communications engineering

Article

A Nature Portfolio journal

https://doi.org/10.1038/s44172-026-00674-9

Adapting general representations of pretrained vision foundation models to seismic understanding

Zhengfa Bi[1,2], Xinming Wu[2] ✉, Nuo Chen[2] & Nori Nakata[1,3]

Interpreting seismic data remains challenging because most learning-based approaches rely on specific architectures that require retraining for each task and generalize poorly across subsurface regimes when training data are limited. Here, we introduce a cross-domain transfer learning framework that adapts a vision foundation model for seismic understanding. A lightweight seismic-to-vision bridge maps seismic data into the representational space of pretrained backbone, while low-rank adaptation and prefix tuning enable efficient parameter updates that preserve general visual priors. Geological constraints are incorporated via stratigraphic prompting that injects stratigraphic order into the latent space, guiding predictions to respect structural consistency. A task-adaptive decoder enables flexible downstream applications, including facies segmentation, structural interpretation, and property inversion. The framework outperforms baseline models across diverse tasks and datasets with fewer parameters, demonstrating that adapting pretrained vision foundation model via geologically informed prompting and lightweight tuning enables transferable seismic interpretation and broader data-driven geoscientific applications.

Seismic interpretation plays a central role in subsurface characterization, offering critical insights into lithostratigraphic sequences, depositional systems, and the spatial organization of key petroleum system elements[1,2]. Identifying seismic facies and structural features supports sedimentological and stratigraphic analysis by capturing spatial variations in reflection attributes such as amplitude strength, reflection continuity, and spectral or phase characteristics. Despite its importance, deriving geologically meaningful patterns from seismic data remains inherently challenging. Complex depositional settings and the progressive loss of resolution with depth often obscure crucial subsurface features. At the same time, the scale and dimensionality of seismic datasets continue to increase, while the number of trained interpreters remains relatively constant.

Although traditional experiential learning partially addresses this expertise gap, a promising alternative lies in encoding interpretive knowledge into intelligent systems capable of learning directly from data[3–5]. Recent breakthroughs in artificial intelligence, particularly deep learning, have transformed pattern recognition across scientific disciplines[6–10], including seismology[11–13], atmospheric sciences[14,15], planetary studies[16,17], and geologic modeling[18–20].

In seismic interpretation, supervised deep learning has become the prevailing paradigm due to its capacity to model complex nonlinear relationships. Nevertheless, the effectiveness of such models is heavily constrained by their reliance on large, high-quality labeled datasets. This dependence introduces issues such as subjectivity in labeling, limited scalability, and inconsistent taxonomies across regions and interpreters. To alleviate the scarcity of annotated data, numerous studies[21,22] have adopted synthetic datasets crafted with domain expertise. While synthetic data can provide scale and control, models trained solely on artificial examples often suffer from poor generalization to field data, especially in structurally complex or previously unseen geological settings[22,23]. This limitation arises as synthetic datasets typically generated through predefined workflows and parametric models may struggle to capture the full diversity of real-world seismic patterns.

The recent emergence of foundation models offers a promising direction for overcoming these limitations. Pretrained on large-scale and diverse datasets, these models can learn generalized and expressive feature representations, enabling strong transferability across a wide range of downstream tasks[24,25]. Their representational capacity driven by billions or even trillions of parameters, allows them to capture spatial patterns beyond the ability of conventional task-specific models. Representative examples such as GPT-4[26], PaLM[27], DINOv2[28], and SAM[29] have set new performance

[1]Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [2]State Key Laboratory of Precision Geodesy, School of Earth and Space Sciences, University of Science and Technology of China, Hefei, China. [3]Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA. ✉e-mail: xinmwu@ustc.edu.cn

benchmarks in language understanding, image segmentation, and multimodal reasoning[30,31].

As most foundation models have been developed beyond geosciences, their extension to seismic interpretation presents a natural yet underexplored opportunity. Preliminary studies have begun bridging this gap[32,33], but widespread adoption has been limited by the lack of massive and well-curated labeled data, and the prohibitive computational cost of full model pretraining. As advances in data acquisition improve the resolution and enrich the content of data, and as model architectures continue to evolve, the need to repeatedly retrain domain-specific models becomes increasingly unsustainable.

A practical alternative is to reuse foundation models pretrained on large-scale natural image datasets and adapt them for seismic applications via transfer learning. Similar strategies have demonstrated success in other geoscientific applications, such as global seamount detection[34]. Despite differences between natural and seismic imagery, they share similarities in their dense, array-based representations. Seismic profiles resemble grayscale images composed of continuous pixel intensities. Furthermore, common AI tasks such as classification, regression, and anomaly detection are shared across both domains. These structural parallels suggest that vision foundation models can be effectively adapted for seismic interpretation. Recent studies[35,36] have explored prompt-guided fine-tuning of vision models to enable instance segmentation and local analysis of seismic features. However, these approaches stop at vision pattern recognition, and post manual interpretation is required to link predictions to geological meanings.

In this study, we introduce a transfer learning framework that repurposes vision foundation models as seismic backbones through parameter-efficient adaptation and geological prompting to address core tasks in seismic interpretation. A seismic bridge aligns amplitude inputs with the pretrained representation space, while low-rank adaptation (LoRA)[37] and prefix tuning[38] inject learnable flexibility into attention modules without retraining billions of parameters. This strategy retains the rich representational capacity of the foundation model while adapting it to domain-specific needs. Beyond representation transfer, we embed structural priors into the latent space of the foundation model, which bridges the gap between data-driven pattern recognition and physics-guided interpretation. By encoding stratigraphic ordering and continuity, these priors guide the transformer's attention toward geologically consistent regions and act as soft constraints during optimization. Our framework also couples with a task-adaptive decoder that accommodates both segmentation and regression, which enables broad applicability across diverse interpretation scenarios.

