can be broadly divided into two categories: methods trained on field data and methods trained on synthetic data. Lu et al. [25] adopted a modern, well-imaged 3-D seismic dataset as the training labels for a generative adversarial network (GAN [26]). Liu et al. [27] proposed a weakly supervised approach to construct unpaired training samples based on high-resolution field datasets, which was then used to train a CycleGAN [28]. While methods trained on field data can capture field complexities, they often face challenges such as the scarcity of high-quality labels and difficulties in handling inputs with significant frequency variations. Weakly supervised methods, in particular, may struggle with the realism of generated data due to the lack of strict pairing between low- and high-resolution samples. Cheng et al. [29] proposed a self-supervised approach that treats resolution enhancement as a frequency extension task, using iterative refinement and multi-loss training to reconstruct high-frequency components.

To address the issue of limited labeled data, Li et al. [30] proposed a strategy that utilizes synthetic data for training, which has since inspired numerous follow-up studies and improvements [31], [32], [33]. Since synthetic data inevitably differ from field data, methods relying on synthetic data often face challenges in generalization. To mitigate this issue, Choi et al. [34] utilized a 1D U-Net framework that explicitly incorporates field data features to improve seismic resolution. Zhang et al. [35] leverage the domain adaptation scheme to bridge the gap between synthetic data and field data.

Recently, several approaches [36], [37] have been proposed to enhance seismic image resolution by leveraging the generative capabilities of diffusion models [38], [39]. Despite these advancements, a significant limitation of current synthetic data, typically generated using the method proposed by Wu et al. [3], is that they tend to exhibit overly smooth reflectors and uniform amplitudes. This discrepancy can result in overly smoothed outputs when applied to field data with significant amplitude variations, such as those containing channel features, potentially damaging the original signal.

C. Differences Between Natural Image and Seismic Super-Resolution

Compared to natural image super-resolution, seismic super-resolution is still in its early stages. The CV community has benefited from large-scale datasets, well-established evaluation metrics, and rapid advances in variety models [5]. Recent deep learning studies in remote sensing and hyperspectral image [40], [41], [42] also demonstrate progress in spatial–spectral modeling. In contrast, seismic data lacks abundant high-resolution labels, and synthetic data often oversimplifies field characteristics—particularly in amplitude variation—leading to poor generalization. Moreover, natural image task emphasizes visual rationality, while seismic applications require preserving subtle structural and amplitude features that are geologically meaningful.

Due to these differences, many successful methods in CV have not yet been fully utilized in the seismic domain. In particular, generative approaches have shown great potential in enhancing texture and spectral richness in images [43], [44], [45], [46], [47]. These capabilities are also highly relevant for seismic data, where reconstructing fine-scale geological features is crucial. We believe that introducing such generative modeling techniques—adapted to the specific characteristics of seismic data—can significantly improve the resolution and realism of seismic.

III. Methods

We first build a realistic synthetic seismic dataset that aligns well with the geological structural and amplitude response of a target survey. Then, an enhanced 2D swin-transformer network is designed to capture spatial features in 3D space. After training the network on the synthetic dataset, we propose a fine-tuning strategy to improve the realism of the super-resolution results based on field datasets.

A. Realistic Synthetic Training Dataset Generation

The synthetic dataset employed by current methods are predominantly generated using the approach proposed by Wu et al. [3]. However, this approach was originally designed for structural interpretation rather than seismic super-resolution. The synthetic data it generates often have uniform amplitudes and overly smooth reflectors, leading to networks that produce smoothed outputs and fail to preserve small-scale discontinuous features, such as channel-related anomalies in seismic profiles. To address this issue, we propose a workflow that generates realistic synthetic data with non-uniform amplitudes and ensures its structural and amplitude characteristics align well with the target seismic data, improving the realism of the training data.

A realistic impedance model is generated through a workflow, as shown in Figure 1. The first step involves multi-information constrained RGT estimation [48], which integrates seismic, interpreted faults (Figure 1a), slope attributes (Figure 1b), and horizons (Figure 1c) to accurately estimate the RGT (Figure 1e) for the target survey. Despite the effort required to obtain these data, they are typically available in mature survey areas. It is important to emphasize that while faults and horizons can significantly improve the accuracy of RGT estimation, they are not strictly required. In practice, seismic and slope attributes are sufficient to produce a reasonably accurate RGT, especially in less mature survey areas where structural interpretations may not be readily available.

After obtaining the RGT, a RGT-guided well interpolation method [49] is applied to generate the impedance model (Figure 1h) using the estimated RGT and well logs (Figure 1f). The method is also not highly sensitive to the number of available wells. In fact, even a single well can provide sufficient general trend information to guide the impedance model interpolation. In most industrial scenarios, at least one well log is typically available, making this approach widely applicable. Although the structural features align well with the target field data, the resulting impedance remains overly smooth using [49]. To mitigate this, we incorporate the Root Mean Square (RMS) attribute (Figure 1d) as a weighting factor, introducing spatial discontinuity and linking impedance values to seismic amplitude variations. Due to the high resolution and fine sampling interval of well logs, it is theoretically possible to interpolate an impedance model at any desired resolution. However, to ensure the stability of subsequent super-resolution experiments, we interpolate the impedance model at a sampling interval of 1 ms, which is half the sampling interval of the seismic data.

To further enhance amplitude heterogeneity and improve synthetic data realism, we insert channels (Figure 1g) generated by [50] into the impedance model (Figure 1h) along horizons, producing a more realistic result (Figure 1i). Since the channels generated by [50] are initially horizontal, we use the RGT to shift the channel masks, ensuring that each channel follows a consistent geological time horizon. After applying slight smoothing to the channel masks, we multiply them with the channel-free impedance model to obtain a final result with clearly defined channels. This workflow allows us to generate multiple impedance models with diverse channel distributions.

Based on the realistic impedance model, a reflectivity volume can be easily generated and further used to create synthetic seismic data. By convolving the reflectivity with different wavelets and applying filtering operations, we obtain synthetic data that closely resembles the target survey. Figure 2 illustrates an example of synthetic data, displaying (from left to right) high-resolution seismic data (1 ms sampling interval) used as labels, low-resolution seismic data with noise

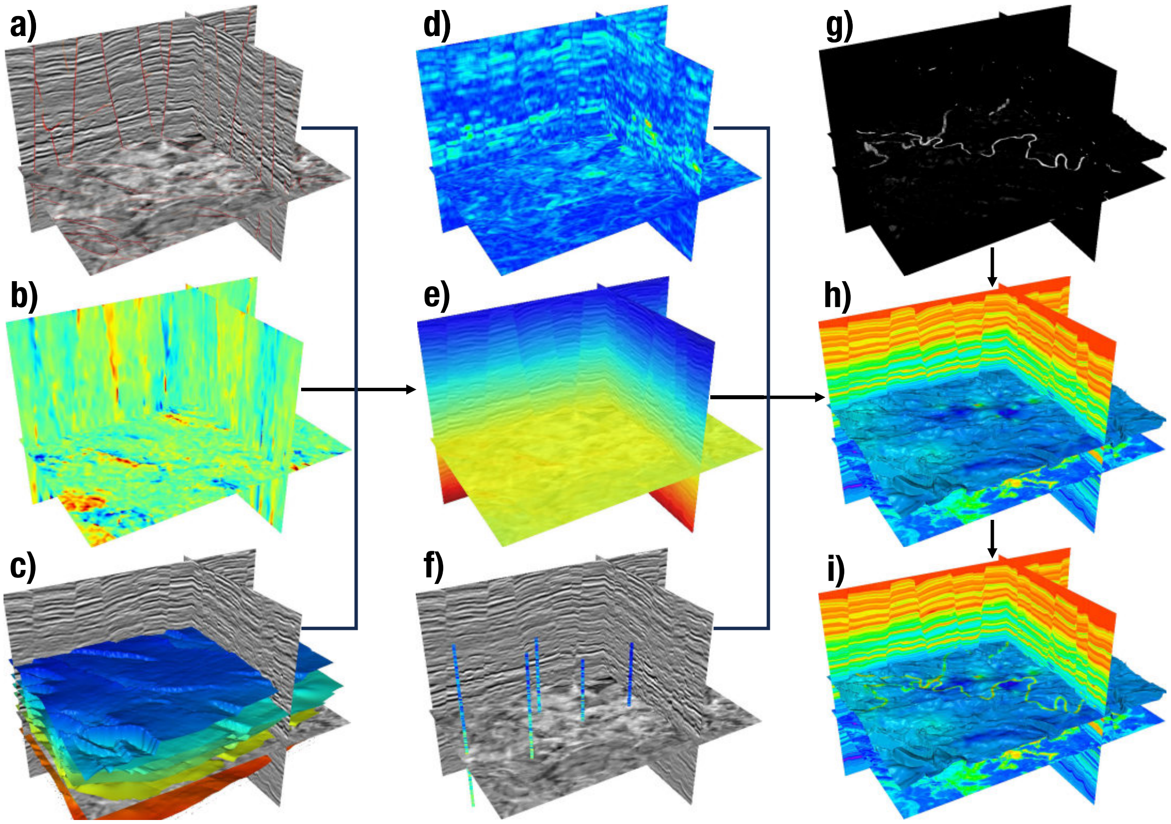

Fig. 1. The Workflow for generating a realistic impedance model. Interpreted faults (a), slope attributes (b), and horizons (c) are integrated to estimate RGT (e). Then, RGT and the RMS attribute (d) are used to guide well logs (f) interpolation, producing a heterogeneous impedance model (h). To further enhance realism, channels (g) are added along horizons, resulting in a more realistic impedance model (i).