We validate our approach across multiple benchmark datasets spanning diverse geological settings, showing that it consistently outperforms both scratch-trained and fully fine-tuned baselines. It achieves superior accuracy with fewer trainable parameters while preserving fine-scale structural details. The results indicate the effectiveness of repurposing vision foundation models in seismic interpretation, and emphasize their potential to advance data-driven geoscientific discovery under supervision-constrained conditions. More broadly, our work positions general-purpose foundation model adaptation as a unifying geologically informed paradigm for data-driven discovery in the geosciences.

Methods

We introduce a transfer learning framework that employs the representational power of pretrained vision foundation models (Fig. 1). It enables seamless adaptation to the seismic domain regardless of the backbone architecture. Our framework mainly consists of four modules, including a seismic bridge module, a vision foundation encoder, a task-adaptive decoder, and a structure prior prompter. The seismic bridge model aligns the representation space of seismic inputs with that of natural images for effective transfer learning with minimal computational overhead. Embedding a LoRA matrix into the frozen backbone enables efficient cross-domain adaptation without the need for extensive retraining of the foundation models. The decoder generalizes the feature representations to a wide range of seismic interpretation tasks. A prompting module integrates geologically informed priors to enforce physical consistency in predictions.

Bridge model for cross-domain representation alignment

While vision foundation models are originally designed to process natural images composed of three-channel color inputs, seismic data exhibit different semantic characteristics. These differences can impair the foundation model's ability to extract meaningful representations from seismic images. To mitigate this modality gap, we utilize a lightweight bridge model that transforms seismic data into a representation space compatible with the frozen backbone of the pretrained vision model.

This module consists of two convolutional layers followed by normalization and non-linear activation, which project the input seismic data with single-channel grayscale amplitude maps into high-dimensional feature representations. The transformed output mimics the properties expected by vision foundation models. It not only enables seamless integration without modifying the core architecture but also ensures knowledge transfer from large-scale natural image datasets. Moreover, it provides a learnable interface that adapts seismic characteristics to the statistical priors embedded in the foundation model.

Low-rank and prefix adaptation for transfer learning

Foundation models possess billions of parameters and are typically trained on large-scale natural image datasets. In this study, we employ the Segment

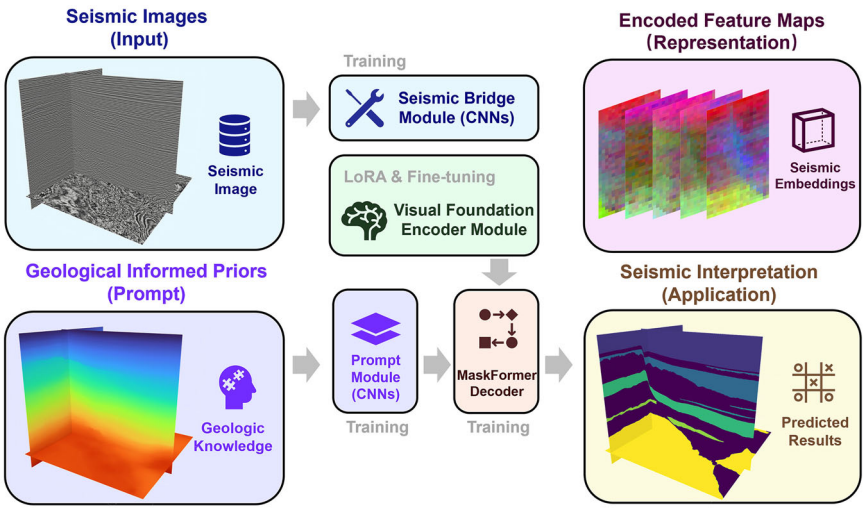

Fig. 1 | Overview of our transfer learning framework for seismic interpretation. Pretrained vision foundation models are adapted to seismic images via a seismic-to-vision representation mapping that aligns seismic data with visual features learned from natural images. Lightweight adaptation modules, including low-rank adaptation and prefix tuning, enable efficient parameter updates while preserving pretrained knowledge. Geologically informed prompting injects stratigraphic order into the latent space, guiding predictions toward structurally consistent representations.

Anything Model (SAM)[29] as a general-purpose feature extractor for seismic interpretation. SAM is a state-of-the-art vision model trained on an extensive dataset comprising over 1.1 billion masks and 11 million images, with powerful vision representations. Despite being optimized for class-agnostic segmentation without any physical understanding, its rich embedding space shows strong potential for adaptation to cross-domain tasks. However, our framework is not limited to SAM and can be extended to any modern vision foundation model, allowing future integration with more powerful architectures as they emerge.

Implementation within the pretrained encoder. Fine-tuning an entire vision foundation model is computationally expensive and prone to overfitting when only limited seismic labels are available. To address this, we employ parameter-efficient adaptation by injecting LoRA and prefix tuning into the attention modules of the pretrained SAM encoder (see Supplementary Information, Notes 1 and 2). These techniques allow the model to acquire seismic-specific representations while preserving the general visual priors learned during large-scale pretraining. Given that most existing foundation models are built upon transformers, this adaptation is broadly extendable to other foundation models.

LoRA introduces lightweight trainable matrices that provide low-rank updates to the query and key projections in each attention block. Rather than modifying the full attention weights, these updates act as small, learnable perturbations applied on top of the frozen pretrained parameters. This design enables the encoder to adapt to seismic structures with only a small fraction of additional parameters. Prefix tuning further enhances flexibility by appending a set of learnable prefix tokens to the key and value sequences. These tokens act as soft prompts that modulate the attention patterns in a task-dependent manner. These two mechanisms are jointly embedded into the SAM encoder to form a parameter-efficient strategy for cross-domain seismic transfer. The detailed mathematical formulations of multi-head attention, LoRA, and prefix construction are provided in the Appendix for reference.

Hybrid adaptation with selective attention tuning. As shown in Fig. 2a, the input features are first embedded with positional encodings to retain spatial context before being passed into the multi-head attention module. They are then projected into queries, keys, and values, which serve as the inputs to the attention block. Within each head, attention weights are applied to the value vectors to yield head-specific outputs. The outputs from all heads are concatenated and passed through a linear projection to form the attention output. The attention output is then combined with the input via a residual connection, followed by a Layer Normalization operation. This output is subsequently passed through a linear layer to model higher-order token-wise interactions. A second residual connection is applied thereafter, yielding the final output of the multi-head attention. We restrict adaptation solely to the attention modules within the image encoder of the foundation model.

To achieve knowledge transfer from vision foundation models to seismic interpretation, we adopt a hybrid adaptation strategy that combines LoRA, prefix tuning, and partial fine-tuning. LoRA and prefix tokens are embedded within the attention layers of the transformer encoder, while the final one or two encoder blocks are unfrozen to allow gradient-based updates. We refer to these configurations as FM-LoRA1, where only the final block is updated, and FM-LoRA2, where the last two blocks are updated. This hierarchical tuning enables a trade-off between efficiency and representational flexibility to retain general vision priors while capturing domain-specific features. For comparison, we denote the encoder backbone trained from scratch as FM-Scratch, and the one fine-tuned from pretrained weights as FM-Finetune. The variant that incorporates prompt embeddings for structural guidance is referred to as FM-Prompt.

Embedding geological knowledge via prompting

We embed geological priors into the latent space of a foundation model through a prompting mechanism (Fig. 1) to reconcile learned visual patterns with physically meaningful features. The priors derived from RGT maps or other attributes encode stratigraphic coherence and depositional chronology, which are often difficult for purely data-driven models to infer.

The prompter consists of convolutional layers that project geological priors into spatially aligned feature maps. These feature maps are then integrated into the image embeddings through positional encodings to guide the transformer's attention towards stratigraphically consistent regions (Fig. 2b). The implementation extends the decoder by introducing additional prompt embeddings that are linearly combined with the image embeddings. A weighting factor further modulates the influence of these priors, preventing overfitting to uncertain interpretations. The enriched latent representations are then translated by the upscaling pathway into high-resolution outputs that preserve fine-scale seismic details and large-scale stratigraphic order.

Task-adaptive decoding with bidirectional feature fusion

The decoder module transforms latent embeddings into task-specific outputs (Fig. 1). It receives seismic representations from the foundation model alongside prompt embeddings encoding prior constraints. These inputs are jointly processed by a MaskFormer-based decoder[39] (Fig. 2b), with a core architecture based on Two-Way Transformers. This architecture allows bidirectional interaction between learnable output tokens and spatially distributed features. It alternates between four operations at each layer, including self-attention among output tokens, cross-attention from tokens to image features, a feedforward multilayer perceptron (MLP), and reverse cross-attention from image features back to tokens. These operations enable global token-level reasoning and localized spatial refinement, which is essential in seismic interpretation where structural features often span multiple spatial scales.

The output embeddings from the Two-way Transformer are refined through a multi-stage convolutional pipeline and integrated self-attention to retain spatial details. Each output token is passed through an MLP to produce weights that encode semantic reasoning for each upsampled feature map. For image-to-image regression tasks, dense predictions are produced via dot products between these weights and the feature maps. This yields spatially resolved outputs with shape $N \times W \times H$, where N denotes the token number and $W \times H$ corresponds to the spatial resolution. For classification and segmentation tasks, token-derived weights are transformed through an MLP followed by non-linear activation to yield class probabilities, producing compact outputs of shape N. By combining structured priors, foundation model features, and token-level reasoning, our approach delivers global semantic understanding with localized spatial features. This design accommodates a broad range of seismic interpretation tasks, including structural delineation and facies classification.

Results

To assess the generalizability of our framework across diverse geologic contexts and task settings, we conduct experiments on three publicly available benchmark datasets. These datasets represent varying levels of structural complexity and depositional diversity, ranging from seismic facies classification to structural interpretation. All models are trained on a single NVIDIA A100 GPU, and the parameter-efficient LoRA configurations decrease peak GPU memory usage by approximately 55% and proportionally accelerate each training epoch. In the following subsections, we provide detailed qualitative and quantitative evaluations for each benchmark and discuss the roles of pretraining and structural prompting in achieving these performance gains.

We employ multiple segmentation metrics to compare different encoder configurations, including FM-Finetune, FM-Scratch, FM-LoRA1, and FM-LoRA2. We benchmark them against a diverse set of modern architectures widely adopted in state-of-the-art geophysical and computer-vision workflows. The comparison includes heavyweight backbones such as deeper residual networks (ResNet[40,41], Xception[42]), hybrid convolution-transformer designs (ConvNeXt[43]), and dilated architectures that enhance multiscale context aggregation (DRN[44]). The evaluation also covers