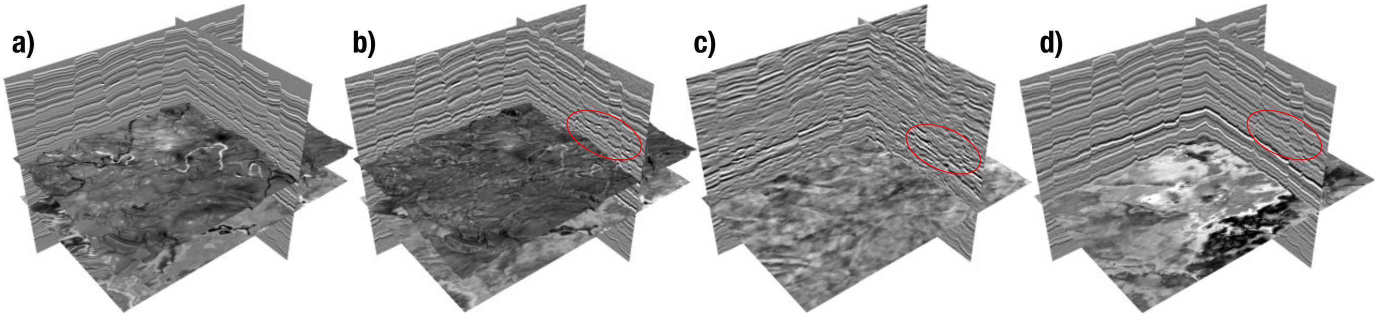

Fig. 2. The comparison of synthetic and field seismic data. a) the synthetic high resolution seismic data, b) the synthetic low resolution seismic data with field noise, c) the field seismic data, d) the low resolution synthetic data generated from Figure 1h (without channels). Our synthetic data exhibits structural consistency with the field data, while the channel features reduce the smoothness of reflection events, preserving small-scale discontinuities.

(2 ms sampling interval), field seismic data (2 ms sampling interval), and low-resolution data from the impedance model illustrated in Figure 1h (i.e., without channels, 2 ms sampling interval). The noise in the low-resolution data is derived from the field data through filtering. By incorporating well logs and interpreted structural information, the generated synthetic datasets align well with the geological structure and well log response characteristics of the target survey. Notably, our synthetic data exhibits amplitude anomalies similar to those observed in the field data, while the synthetic data without channels lacks such features. Although the reflection events in the labels appear slightly smoother compared to the field data, the inclusion of channels ensures that small discontinuous features are preserved.

To further validate the realism of our synthetic data, we provide quantitative and visual comparisons. Figure 3a shows the normalized amplitude distribution of different datasets. The distribution of our synthetic data most closely matches that of the field data. Figure 3b compares the frequency spectra, where the low-resolution synthetic data matches the field data well, confirming frequency consistency. We also compute two quantitative metrics to assess the similarity between synthetic and field data: the Fréchet Inception Distance (FID) [51] and the Kullback-Leibler (KL) divergence. FID is used to evaluate the overall difference between distributions, while KL divergence focuses on the similarity in amplitude distributions. As seismic-specific feature extractors for FID are not yet well established, we approximate FID by sampling 15,000 slices

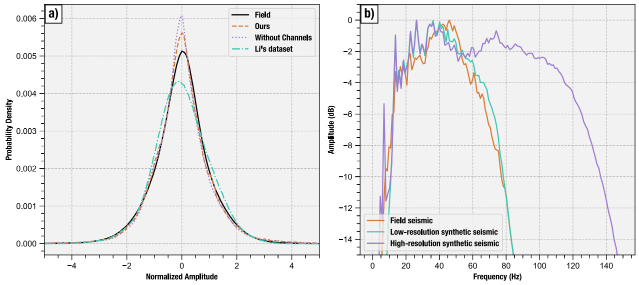

Fig. 3. Comparison of (a) normalized amplitude distributions and (b) frequency spectra among different datasets. The amplitude distribution of our synthetic data most closely matches that of the field data, while the frequency spectrum of the low-resolution synthetic data aligns well with that of the field data. The high-resolution synthetic data contains broader frequency content. This confirms that our synthetic data is well aligned with the field data.

TABLE I

QUANTITATIVE COMPARISON BETWEEN SYNTHETIC AND FIELD SEISMIC

Method	**FID Score**	**KL-Divergence**
Li's dataset [30]	244.51	0.0282
Ours (without Channels)	140.93	0.0119
Ours	**121.78**	**0.0087**

of size 256 × 256 from the 3D volumes and treating them as grayscale images. The results are summarized in Table I, where our method consistently achieves the best scores across both metrics among the compared datasets.

To enhance the generalization capability of the training dataset, we did not enforce strict alignment between the low-resolution synthetic data and the field data. Instead, we defined a range for the frequency band, allowing it to vary randomly within this range across different training samples. By generating channels with diverse morphologies, randomly adjusting the values of channel masks to modify the impedance of the channels, and varying parameters such as wavelets and cutoff frequencies for filtering, we were able to create a diverse set of training data. This diversified data generation strategy can improve the generalization ability of the training set.

B. Enhanced 2D Network With 3D Awareness

Most existed methods utilize CNN-based architectures (such as UNet and GANs). However, Swin-Transformer-based architectures have gained widespread adoption in computer vision super-resolution tasks and have been proven to outperform CNN-based approaches [5], [52]. Therefore, we adopt an architecture with a Swin-Transformer backbone, as shown in Figure 4. It primarily consists of three components: First, low-resolution data (with a size of $1 \times t \times w$) is processed using a simple convolution to extract shallow features. Then, a series of Swin-Transformer-based blocks are employed to extract deep features, which are combined with the shallow features through residual connections (resulting in a size of $c \times t \times w$). Finally, a reconstruction module is applied to generate the super-resolution result (with a size of $1 \times 2\ t \times w$), effectively enhancing the vertical resolution of the seismic data. In this work, we incorporated the Residual Hybrid Attention Groups (RHAG) designed by [5] as the fundamental block for deep feature extraction.

Unlike natural images, seismic data is not two-dimensional but three-dimensional, containing richer spatial information. While directly applying 2D networks to inline and crossline sections individually yields satisfactory results, assembling 2D slices into a 3D volume often introduces stitching artifacts along the assemble direction. These artifacts are not only present in 2D Swin-Transformer-based models but are even more pronounced in CNN-based models. A detailed analysis of these artifacts and their impact on the results will be discussed in the next section. However, extending Swin-Transformer to 3D is highly challenging due to the quadratic increase in memory requirements for the attention mechanism as the number of tokens grows.

To address this issue, we propose an enhanced 2D network with 3D awareness by combining existing 2D Swin-Transformer blocks with a small number of 3D convolutional modules, as illustrated in the bottom branch of Figure 5 for 3D input. Assuming b, c, t, x and i represent the dimension of batch, channels, time, crossline, and inline, respectively. The dimensions of the input and output for each block are shown above and below the block. Through a simple "permute" operation, we reshape the input to $(b\ \ i)\ \ c\ \ t\ \ x$, where $(b\ \ i)$ denotes merging the batch and inline dimensions. This reshaped input is then passed through a 2D Swin-Transformer block to extract features $\mathbf{z}_i$ along the inline slices. Then, a 3D convolutional block is applied to capture spatial features, and the data is reoriented to $(b\ \ x)\ \ c\ \ t\ \ i$ for input into another 2D Swin-Transformer block, which extracts crossline features $\mathbf{z}_j$ and reshapes them to match the dimensions of $\mathbf{z}_i$. Finally, a learnable parameter α is used to fuse the inline features $\mathbf{z}_i$ and crossline features $\mathbf{z}_j$ through:

$$\alpha \mathbf{z}_i + (1 - \alpha)\mathbf{z}_j \tag{1}$$

This approach allows us to leverage 2D Swin-Transformer blocks while effectively extracting spatial features, and avoid stitching artifacts in the final output.

To train this enhanced network on synthetic training dataset, we employed several loss functions, including L1 loss, the Learned Perceptual Image Patch Similarity (LPIPS) loss [53], and GAN loss. The L1 Loss measures the absolute difference between the predicted output $\hat{\mathbf{y}}$ and the ground truth $\mathbf{y}$ to ensure precise reconstruction and structural fidelity:

$$\mathcal{L}_{\text{L1}} = \frac{1}{N} \sum_{i=1}^{N} |\hat{y}_i - y_i| \tag{2}$$

where $\hat{y}_i$ and y_i are the i-th pixel values of the predicted output and ground truth, respectively, and N is the total number of pixels. The LPIPS loss leverages deep features from a pre-trained network to evaluate perceptual similarity, enhancing the visual quality and fine details of the output:

$$\mathcal{L}_{\text{LPIPS}} = \sum_{l} \|\phi_l(\hat{\mathbf{y}}) - \phi_l(\mathbf{y})\|_2^2 \tag{3}$$

where $\phi_l(\hat{\mathbf{y}})$ and $\phi_l(\mathbf{y})$ are the deep features of $\hat{\mathbf{y}}$ and $\mathbf{y}$ extracted from the l-th layer of a pre-trained network,