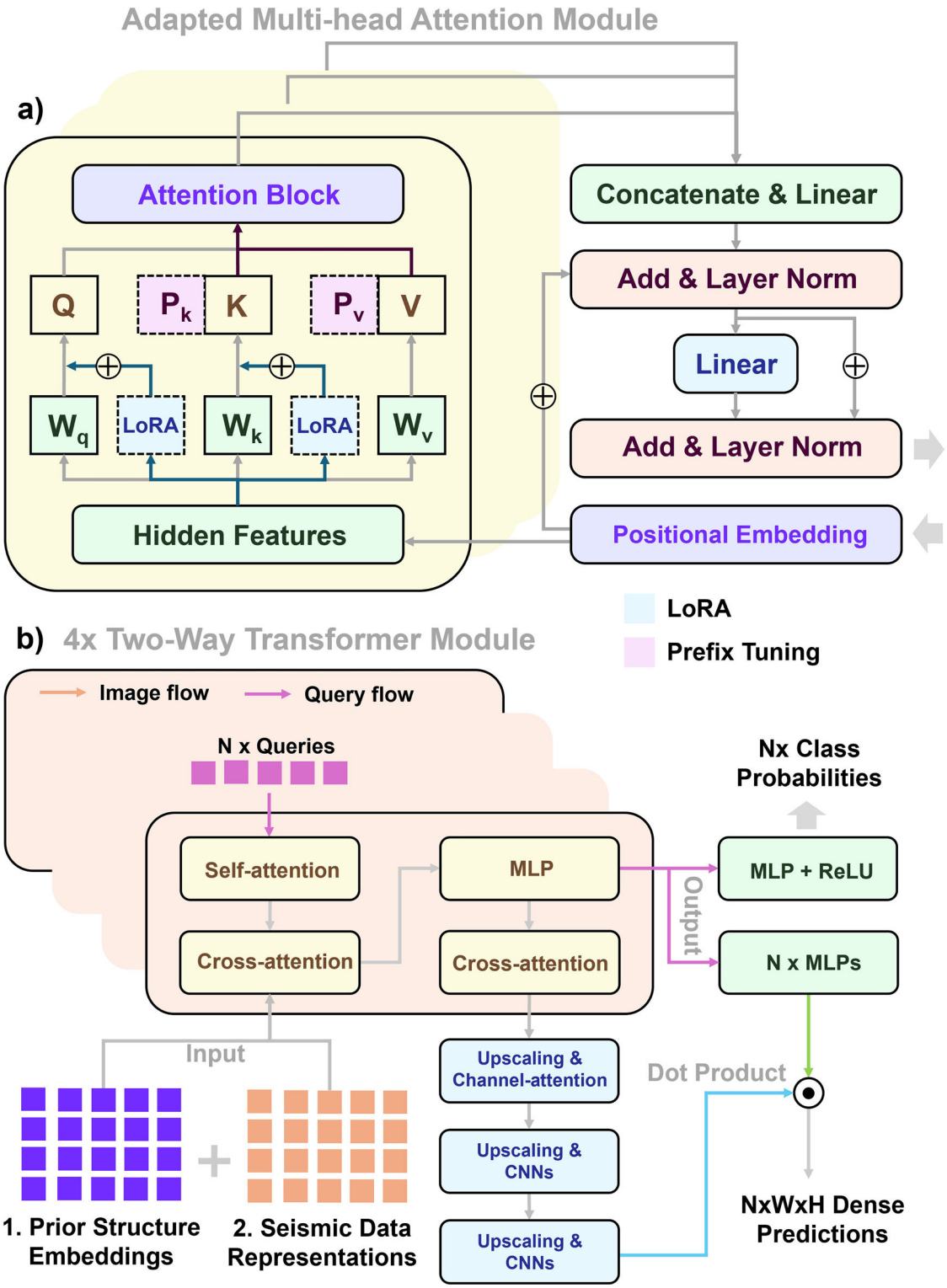

Fig. 2 | Architecture of the proposed framework integrating geologically informed prompting with lightweight adaptation for seismic representation learning. a Low-rank adaptation and prefix tuning are incorporated into attention layers throughout the foundation model encoder to enable efficient parameter updates. **b** The decoder employs Two-way Transformers to achieve bidirectional fusion between output tokens and spatial features, while structure-aware prompting injects geological priors into the learned representations.

computationally efficient lightweight models, including EfficientNet[45] and MobileNet[46]. All baselines utilize official implementations to ensure reproducibility. Model complexity varies across backbones, with FM-LoRA1, FM-LoRA2, EfficientNet, and MobileNet under 100 million parameters, ResNet, DRN, and Xception in the 100 – 120 million range, and ConvNeXt, FM-Finetune, and Scratch above 160 million (Fig. 3a).

Facies segmentation in the Netherlands F3 block

Segmentation is critical in seismic interpretation, including both lithological facies classification and the delineation of intricate geological structures. In this section, we focus on the representative example of seismic facies segmentation. We employ the fully annotated Netherlands F3 block as a benchmark dataset[47–49] to demonstrate the effectiveness of our framework.