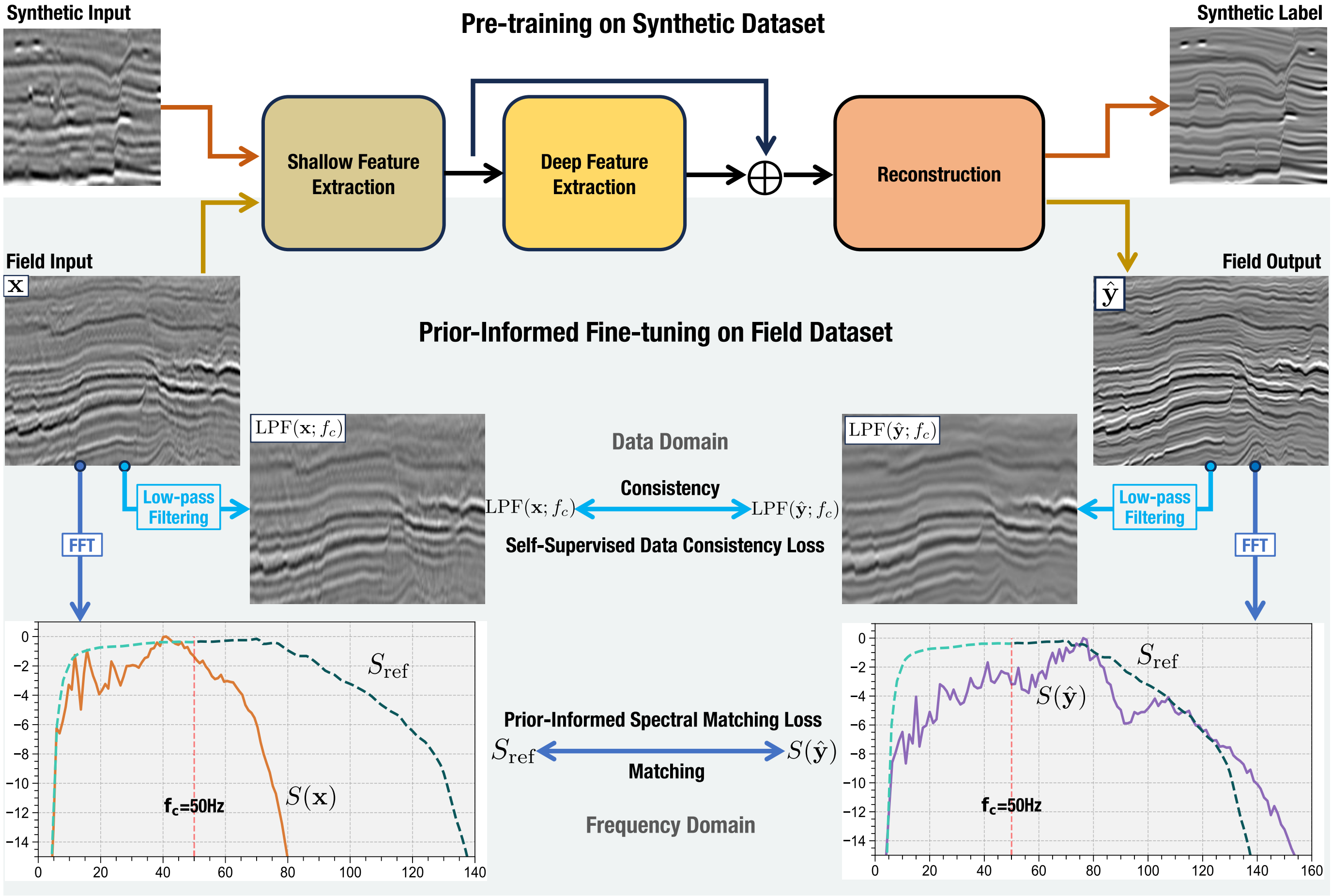

Fig. 4. The proposed network undergoes two training stages: pre-training on synthetic data and prior-informed fine-tuning on field data. During pre-training, the network learns to enhance resolution using synthetic data. In fine-tuning, the network adapts to field data without labels through two prior-informed losses: the Self-supervised Data Consistency Loss, which ensures the low-frequency component of the prediction $\mathrm{LPF}(\hat{\mathbf{y}}; f_c)$ matches that of the input data $\mathrm{LPF}(\mathbf{x}; f_c)$, and the Prior-Informed Spectral Matching Loss, which aligns the spectrum of the output $S(\hat{\mathbf{y}})$ with an expected spectrum S_{ref} derived from the input data $S(\mathbf{x})$.

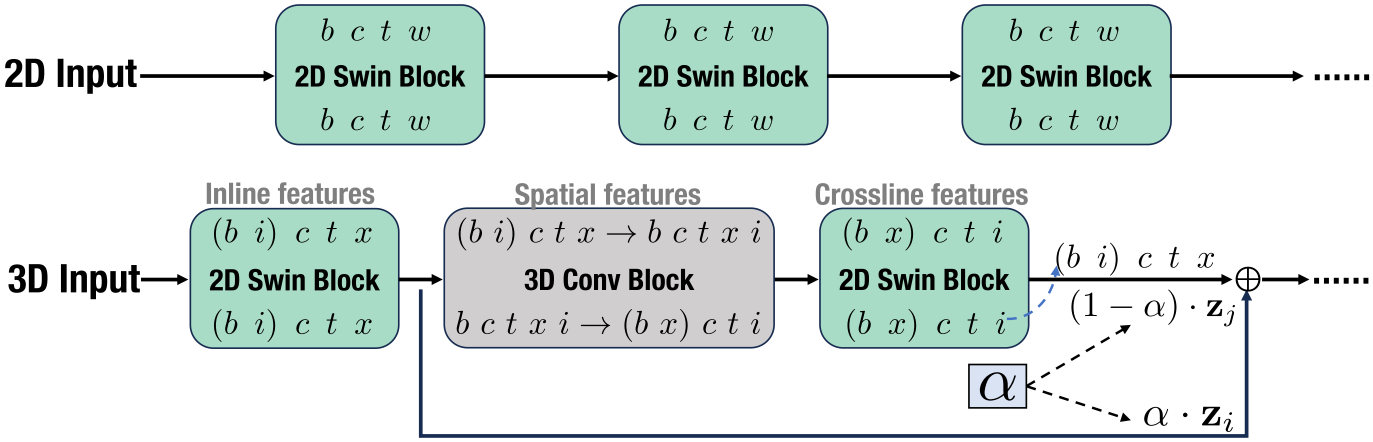

Fig. 5. To extract 3D spatial features instead of 2D sectional features, we extend the 2D network to a 3D network based on the 2D swin-transformer and 3D convolution blocks. Each block in the figure shows the input dimensions above and the output dimensions below, where b, c, t, x, and i represent batch, channel, time, xline, and inline dimensions, respectively. We first extract features $\mathbf{z}_i$ along the inline direction ($t-x$ slices), by a 3D convolution to capture spatial features. The data is then transposed to extract features $\mathbf{z}_j$ along the crossline direction ($t-i$ slices). Finally, a learnable parameter α is used to fuse the two feature representations. We refer to the model with the 2D input branch as *2D HAT* and the model with the 3D input branch as *Ours*.

and $\|\cdot\|_2$ denotes the L2 norm. The GAN loss, implemented through a discriminator network, encourages the generation of realistic and sharp results by distinguishing between real and synthetic data. The generator loss is defined as: The generator loss is defined as:

$$\mathcal{L}_{\mathrm{GAN}} = \mathrm{BCE}\big(D(\hat{\mathbf{y}}),\,1\big), \tag{4}$$

Here, $D(\hat{\mathbf{y}})$ is the discriminator's output for the predicted output $\hat{\mathbf{y}}$, and N is the total number of samples. The discriminator loss is defined as:

$$\mathcal{L}_{\mathrm{D}} = \mathrm{BCE}\big(D(\mathbf{y}), 1\big) + \mathrm{BCE}\big(D(\hat{\mathbf{y}}), 0\big), \tag{5}$$

where $D(\mathbf{y})$ is the discriminator's output for the ground truth image $\mathbf{y}$.

C. Prior-Informed Fine-Tuning With Field Data

While our synthetic data is of high quality, it still suffers from gaps compared to field data. These gaps may lead to overly idealized artifacts in some areas of the results, where the output resembles the synthetic data but lacks realism. Additionally, when applied to field datasets other than the target survey, especially in complex regions with significant differences from the training data or with varying frequency components (e.g., between the top and base, where frequencies differ significantly), substantial artifacts may arise. These artifacts not only reduce the realism of the results but may also damage the original structural and amplitude information in the seismic data. These limitations hinder the broader application of deep learning-based seismic data resolution enhancement methods.

To address these issues, we propose a prior-informed fine-tuning approach using field data to enhance the realism of the super-resolution results while effectively preserving the original information. This method is not limited to the previously mentioned model and can be applied to fine-tune any super-resolution network, ensuring both improved realism and fidelity to the input data. This method does not require labeled data and is guided by two prior-informed losses: the **S**elf-**S**upervised **D**ata **C**onsistency (**SSDC**) loss in the data domain and the **P**rior-**I**nformed **S**pectral **M**atching (**PISM**) loss in the frequency domain, as shown in Figure 4.

In the data domain, the SSDC loss ensures that the network's output share the same low-frequency components with the input. This guarantees that the low-frequency information of the input is preserved and not compromised in the output. The formula for this loss is as follows:

$$\mathcal{L}_{\mathrm{low}} = \frac{1}{N}\sum_{i=1}^{N}(\mathrm{LPF}(\mathbf{x}; f_c)_i - \mathrm{LPF}(\hat{\mathbf{y}}; f_c)_i)^2 \tag{6}$$

where $\mathrm{LPF}(\cdot; f_c)$ denotes the **l**ow-**p**ass **f**iltering operation with a cutoff frequency f_c, $\mathbf{x}$ is the input data, $\hat{\mathbf{y}}$ is the output data from the network, $\mathrm{LPF}(\mathbf{x}; f_c)$ and $\mathrm{LPF}(\hat{\mathbf{y}}; f_c)$ are the low-pass filtered data of $\mathbf{x}$ and $\hat{\mathbf{y}}$, respectively, and N is the total number of data samples.

While the SSDC loss ensures that the low-frequency components of the output remain consistent with the input, relying solely on this loss can lead to the gradual decay of high-frequency information. Over time, this may cause the model to collapse, resulting in outputs that are identical to the inputs and losing the ability to enhance resolution. To address this issue, we introduce a frequency-domain constraint as the PISM loss. Based on the spectrum of the input data $S(\mathbf{x})$, we design an expected broadened spectrum S_{ref} as the reference spectrum. This constraint aims to match the spectrum of the output data $S(\hat{\mathbf{y}})$ to the reference spectrum in the frequency domain, ensuring that the super-resolution results incorporate realistic high-frequency information. The formula for this loss is as follows:

$$\mathcal{L}_{\mathrm{spectrum}} = \frac{1}{M}\sum_{j=1}^{M}\big(S(\hat{\mathbf{y}})_j - S_{\mathrm{ref},j}\big)^2 \tag{7}$$

where M is the number of frequency points in the spectrum. The matching process can be applied to either the full spectrum or a specific portion of the spectrum that requires adjustment, allowing the model to focus on enhancing particular frequency ranges.

IV. Experiments And Discussion

In this section, we first present detailed descriptions of the dataset, model architecture, and training setup, followed by evaluations conducted on both synthetic and field data. To more comprehensively demonstrate the effectiveness of our approach, we quantitatively compare three key aspects: the model's performance on synthetic data, the fidelity of the outputs on field data, and the extent of frequency enhancement achieved on field data.

A. Implementation Details

To validate our method, we conducted training and testing with the following details:

Synthetic Training Dataset. To estimate the RGT, we used six horizons, and faults were estimated using FaultSeg3D plus [54]. Subsequently, we employed RGT and RMS-guided well log interpolation to generate an impedance model with dimensions (inline, crossline, time) = (401, 601, 878) and a temporal sampling interval of 1 ms. We then randomly generated 60 channel masks with varying numbers of channels to create impedance models containing channels, from which synthetic data were constructed. From each synthetic dataset, we extracted 30 labeled volumes of size 32 × 64 × 720, resulting in a total of 1,800 training samples. During training, each sample was randomly cropped into sub-volumes of size 32 × 64 × 128 (i.e., $i \times x \times t$) for input to the network. For 2D testing experiments, the label size was set to 64 × 128. Since seismic super-resolution is a low-level task that does not rely on large-scale global context, smaller training sample sizes are typically sufficient. This practice is also common in natural image super-resolution tasks. Moreover, given the use of the Swin Transformer architecture, increasing the sample size significantly raises GPU memory consumption, especially for 3D data. The RGT estimation, well-log-guided impedance interpolation, and subsequent synthetic data generation can be completed within half an hour on a 64-core Intel Xeon server.

Network Details. Our network is based on the HAT architecture [5], with two key modifications: (1) the deep feature extraction module is adapted (Figure 5) to capture 3D features, and (2) PixelShuffle module is modified to perform upsampling only along the time axis. The model input has a shape of, and the output has a shape of, indicating a 2× super-resolution along the time dimension. During training, the batch size is set to 1. The input is formatted as a 4D tensor with shape $((b\ i), 1, t, x)$, and the outputs with shape $((b\ i), 1, 2t, x)$.

Training Setup. Before feeding to the network, we normalized the input data using Z-score standardization, transforming each feature to have zero mean and unit variance. The model was trained for 150 epochs using the Adam [55] optimizer with an initial learning rate of 2×10^{-4}. We applied a simple learning rate decay strategy, reducing the learning rate by a factor of 0.5 at specific epochs during training. For the LPIPS loss, AlexNet [56] was used as the feature extractor. A simple UNet [57] was employed as the discriminator to compute the GAN loss. To ensure the model captures structural consistency from both inline and crossline perspectives, we compute the LPIPS and GAN losses using both $((b\ \ i), 1, 2t, x)$ and its transposed form $((b\ \ x), 1, 2t, i)$, and average the results. This dual-view strategy improves the generalization of the model without introducing additional network complexity. Aside from these adjustments, the overall network structure adheres to standard implementation practices used in natural image super-resolution and does not involve any non-conventional components.

Fine-Tuning Setup. During fine-tuning, the input data size was set to (32 × 32 × 128), with the output size being (32 × 32 × 256). To avoid large frequency intervals when computing the spectrum, the data were padded to a length of 512. Only the SSDC Loss and PISM Loss were used, without the losses from the training stage. The pre-training took 16 hours, and the fine-tuning took 28 minutes on a single NVIDIA H20 GPU.

Selected Models for Comparison. Due to the limited availability of open-source models in the geophysical field, we compared our method with two representative approaches: an open-source model [30] and a recently published diffusion-based method [36]. The code and model weights for the diffusion-based method were obtained by contacting the authors. For brevity, we refer to these methods as Li2021 and Xiao2024, respectively, in the following discussion. In addition, we included comparisons with several widely used general-purpose image super-resolution models, including EDSR [58], RCAN [43], SwinIR [52], HAT [5], and ATD [59], to further validate the effectiveness of our method beyond the seismic domain.

B. Comparsion and Results on the Target Survey

We first evaluate various methods on the synthetic test set, which consists of 200 samples. The PSNR and SSIM scores for each method are reported in the first column of Table II. To enable a fair comparison with Li2021 and Xiao2024, we retrain both models on our dataset, and denote the results as Li2021-retrained and Xiao2024-retrained, respectively. As shown in Table II, our method significantly outperforms Li2021, Xiao2024, and the 2D HAT in both PSNR and SSIM. Moreover, the 2D HAT model, which is based on a Swin Transformer, achieves better performance than the two U-Net-based approaches. After retraining, both Li2021 and Xiao2024 show improvements over their original pretrained versions.

Then, we applied the trained models to this target field data to evaluate the effectiveness of our method. Figure 6 shows the results of applying the model to the entire 3D dataset: Figure 6a displays the original data; Figure 6b shows the results of the 2D model (without fine-tuning), where slices were processed individually and then aggregated along the assemble direction; Figure 6c presents the results of the 3D model (without fine-tuning); and Figure 6d shows the results of the 3D model after prior-informed fine-tuning using field data. From the vertical sections, it is evident that both the 2D and 3D models significantly improve the vertical resolution. In the time slices, the super-resolution results enhance the clarity of structural features, such as the channels indicated by the red arrows, which become more continuous and well-defined. However, the results without fine-tuning appear overly clean and idealized, lacking realism. After prior-informed fine-tuning, the super-resolution results exhibit greater authenticity, closely resembling the characteristics of field data rather than synthetic data, such as more natural textures and realistic amplitude variations.

TABLE II
PERFORMANCE COMPARISON ACROSS METHODS

Method	Synthetic PSNR/SSIM	Fidelity PSNR/SSIM	Spectral Extension Hz
Li2021	22.48 / 0.5641	27.07 / 0.8901	15.90
Li2021-retrained	24.66 / 0.6138	29.51 / 0.9032	20.12
Xiao2024	21.80 / 0.3250	27.28 / 0.8442	3.91
Xiao2024-retrained	25.01 / 0.6854	29.74 / 0.8913	18.24
EDSR	23.27 / 0.6013	29.38 / 0.8932	18.12
RCAN	25.96 / 0.7321	28.68 / 0.8918	23.09
SwinIR	29.57 / 0.8459	31.98 / 0.9276	21.86
2D HAT	30.19 / 0.8416	33.48 / 0.9358	20.65
ATD	30.38 / 0.8532	30.475 / 0.9109	30.65
Our pretrain	**34.98 / 0.9523**	34.15 / 0.9179	30.96
Ours	–	**36.51 / 0.9724**	**38.18**

We extracted an inline slice and evaluated it alongside other methods, including their frequency spectra, as illustrated in Figure 7. It is worth noting that for Figure 7e and 7f, we utilized the source code and model weights of Li2021 and the diffusion-based method Xiao2024, respectively. These models were trained on synthetic data generated using the approach described in [3]. While this comparison may not be entirely equitable, the development of more realistic datasets remains one of our key contributions. In contrast to our method, both alternative approaches yield overly smooth and synthetic-like results when encountering data with rapid amplitude variations and discontinuous reflectors. Numerous small-scale discontinuous features (e.g., small channels) are erased, significantly compromising the original information. Specifically, Figure 7e displays pronounced artifacts. On the other hand, our approach successfully preserves these discontinuous features, retains the valid signals in the original data, and enhances the resolution of the seismic data. Relative to the results without fine-tuning (Figure 7b), the fine-tuned outputs demonstrate more natural textures and realistic amplitude variations.

To more fairly and quantitatively evaluate the fidelity of each method—that is, the ability to preserve meaningful signals and avoid damaging fine-scale features such as small