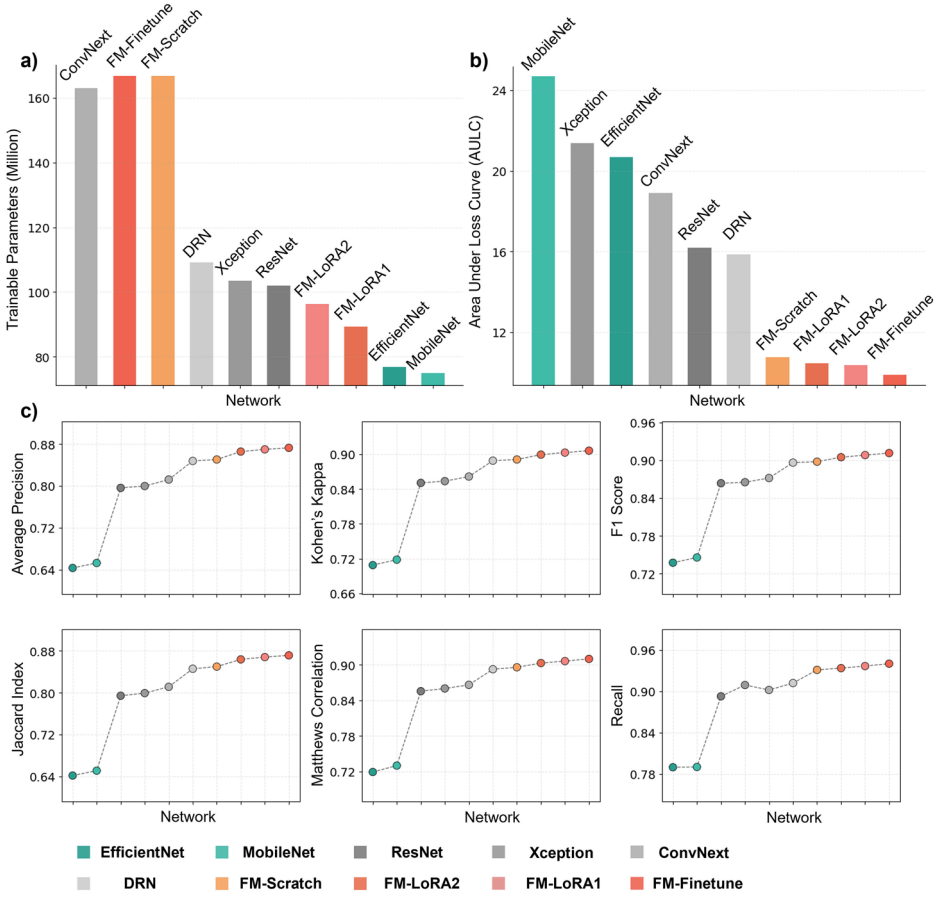

Fig. 3 | Quantitative analysis of segmentation performance and model complexity. a Number of trainable parameters for each backbone configuration. **b** Area Under Loss Curve (AULC) for different model configurations. **c** Evaluation metrics across these configurations. LoRA1 achieves a strong trade-off between performance and model complexity, outperforming both lightweight and heavyweight baselines.

This dataset comprises a post-stack time-migrated seismic volume that spans approximately 16km × 24km, with dimensions of 255 × 701 × 401 in depth, crossline, and inline directions. The corresponding facies volume derived from 26 boreholes contains seven lithostratigraphic units, including the Upper North Sea, Middle North Sea, Lower North Sea, Chalk, Rijnland, Scruff, and Zechstein formations. Because the boundary between the Chalk and Rijnland units is difficult to distinguish, the two Cretaceous intervals are merged into a single facies class. It encompasses a geologically complex region located at the boundary between the Dutch Central Graben and the Step Graben in the North Sea, where tectonic processes and salt diapirism have produced pronounced structural heterogeneity.

To construct the training and validation sets for our experiments, we extract a total of 169 seismic and facies sections of shape 255 × 701 samples from the 3-D volumes. These slices are extracted at varying azimuthal orientations and positions to enhance structural diversity. We rotate through multiple azimuths at 10° increments and sample neighboring planes at intervals of 10 grid units. Among the extracted slices, 80% are randomly assigned to the training set, while the remaining 20% are reserved for validation.

We employ our framework to train models with various encoder backbones, each augmented with our proposed adaptation modules. Their backbone encoders are either trained from scratch or fine-tuned depending on the transfer learning strategy, while the adaptation modules are initialized from scratch. Training is performed using the AdamW optimizer with a learning rate of 1×10^{-4} and a weight decay of 1×10^{-2}. The learning rate is adaptively reduced using a scheduler with a patience of 10 epochs and a decay factor of 0.5. The total number of training epochs is set to 100, and the batch size is set proportionally to the number of available GPUs, using 24 samples per GPU. Although the epoch count may appear high given the use of pretrained backbones, this choice ensures that all architectures under comparison, such as those initialized from scratch or equipped with different encoder depths, reach stable convergence.

A hybrid loss function combining focal loss and Dice loss is used to address class imbalance and identify the boundary of seismic facies by optimizing the overlap between predicted and ground truth masks. The focal loss is formulated as:

$$\mathcal{L}_{\text{focal}} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C}\alpha_c\,(1-p_{i,c})^{\gamma}\,y_{i,c}\log(p_{i,c}), \tag{1}$$

where N is the total number of pixels, C is the number of facies classes, $p_{i,c} \in [0, 1]$ is the predicted probability for class c at pixel i, and $y_{i,c} \in \{0, 1\}$ is the corresponding one-hot encoded ground truth. The factor $\alpha_c \in [0, 1]$ compensates for class imbalance, and $\gamma > 0$ is the focusing parameter that down-weights well-classified examples. The Dice loss is defined as:

$$\mathcal{L}_{\text{dice}} = 1 - \frac{2\sum_{i=1}^{N}\sum_{c=1}^{C}p_{i,c}\,y_{i,c} + \epsilon}{\sum_{i=1}^{N}\sum_{c=1}^{C}p_{i,c} + y_{i,c} + \epsilon}, \tag{2}$$

where ϵ is a small constant added for numerical stability. The overall facies segmentation loss is then expressed as a weighted combination of the two components:

$$\mathcal{L}_{\text{facies}} = \lambda_{\text{focal}}\,\mathcal{L}_{\text{focal}} + \lambda_{\text{dice}}\,\mathcal{L}_{\text{dice}}, \tag{3}$$

where $\lambda_{\text{focal}} = 20.0$ and $\lambda_{\text{dice}} = 1.0$ control the relative contributions of each term.

To further assess convergence behavior under a unified training budget, we compute the Area Under the Loss Curve (AULC) for all backbone and adaptation configurations (Fig. 3b). AULC provides an integrated

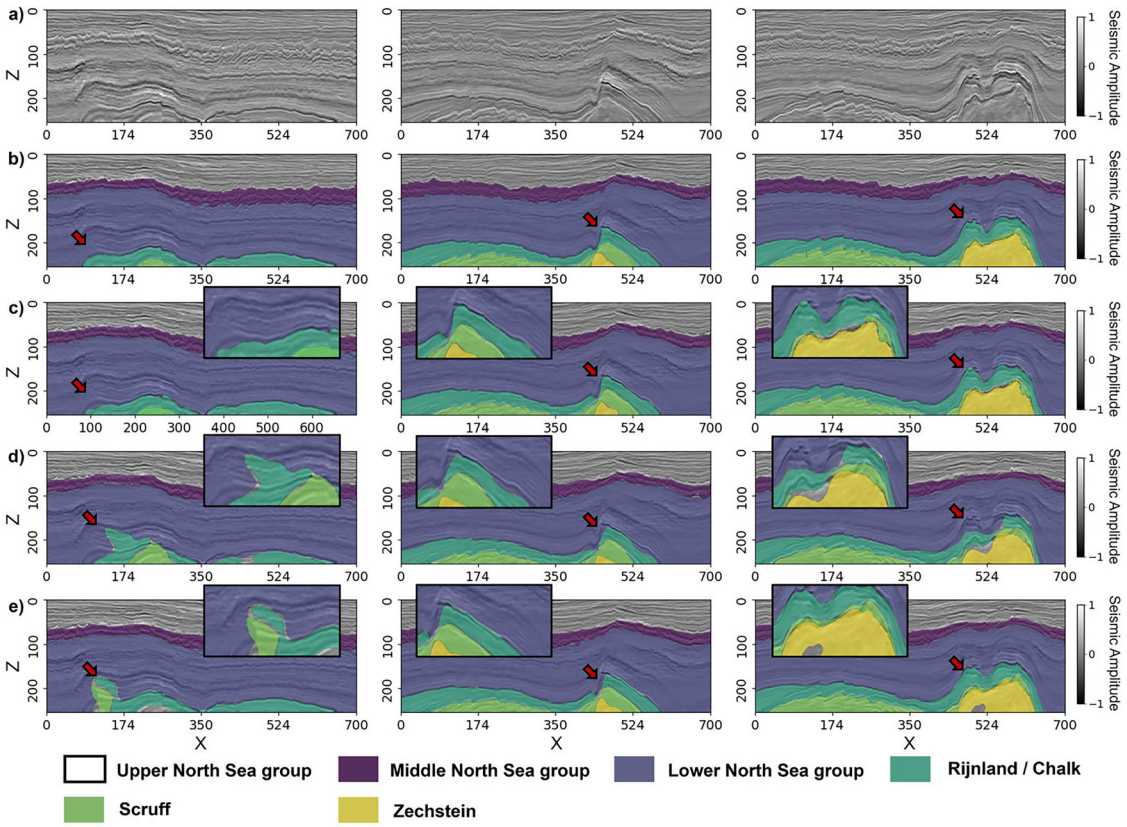

Fig. 4 | Qualitative comparison of facies segmentation results on slices from the validation set of the Netherlands F3 block. a Seismic profile. **b** Ground-truth facies labels with six lithostratigraphic units. **c** Predicted facies from our method using LoRA1. **d**, **e** Predictions from DRN and ResNet, respectively. Red arrows indicate regions affected by salt tectonics and structural deformation, where baseline models fail to delineate facies boundaries, while FM-LoRA1 produces stratigraphically aligned predictions. The black insets show zoomed-in views of the regions indicated by the arrows.

measure of optimization efficiency, where smaller values indicate faster descent and more rapid stabilization of the loss. Across the evaluated models, the configuration with full fine-tuning (FM-Finetune) exhibits the lowest AULC, reflecting the fastest convergence among all tested variants. The LoRA-based models (FM-LoRA1 and FM-LoRA2) follow closely, showing faster loss reduction than both scratch-trained models and conventional baselines. The improved convergence of these configurations is consistent with their initialization from pretrained vision backbones, indicating that the transferred priors provide a favorable optimization starting point for seismic tasks.

Qualitative comparisons demonstrate the effectiveness of our approach (Fig. 4). FM-LoRA1 exhibits superior detection of complex facies boundaries and preserves stratigraphic continuity, outperforming DRN and ResNet, which struggle in salt-intruded or tectonically uplifted regions (red arrows). This is evident in boundary fragmentation and geologically implausible facies assignments. Quantitative results in Fig. 3c support these observations. FM-LoRA1 consistently achieves the highest scores across metrics, which indicates the benefit of lightweight adaptation using a pretrained foundation models. In contrast, ResNet and DRN suffer from overfitting due to the limited amount of labeled data, which leads to poor generalization on unseen validation slices. While full fine-tuning (FM-Finetune) yields competitive performance, it requires substantially more parameters (above 160 million). In contrast, FM-LoRA1 reaches comparable results with a significantly smaller model footprint, demonstrating the efficiency and scalability of our approach.