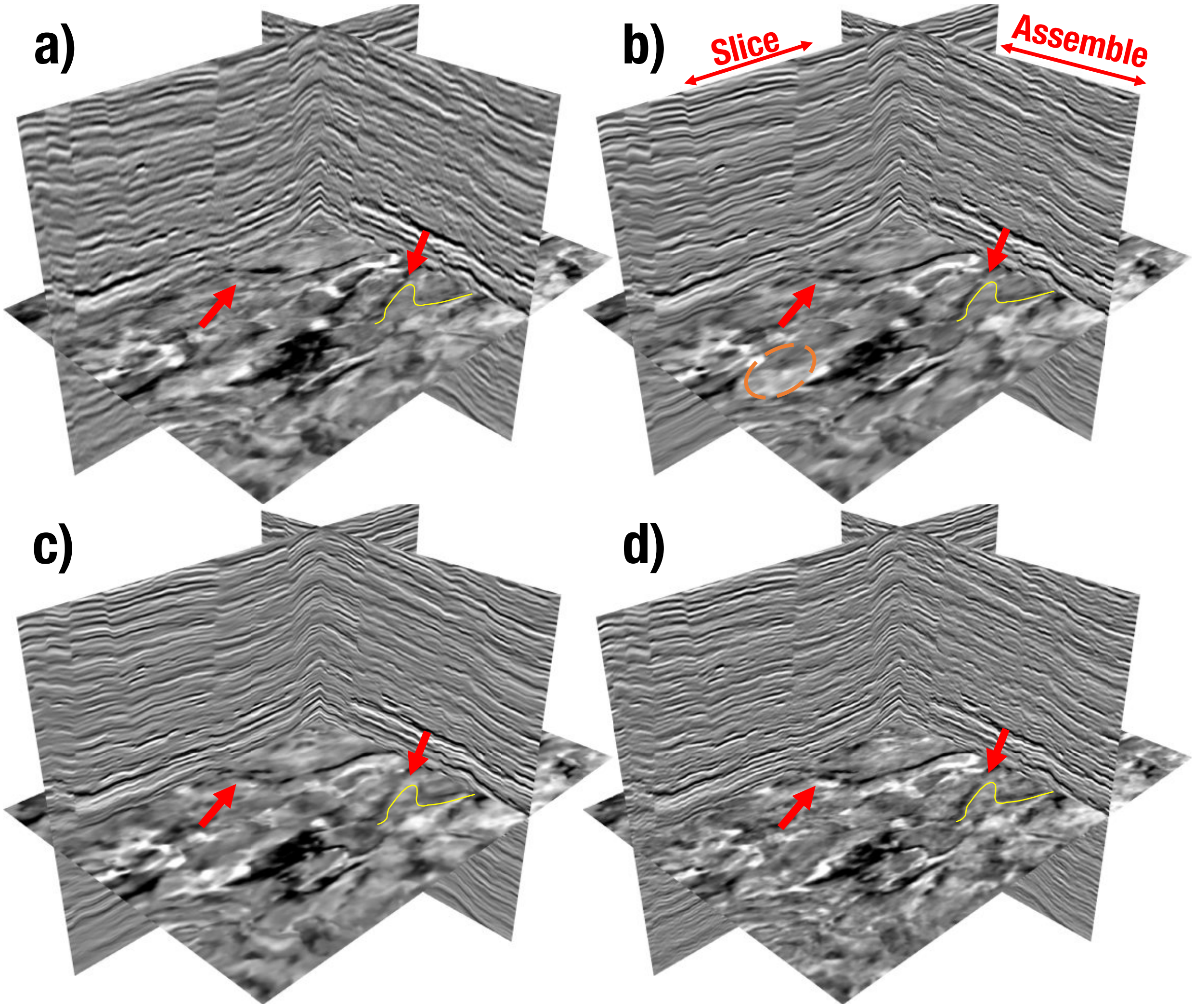

Fig. 6. Results of applying the model to the entire 3D dataset. (a) Original data; (b) 2D HAT model results (no fine-tuning), where inline slices are processed individually and aggregated along the assemble direction, leading to noticeable stitching artifacts; (c) Enhanced 2D model (3D input) results (no fine-tuning); (d) Enhanced 2D model results after fine-tuning. Both 2D and enhanced 2D models improve vertical resolution, while fine-tuning enhances realism.

channels—we report fidelity-related metrics in the second column of Table II. For this, we apply a low-pass filter to the super-resolution results using the frequency spectrum of the input data. This removes the newly generated high-frequency components. We then compute the PSNR and SSIM between the filtered result and the raw data. These metrics effectively indicate how well each method preserves the original seismic information. As shown in Table II, our method achieves the best fidelity, especially after being fine-tuned with the SSDC loss. Even without fine-tuning, the pretrained version still performs better than all the baselines. After retraining on our more realistic dataset, both Li2021 and Xiao2024 show noticeable improvements in fidelity compared to their original versions.

The frequency spectrum analysis (Figure 7d) further supports the above conclusions. In the low-frequency range, the spectral curves of Figure 7e and 7f (dashed lines) diverge markedly from those of the raw data, indicating a loss of original information. In contrast, our results, whether fine-tuned or not, exhibit strong alignment with the raw data in the low-frequency range, underscoring the ability to preserve the original information effectively. Although the spectrum of Figure 7b is broader than that of the raw data, it shows an anomalous dip in the mid-frequency range. After fine-tuning, the spectrum of prediction aligns more closely with the ideal spectrum we designed based on prior knowledge, filling in the dip and producing a more reasonable frequency distribution.

To evaluate the frequency enhancement capability of each method, we use the–5 dB point in the amplitude spectrum as a reference threshold. We then measure how much each method extends the high-frequency cutoff compared to the raw data. This provides a quantitative assessment of the model's ability to enhance resolution in the frequency domain. The results are reported in the last column of Table II. For example, the raw data has a–5 dB cutoff at 65.55 Hz, while our method extends it by 35.94 Hz, reaching up to 101.49 Hz. Among all compared methods, our approach achieves the largest frequency extension, demonstrating its superior performance in high-frequency reconstruction.

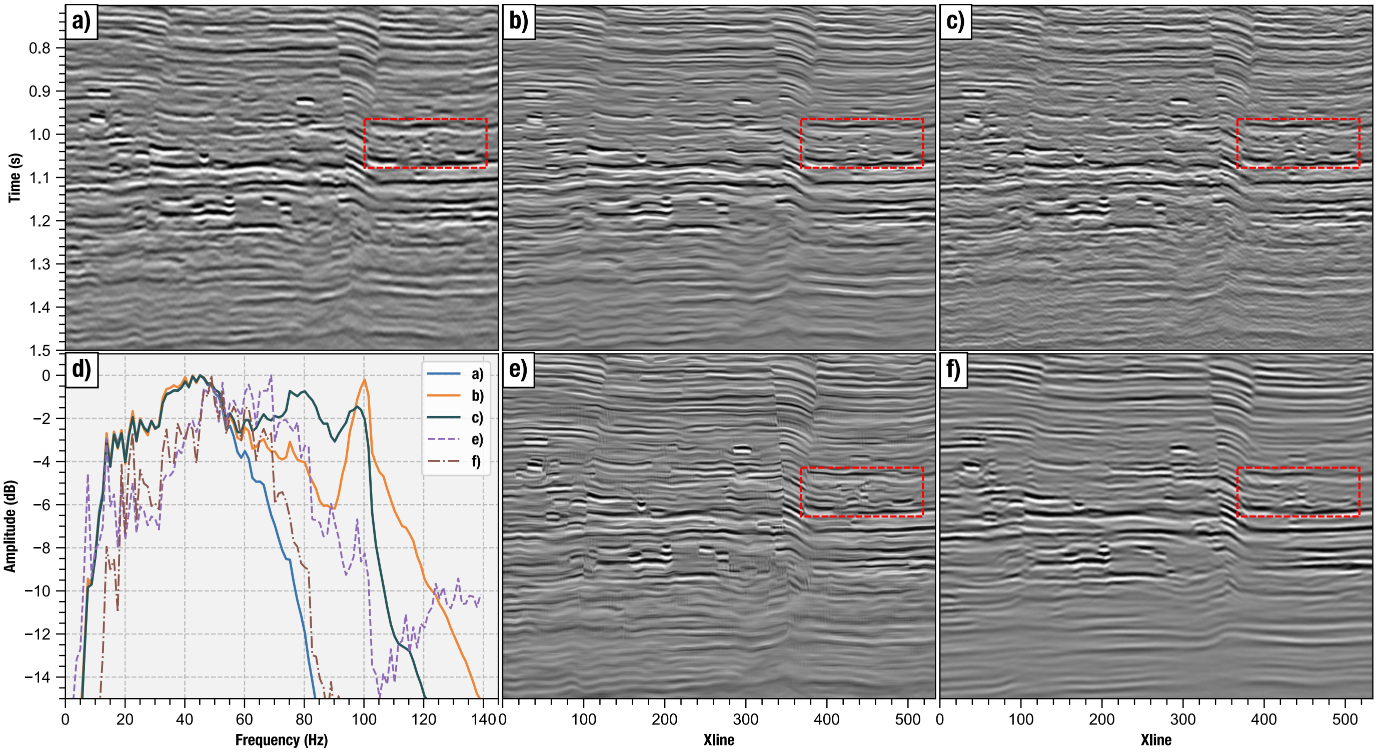

Fig. 7. Comparison of super-resolution results and frequency spectra. (a) The low-resolution data; (b) Our method without fine-tuning; (c) Our method after fine-tuning; (d) Frequency spectra comparison; (e) Results from [30]; (f) Results from the diffusion-based method [36].

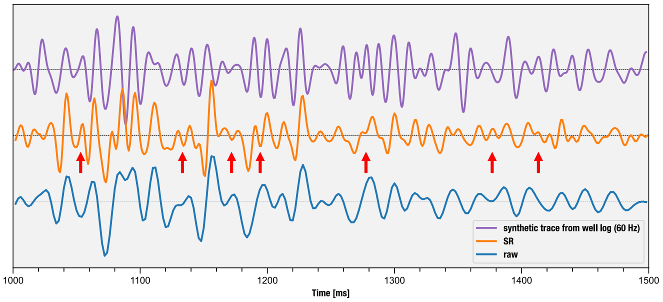

Fig. 8. Comparison of seismic traces at a blind well. Purple: synthetic trace from well log using a 60 Hz Ricker wavelet; orange: our fine-tuned result; blue: raw field data. The super-resolution result matches raw data amplitudes and reveals thin layers (red arrows) aligned with the well log but absent in raw data.

Lastly, we extracted traces from a blind well location (not used in training dataset generation) and analyzed them against well log data. For reference, we convolved the well log's reflectivity curve with a 60 Hz Ricker wavelet to generate a synthetic seismic trace (purple curve in Figure 8). The orange curve corresponds to Figure 6d (our fine-tuned result), while the blue curve represents the raw field data. The amplitude variations in our super-resolution result closely mirror those of the raw data, as our method effectively retains the low-frequency components. However, the super-resolution result reveals numerous thin layers that are absent in the raw data but align well with the synthetic trace derived from the well log, particularly at the locations marked by the red arrows. These thin layers, though missing in the raw data, are clearly visible in both the super-resolution result and the well log.

C. Comparison of Efficiency

We compare the efficiency of our method against other models in terms of several key metrics, including computational complexity (FLOPs), number of parameters, inference time, GPU memory consumption, and the presence of stitching artifacts when using patch-wise prediction. All tests were conducted on an NVIDIA H20 GPU. It is important to note that Xiao2024, as a diffusion-based method, has relatively low cost per denoising step. However, due to its iterative nature, the total inference time is significantly longer. Moreover, because of the stochastic nature of diffusion, it produces the most noticeable stitching artifacts when applied in patches. As shown in Table III, although our method is not the most efficient in terms of inference time or memory usage, it introduces nearly invisible stitching artifacts. This makes it well-suited for block-wise prediction without introducing noticeable defects. In seismic exploration, data processing is not a real-time task. The entire workflow—from raw data acquisition to final imaging—often takes weeks or even months. In this context, the inference time of our method is already very fast, fully acceptable, and not a significant limitation.