Structural interpretation in the Teapot Dome dataset

To demonstrate the applicability of our framework beyond facies segmentation, we evaluate it on the publicly available Teapot Dome dataset[50,51] for structural interpretation. This dataset has been widely used in seismic attribute analysis and fracture characterization, and serves as a benchmark for seismic interpretation tasks, including Relative Geologic Time (RGT)[52–54] estimation and borehole-constrained property inversion. We utilize the models using different encoder backbones to estimate RGT images. RGT images reflect a continuous stratigraphic ordering of seismic reflections, assigning each voxel a relative age that increases monotonically with depth along locally coherent structural directions. This attribute captures the depositional chronology encoded implicitly within the seismic volume and serves as a powerful descriptor of stratigraphic architecture. The dataset consists of a post-stack time-migrated seismic volume with dimensions of $401 \times 357 \times 161$ samples in depth, crossline, and inline directions. To construct training and validation sets, we extract 139 seismic-RGT slice pairs with a size of 401×357 from multiple azimuths sampled at 10° increments, and spaced at 10 grid units. Among these slices, 80% are randomly assigned to training, and the remaining 20% are reserved for validation. Model training is conducted using the AdamW optimizer with the same hyperparameters as in prior experiments. As this is a regression task, we adopt the Mean Squared Error (MSE) loss function.

Compared to ResNet, our lightweight adaptation model FM-LoRA1 demonstrates an improved ability to capture structural features, producing RGT profiles that align more closely with the ground truth across multiple validation slices (Fig. 5a–d). In structurally complex regions such as anticlines and synclines or low-quality regions, ResNet struggles to follow reflectors (red arrows in Fig. 5), whereas FM-LoRA1 maintains structural coherence and yields geologically consistent stratigraphic layering. Quantitative comparisons (Table 1) further support these observations. Across multiple regression metrics, LoRA configurations consistently outperform both fine-tuned and scratch-trained counterparts. Such a performance gap is not caused by differences in model capacity but instead reflects the challenge of overfitting and poor generalization in training on limited data.

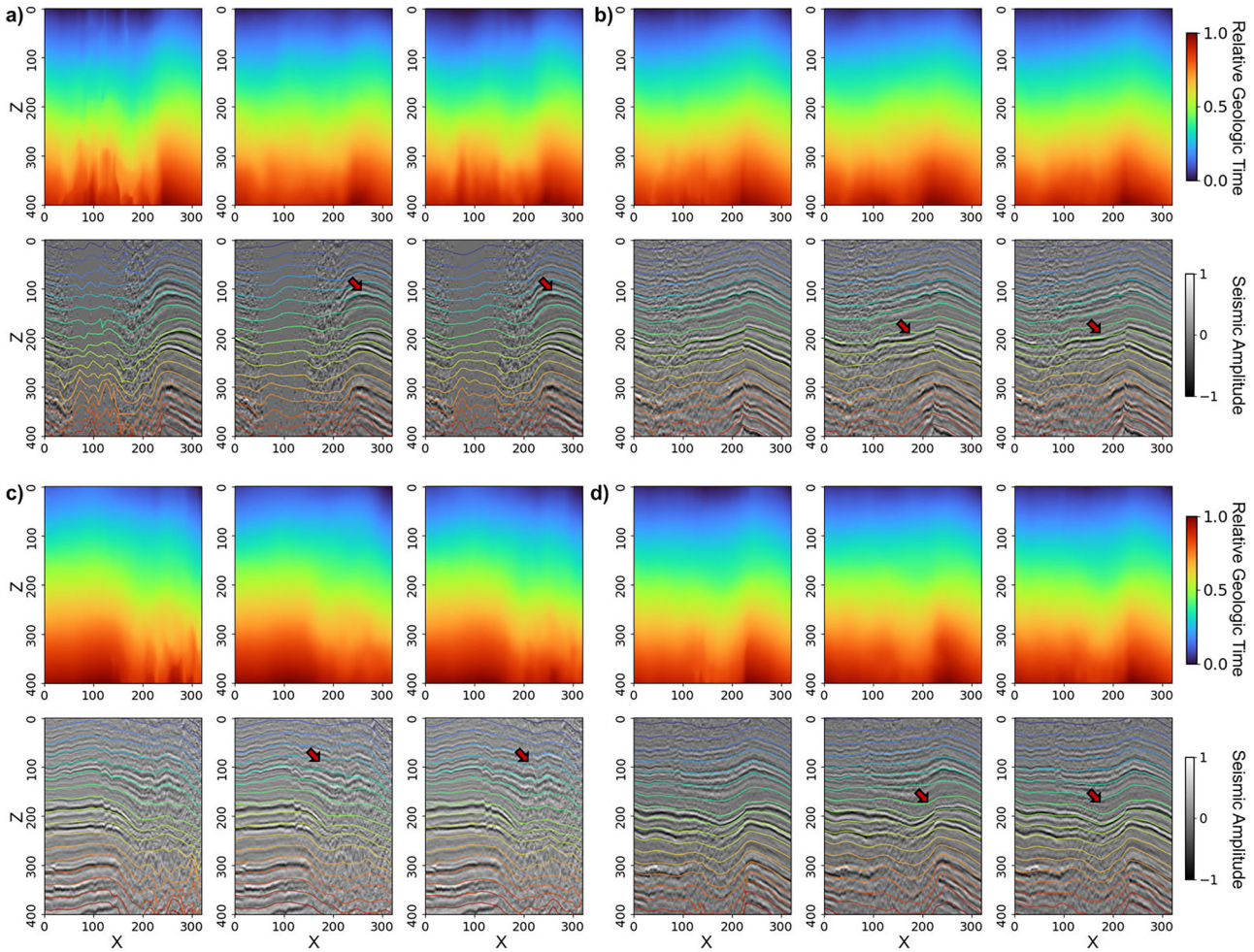

Fig. 5 | Comparisons of Relative Geologic Time (RGT) predictions on validation slices from the Teapot Dome dataset. a–d Represent different slices sampled across the volume. For each subfigure, the top row displays the ground truth RGT image (left), the RGT prediction from ResNet (middle) and LoRA1 (right). The bottom row shows the corresponding seismic sections overlaid with horizons extracted from the ground truth (left), ResNet (middle) and LoRA1 (right). Red arrows indicate structurally complex regions or low signal-to-noise zones where ResNet fails to continuously track reflectors, while FM-LoRA1 yields geologically plausible stratigraphic layers.