D. Ablation Study

1) Training Data With Channels Prevents Overly Smooth Results: As shown in Figure 7e and f, if the training data consist solely of amplitude-balanced data with smooth reflectors, the model tends to erase small discontinuous features, resulting in overly smooth outputs. This is particularly evident in the regions marked by the red dashed boxes. However, since these

TABLE III
COMPUTATIONAL ANALYSIS (INPUT: 64 × 256 × 256)

Method	FLOPs	Params	Time	VRAM	Tiling
	(T)	(M)	(s)	(GB)	Artifacts
Li2021	9.5	39.8	0.87	13.8	Severe
Li2021-retrained	9.5	39.8	0.87	13.8	Visible
Xiao2024*	18.2×100	105.4	0.65×100	18.7	Severe
Xiao2024-retrained*	18.2×100	105.4	0.65×100	18.7	Severe
EDSR	5.4	1.3	0.53	6.2	Visible
RCAN	63.9	15.4	3.5	6.1	Visible
SwinIR	48.9	11.6	4.9	21.4	Visible
2D HAT	85.6	20.3	7.7	81.6	Visible
ATD	76.4	19.7	18.5	65.1	Visible
Ours	111.2	13.1	8.9	86.7	Invisible

Tested on H20 GPU, and Xiao2024 is a diffusion model (at least 100 steps denoise). FLOPs shown as per-step×total steps.

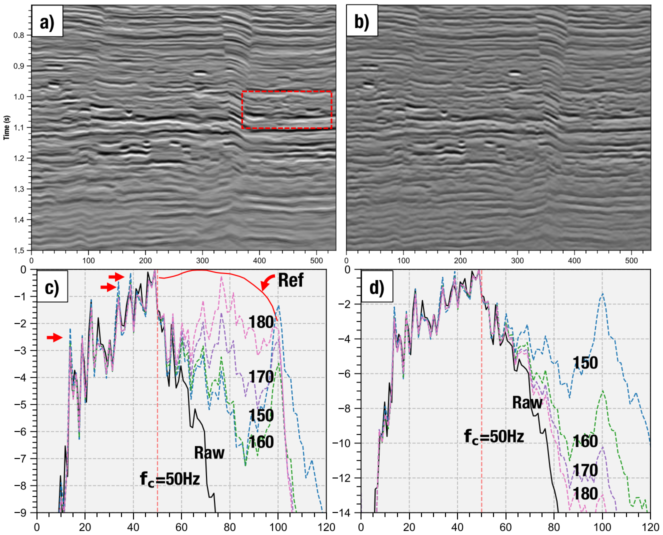

Fig. 9. Ablation study. (a) Effect of training data without channels; (b) Result without pretraining; (c) Frequency spectra during fine-tuning; (d) Frequency spectra during fine-tuning without Prior-Informed Spectral Matching loss.

models were not trained on data constructed from the target survey, we conducted a more equitable comparison by training the network on synthetic data without channel features. The results are shown in Figure 9a. Compared to Figure 7b, the results from training data without channels fail to preserve small-scale channel features.

2) Enhanced 2D Network Eliminates Stitching Artifacts: Even when following the approach of [27], which involves extracting training data from both inline and crossline directions, stitching artifacts still appear along the assemble direction when applied to field data, as shown in Figure 6b (produced by the 2D input branch in Figure 4). The result exhibits distinct characteristics in the slice and assemble directions: the slice direction appears clean and clear, while the assemble direction shows noticeable stitching artifacts, manifesting as fine linear traces in the time section (indicated by the orange dashed circle). In contrast, our enhanced

TABLE IV
ABLATION STUDY

Method	Fidelity (PSNR/SSIM)	Spectral Extension (Hz)
Without Channels	32.15 / 0.9266	7.1
Without Pretraining	22.89 / 0.6350	5.6
Epoch 150–180 (Fine-tuning)		
Epoch 150	33.52 / 0.9417	23.14
Epoch 160	35.27 / 0.9691	30.61
Epoch 170	35.99 / 0.9753	32.95
Epoch 180	36.11 / 0.9728	38.10
Epoch 150–180 (Without PISM Loss)		
Epoch 150	34.48 / 0.9506	23.10
Epoch 160	36.07 / 0.9699	2.93
Epoch 170	38.14 / 0.9794	0.0
Epoch 180	39.18 / 0.9822	0.0

The first block corresponds to Figures 9a-b, the second to Figure 9c, and the third to Figure 9d.

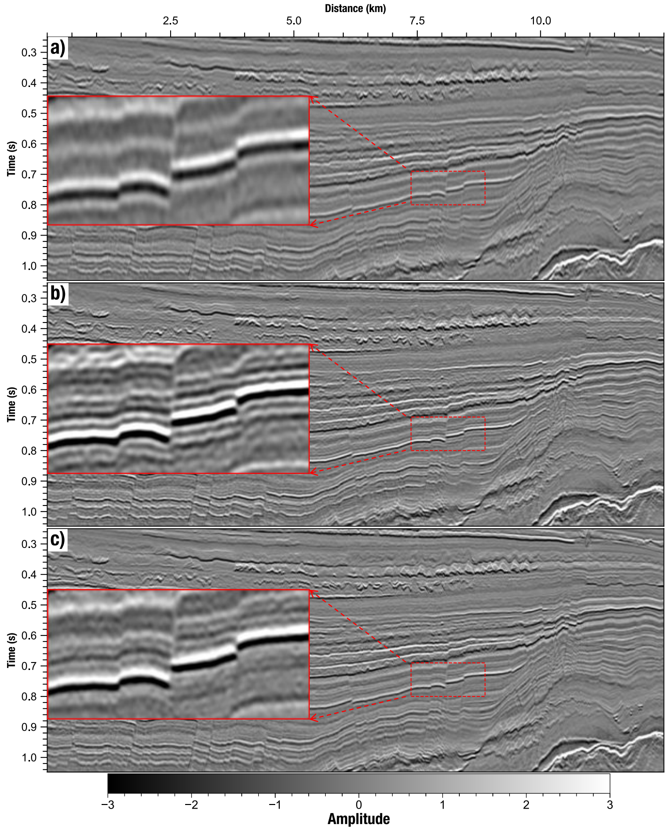

Fig. 10. Super-resolution results on a broad-frequency dataset. (a) Low-resolution input, generated by applying a low-pass filter to the high-resolution data; (b) Fine-tuned super-resolution result; (c) High-resolution data as a reference. The fine-tuned result enhances resolution and maintains consistency with the reference data, particularly in locally magnified regions.

2D network with 3D awareness effectively eliminates these artifacts (Figure 6c).

3) Pretraining on Synthetic Data Provides a Strong Initialization for Fine-Tuning: Without the pretraining step on synthetic data, directly applying the fine-tuning strategy

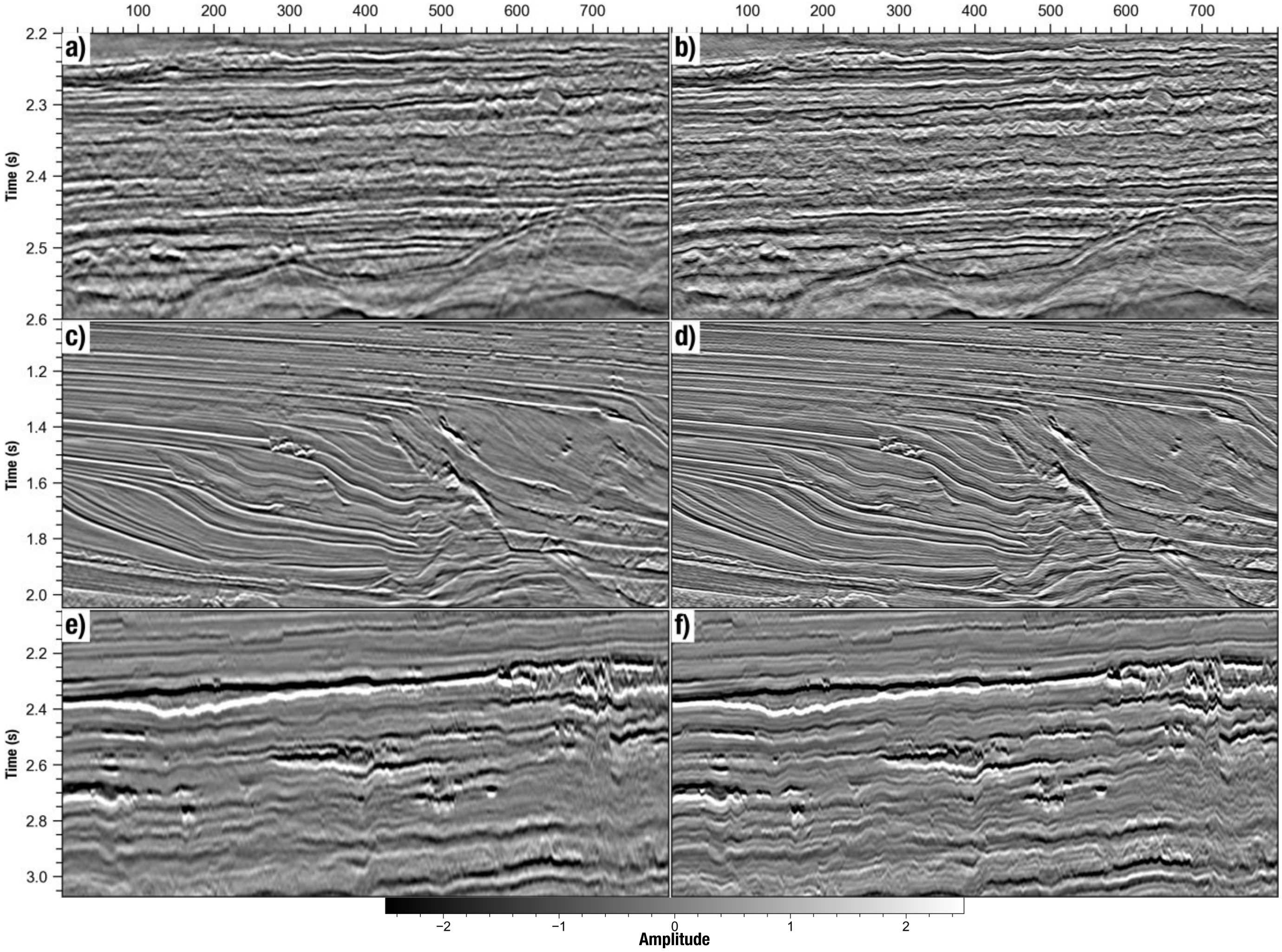

Fig. 11. Super-resolution results on three diverse field seismic datasets. The left column shows the raw field data, which exhibit significant differences from each other and from the data shown in Figure 6. The right column presents the high-resolution results obtained by our method, demonstrating its ability to generalize across varied data distributions.

struggles to produce high-fidelity super-resolution results, even though the low-frequency information can be preserved through the first loss function. The pretraining stage provides a strong initialization for fine-tuning, as the initial result already contains rich high-frequency information, which can be further refined to achieve greater realism. However, if the starting point for fine-tuning lacks high-frequency information, the results may become distorted. As shown in Figure 9b, while the overall structure is maintained, the amplitude variations appear unnatural. The observation is further supported by the quantitative metrics in Table IV. This limitation also poses challenges when applying the method to the field datasets with inherently broad frequency bands, as discussed in detail in the Limitations section.

4) Fine-Tuning Enhances the Realism of Super-Resolution Results: To understand the fine-tuning process, we visualize the frequency spectra of the field data at different epochs, as shown in Figure 9c. The numbers on the curves indicate the epoch, with fine-tuning starting from the 149th epoch. Each epoch consists of 50 iterations, with a batch size of 1 and f_c set to 50 Hz. Initially, the spectrum of the results shows a significant dip between 50-100 Hz. In the first 10 epochs, the SSDC loss dominates, improving the alignment of the low-frequency components with the original data. Once the low-frequency information is well-fitted, the PISM loss begins to play a significant role, gradually elevating the frequency components in the dip region. Table IV provides additional evidence for this trend. This process ultimately yields a more reasonable and realistic frequency distribution.

5) Prior-Informed Spectral Matching Loss Preserves High-Frequency Information: When only the SSDC loss is used, the model lacks constraints on high-frequency information, leading to a tendency to produce outputs identical to the input, resulting in model collapse. As shown in Figure 9d and Table IV, the high-frequency information in the results gradually diminishes during the fine-tuning process, almost completely disappearing by the 180th epoch. This highlights the critical role of the PISM loss in preserving high-frequency details.

Overall, the three components work together to produce high-quality results. Among them, the prior-informed fine-tuning has the strongest impact on field-data performance, as

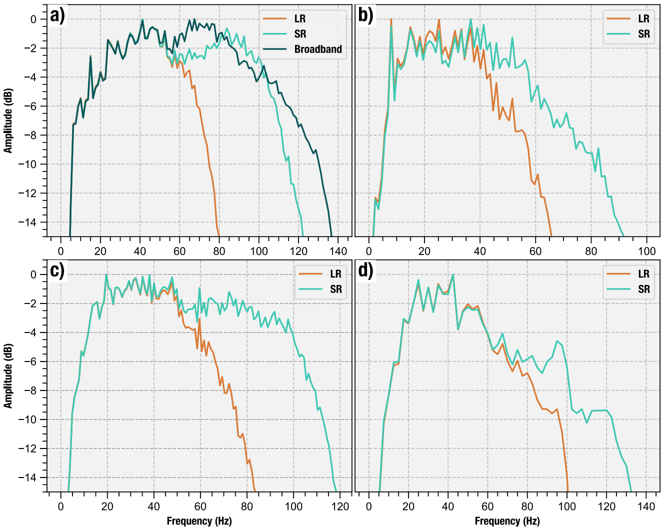

Fig. 12. Frequency spectra of the four datasets. (a) Broad-frequency dataset; (b-d) Three additional diverse datasets. The spectra demonstrate that our method preserves low-frequency components while extending the high-frequency range.

the SSDC and PISM losses inject essential domain knowledge that preserves reliable low-frequency information and guides physically meaningful spectral enhancement.

E. Cross-Survey Generalization

To test the proposed prior-informed fine-tuning strategy, we found that it exhibits strong cross-survey generalization capabilities, enabling it to handle diverse seismic data, even when their characteristics differ significantly from the target survey discussed earlier. We validated this claim on several field datasets.

First, we used a modern, high-resolution seismic dataset with a very broad frequency band (Figure 10c) as the reference or label. We applied a low-pass filter to this dataset to generate low-resolution seismic data (Figure 10a). Figure 10b shows the super-resolution results after fine-tuning with the low-resolution data. Despite the significant differences in characteristics compared to the previously used data, our fine-tuned method successfully enhanced the resolution, achieving results consistent with the true broad-frequency reference data. This is particularly evident in the locally magnified regions, demonstrating that our method not only improves resolution but also maintains high authenticity.

In addition, we present three additional field seismic datasets (left column of Figure 11) and their corresponding super-resolution results after fine-tuning (right column of Figure 11). These datasets exhibit diverse characteristics: the first is from a land survey, while the other two are from marine surveys. The first two datasets have a sampling interval of 2 ms, whereas the third is sampled at 1 ms. Despite these differences in acquisition type and sampling rate, our method consistently produces reliable and realistic super-resolution results without introducing model-like artifacts or significant false features.

The frequency spectra of these four datasets are shown in Figure 12. For the broad-frequency dataset (Figure 12a), the predicted results maintain strong consistency with the low-resolution data in the low-frequency range, while the overall spectrum closely matches the high-resolution reference data, demonstrating effective frequency band extension. For the other three datasets (Figure 12b-d), our method similarly preserves the low-frequency components of the original data while extending the high-frequency range, resulting in spectra that align well with the expected broadened characteristics. These spectral comparisons confirm that our approach not only enhances resolution but also maintains the integrity of the original information across diverse datasets.

F. Limitations

Although we have demonstrated the effectiveness of our method on multiple diverse field datasets, we clearly recognize that it still has certain limitations. First, during the fine-tuning stage, the weights of the two losses are challenging to balance. Different datasets may require repeated adjustments, as an excessively large weight for the SSDC Loss can cause the model to lose its ability to extend high frequencies, while an overly small weight may suppress the low-frequency components. Second, if the frequency range of the data is already very broad, resulting in the pre-trained network output lacking extended high-frequency signals, fine-tuning may have limited effect. As discussed in the Ablation Study, fine-tuning can adjust high-frequency information but cannot generate high-frequency signals from scratch.

In addition, the PISM Loss requires a user-defined reference spectrum S_{ref}. The spectrum follows the typical shape of controlled-source seismic signals, where a central frequency dominates and the energy tapers toward both sides. In practice, S_{ref} is chosen in the same way commonly used in geophysical processing. It can be specified by manually drawing a smooth spectral envelope or by adopting the spectrum of an Ormsby wavelet. These choices provide a simple and physically reasonable prior for guiding the spectral enhancement.

Finally, the field seismic data is non-stationary, with broader frequency content at the top and narrower frequency content at the bottom. Therefore, our method is best applied within a relatively small time window. When applying to the data with a large time window, a potential solution is to set different f_c and reference spectra S_{ref} for different time intervals and avoid applying the PISM Loss to the base area. It is worth emphasizing that this limitation does not significantly reduce the practical value of our method. In most real-world scenarios, geophysicists are primarily interested in specific target intervals, such as reservoir zones, rather than the full survey. Accurate super-resolution in these focused time windows is often sufficient and meaningful for interpretation and inversion.

V. Conclusion

In this study, we proposed a framework for improving the vertical resolution of seismic data, addressing challenges such as unrealistic outputs and limited generalization in deep learning-based methods. Our approach consists of three main components: realistic synthetic data generation, an enhanced 2D network with 3D awareness, and a prior-informed fine-tuning strategy. The synthetic data generation process incorporates structural and amplitude characteristics of field surveys, preserving small-scale features and avoiding overly smooth outputs. The enhanced 2D network, based on the Swin-Transformer architecture, captures 3D spatial features while maintaining computational efficiency and eliminating stitching artifacts. Most importantly, the prior-informed fine-tuning strategy plays a central role in achieving high-fidelity results, as it uses field data and is guided by Self-Supervised Data Consistency and Prior-Informed Spectral Matching losses, ensuring that the results retain original information while yielding a spectral distribution consistent with physical expectations.

Experiments on multiple field datasets demonstrated the effectiveness of our method. Compared to existing methods, our approach preserved small-scale discontinuities and maintained amplitude variations more effectively. The fine-tuning strategy further enhanced the realism of the outputs, making the method applicable to a variety of seismic datasets.

Although effective, the method still has some limitations. The fine-tuning loss weights may require manual adjustment for different datasets, and the PISM Loss depends on a user-defined reference spectrum. In addition, the non-stationary nature of seismic data means the method is most suitable for relatively small time windows. Future work may focus on more adaptive and data-driven strategies to reduce these dependencies and improve robustness across surveys.

In conclusion, our framework provides a practical solution for seismic resolution enhancement. By combining realistic synthetic data generation, an advanced network design, and unsupervised fine-tuning, we developed a method that improves resolution while maintaining the fidelity of the results. This approach has potential applications in diverse geological settings for seismic interpretation and reservoir characterization.

Acknowledgment

The authors thank the USTC supercomputing center for providing computational resources for this project. Finally, the authors are grateful to the anonymous reviewers and the editors for their insightful comments and suggestions, which have greatly improved this manuscript.

References

[1] X. Zhang and X. Zheng, "Thin-bed identification based on attribute difference between far and near offset within prestack data: A model study," in *Proc. SEG Tech. Program Expanded Abstr.*, Jan. 2007, pp. 293–297.

[2] M. B. Widess, "How thin is a thin bed?," *Geophysics*, vol. 38, no. 6, pp. 1176–1180, Dec. 1973.

[3] X. Wu, Z. Geng, Y. Shi, N. Pham, S. Fomel, and G. Caumon, "Building realistic structure models to train convolutional neural networks for seismic structural interpretation," *Geophysics*, vol. 85, no. 4, pp. WA27–WA39, Jul. 2020.

[4] Z. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 10012–10022.

[5] X. Chen, X. Wang, J. Zhou, Y. Qiao, and C. Dong, "Activating more pixels in image super-resolution transformer," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 22367–22377.

[6] Y. Fu, T. Zhang, L. Wang, and H. Huang, "Coded hyperspectral image reconstruction using deep external and internal learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 7, pp. 3404–3420, Jul. 2022.

[7] M. D. Sacchi, "Reweighting strategies in seismic deconvolution," *Geophys. J. Int.*, vol. 129, no. 3, pp. 651–656, Jun. 1997.

[8] K. Zhang, Y. Li, G. Liu, and Y. Jia, "Multiresolution seismic signal deconvolution," *Chin. J. Geophys.*, vol. 42, no. 4, pp. 529–535, 1999.

[9] S. D. Billings, B. R. Minty, and G. N. Newsam, "Deconvolution and spatial resolution of airborne gamma-ray surveys," *Geophysics*, vol. 68, no. 4, pp. 1257–1266, Jan. 2003.

[10] A. Gholami and M. D. Sacchi, "Fast 3D blind seismic deconvolution via constrained total variation and GCV," *SIAM J. Imag. Sci.*, vol. 6, no. 4, pp. 2350–2369, Jan. 2013.

[11] Y. Sui and J. Ma, "Blind sparse-spike deconvolution with thin layers and structure," *GEOPHYSICS*, vol. 85, no. 6, pp. V481–V496, Nov. 2020.

[12] Y. Zhang, H. Zhou, Y. Wang, M. Zhang, B. Feng, and W. Wu, "A novel multichannel seismic deconvolution method via structure-oriented regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5910410.

[13] E. Kjartansson, "Constant Q-wave propagation and attenuation," *J. Geophys. Res., Solid Earth*, vol. 84, no. B9, pp. 4737–4748, Aug. 1979.

[14] Y. Wang, "Inverse Q-filter for seismic resolution enhancement," *Geophysics*, vol. 71, no. 3, pp. V51–V60, Jan. 2006.

[15] I. L. S. Braga and F. S. Moraes, "High-resolution gathers by inverse q filtering in the wavelet domain," *Geophysics*, vol. 78, no. 2, pp. V53–V61, Mar. 2013.

[16] Y.-J. Xue, J.-X. Cao, and X.-J. Wang, "Inverse q filtering via synchrosqueezed wavelet transform," *Geophysics*, vol. 84, no. 2, pp. V121–V132, Mar. 2019.

[17] Y. Li et al., "Extended stable factor method for the inverseQ-filter," *Geophysics*, vol. 85, no. 3, pp. T155–T163, May 2020.

[18] M. Smith, G. Perry, J. Stein, A. Bertrand, and G. Yu, "Extending seismic bandwidth using the continuous wavelet transform," *First Break*, vol. 26, no. 6, pp. 97–102, Jun. 2008.

[19] M. C. de Matos and K. J. Marfurt, "Inverse continuous wavelet transform, 'deconvolution,'" in *Proc. SEG Tech. Program Expanded Abstr.*, 2011, pp. 1861–1865.

[20] J. Qi, B. Zhang, H. Zhou, and K. Marfurt, "Attribute expression of fault-controlled karst—Fort worth Basin, Texas: A tutorial," *Interpretation*, vol. 2, no. 3, pp. SF91–SF110, Aug. 2014.

[21] D. Agustianto and Y. Sun, "Seismic resolution enhancement via modified s-transform spectral balancing to improve interpretational aspect of the carbonate reservoir," in *SEG Tech. Program Expanded Abstr. 2018*, pp. 3332–3336, Society of Exploration Geophysicists, 2018.

[22] G. Blache-Fraser and J. P. Neep, "Increasing seismic resolution using spectral blueing and colored inversion: Cannonball field, trinidad," in *Proc. SEG Tech. Program Expanded Abstr.*, 2004, pp. 1794–1797.

[23] S. H. Kazemeini, C. Yang, C. Juhlin, S. Fomel, and C. Cosma, "Enhancing seismic data resolution using the prestack blueing technique: An example from the ketzin CO_2 injection site, Germany," *Geophysics*, vol. 75, no. 6, pp. V101–V110, Jan. 2010.

[24] P. Zhang, L. Liu, Z. Jiang, and Z. Zhang, "Application of seismic spectral blueing based on generalized S-transform in high-quality reservoir prediction in gas cloud area," in *Proc. SEG Int. Expo. Annu. Meeting*, 2022, pp. 1418–1422.

[25] P. Lu, M. Morris, S. Brazell, C. Comiskey, and Y. Xiao, "Using generative adversarial networks to improve deep-learning fault interpretation networks," *Lead. Edge*, vol. 37, no. 8, pp. 578–583, Aug. 2018.

[26] I. Goodfellow et al., "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014, pp. 2672–2680.

[27] D. Liu, W. Niu, X. Wang, M. D. Sacchi, W. Chen, and C. Wang, "Improving vertical resolution of vintage seismic data by a weakly supervised method based on cycle generative adversarial network," *Geophysics*, vol. 88, no. 6, pp. V445–V458, Nov. 2023.

[28] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2242–2251.

[29] S. Cheng, H. Zhang, and T. Alkhalifah, "Self-supervised seismic resolution enhancement," *IEEE Trans. Geosci. Remote Sens.*, vol. 63, 2025, Art. no. 5904115.

[30] J. Li, X. Wu, and Z. Hu, "Deep learning for simultaneous seismic image super-resolution and denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–11, 2022.

[31] F. Min, L. Wang, S. Pan, and G. Song, "D2UNet: Dual decoder U-Net for seismic image super-resolution reconstruction," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5906913.

[32] R. Zhou, C. Zhou, Y. Wang, X. Yao, G. Hu, and F. Yu, "Deep learning with fault prior for 3-D seismic data super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5907416.

[33] L. Lin, Z. Zhong, C. Cai, C. Li, and H. Zhang, "SeisGAN: Improving seismic image resolution and reducing random noise using a generative adversarial network," *Math. Geosci.*, vol. 56, no. 4, pp. 723–749, May 2024.

[34] Y. Choi, Y. Jo, S. J. Seol, J. Byun, and Y. Kim, "Deep learning spectral enhancement considering features of seismic field data," *Geophysics*, vol. 86, no. 5, pp. V389–V408, Sep. 2021.

[35] H. Zhang, T. Alkhalifah, Y. Liu, C. Birnie, and X. Di, "Improving the generalization of deep neural networks in seismic resolution enhancement," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023.

[36] Y. Xiao, K. Li, Y. Dou, W. Li, Z. Yang, and X. Zhu, "Diffusion models for multidimensional seismic noise attenuation and superresolution," *Geophysics*, vol. 89, no. 5, pp. V479–V492, Jan. 2024.

[37] H.-R. Zhang, Y. Liu, Y.-H. Sun, and G. Chen, "SeisResoDiff: Seismic resolution enhancement based on a diffusion model," *Petroleum Sci.*, vol. 21, no. 5, pp. 3166–3188, Oct. 2024.

[38] J. Sohl-Dickstein, "Deep unsupervised learning using nonequilibrium thermodynamics," in *Proc. Int. Conf. Mach. Learn.*, 2024, pp. 2256–2265.

[39] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 6840–6851.

[40] K. Ren, W. Sun, G. Yang, X. Meng, J. Peng, and H. Li, "Multistage hybrid denoising network for satellite hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5534717.

[41] K. Ren, W. Sun, X. Meng, and G. Yang, "GCM-PDA: A generative compensation model for progressive difference attenuation in spatiotemporal fusion of remote sensing images," *IEEE Trans. Image Process.*, vol. 34, pp. 3817–3832, 2025.

[42] Y. Ning, J. Peng, Q. Liu, W. Sun, and Q. Du, "Domain invariant and compact prototype contrast adaptation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5510214.

[43] M. Zhang, N. Wang, Y. Li, and X. Gao, "Deep latent low-rank representation for face sketch synthesis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 10, pp. 3109–3123, Oct. 2019.

[44] M. Zhang, N. Wang, Y. Li, and X. Gao, "Bionic face sketch generator," *IEEE Trans. Cybern.*, vol. 50, no. 6, pp. 2701–2714, Jun. 2020.

[45] M. Zhang, N. Wang, Y. Li, and X. Gao, "Neural probabilistic graphical model for face sketch synthesis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 7, pp. 2623–2637, Jul. 2020.

[46] Z. Yue, K. Liao, and C. C. Loy, "Arbitrary-steps image super-resolution via diffusion inversion," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2025, pp. 23153–23163.

[47] L. Zhang, W. You, K. Shi, and S. Gu, "Uncertainty-guided perturbation for image super-resolution diffusion model," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2025, pp. 17980–17989.

[48] X. Wu, Y. Li, and P. Sawasdee, "Toward accurate seismic flattening: Methods and applications," *Geophysics*, vol. 87, no. 5, pp. IM177–IM188, Sep. 2022.

[49] Z. Bi, X. Wu, Y. Li, S. Yan, S. Zhang, and H. Si, "Geologic-time-based interpolation of borehole data for building high-resolution models: Methods and applications," *Geophysics*, vol. 87, no. 3, pp. IM67–IM80, May 2022.

[50] G. Wang, X. Wu, and W. Zhang, "Cigchannel: A massive-scale 3D seismic dataset with labeled paleochannels for advancing deep learning in seismic interpretation," *Earth Syst. Sci. Data Discuss.*, vol. 2024, pp. 1–27, 2024.

[51] N. B. Bynagari, "GANs trained by a two time-scale update rule converge to a local Nash equilibrium," *Asian J. Appl. Sci. Eng.*, vol. 8, no. 1, pp. 25–34, Apr. 2019.

[52] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, "SwinIR: Image restoration using Swin transformer," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2021, pp. 1833–1844.

[53] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 586–595.

[54] Y. Li, X. Wu, Z. Zhu, J. Ding, and Q. Wang, "FaultSeg3D plus: A comprehensive study on evaluating and improving CNN-based seismic fault segmentation," *Geophysics*, vol. 89, no. 5, pp. 1–57, Jan. 2024.

[55] D. P. Kingma, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[56] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 60, 2017, pp. 84–90.

[57] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*, Munich, Germany, 2015, pp. 234–241.

[58] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 1132–1140.

[59] L. Zhang, Y. Li, X. Zhou, X. Zhao, and S. Gu, "Transcending the limit of local window: Advanced super-resolution transformer with adaptive token dictionary," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 2856–2865.