Table 1 | Performance comparison of network backbones on the validation set

Backbone	Parameters (M)	MAE ↓	MSE ↓	PSNR (dB) ↑	RMSE ↓	SSIM ↑
FM-LoRA1	92.7	0.0439	0.0037	**40.71**	0.0505	0.9960
FM-LoRA2	99.8	**0.0408**	**0.0033**	39.98	**0.0487**	**0.9961**
FM-Finetune	170.3	0.0660	0.0086	39.83	0.0720	0.9957
FM-Scratch	170.3	0.0433	0.0032	37.77	0.0525	0.9949
ResNet	111.3	0.0554	0.0060	36.44	0.0663	0.9931
DRN	118.3	0.0546	0.0054	37.87	0.0635	0.9948
ConvNeXt	172.2	0.0532	0.0052	34.92	0.0644	0.9915
Xception	112.8	0.0708	0.0095	32.11	0.0865	0.9899
MobileNet	84.3	0.0697	0.0096	31.87	0.0864	0.9894
EfficientNet	86.0	0.0724	0.0107	32.67	0.0901	0.9910

Bold values indicate the best performance for each metric.
Lower values indicate better performance for mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE); higher values indicate better performance for peak signal-to-noise ratio (PSNR, dB) and structural similarity (SSIM).

Under data-scarce conditions, our lightweight adaptation strategy using pretrained foundation model enables robust generalization to previously unseen structural patterns. In contrast, large models trained from scratch struggle to generalize reliably, and lightweight models without pretraining lack sufficient representational capacity.

To evaluate the behavior of the proposed framework under restricted supervision, we further conduct experiments in which the proportion of available training slices is systematically reduced from 80% down to 10%, with the remaining slices reserved for validation. The full dataset contains only 63 profiles extracted from multiple azimuths across the entire volume. To mitigate variability introduced by random train-validation partitions, we repeat each experiment three times and report the averaged metrics. Across this sequence of increasingly constrained training scenarios, the foundation-model-based approach exhibits a high degree of resilience to reduced supervision. In the low-data regimes (10–30%), where baseline architectures show substantial deterioration due to limited labeled information, our method retains consistently strong performance across all evaluation metrics, including RMSE (Fig. 6a), MAE (Fig. 6b), SSIM (Fig. 6c), and PSNR (Fig. 6d). An ablation study comparing models initialized with pretrained weights against those trained from scratch (FM-Scratch) further confirms that pretrained visual priors enhance robustness when supervision is sparse.

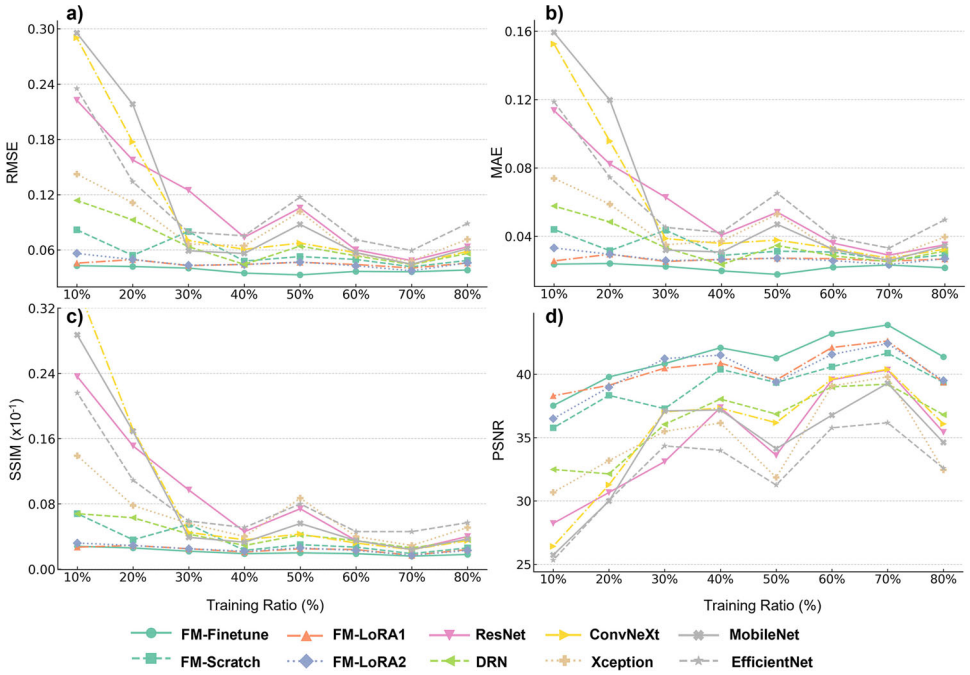

Fig. 6 | Evaluation of model performance under varying amounts of labeled training data. **a–d** Root mean square error (RMSE), mean absolute error (MAE), structural similarity (SSIM), and peak signal-to-noise ratio (PSNR) as functions of the fraction of available training slices, with the training data progressively reduced from 80% to 10%, revealing the robustness of the model under limited supervision.

The stability of these metrics reflects the model's ability to utilize pretrained visual priors to compensate for the lack of extensive annotated data. These results highlight the effectiveness of foundation-model transfer learning for generalizable seismic interpretation, and its potential in realistic conditions where dense labeling is rarely feasible.

Prompt-guided segmentation in the Parihaka dataset

To further evaluate the generalization capability of our framework across diverse depositional settings, we conduct experiments on the publicly available Parihaka 3-D seismic dataset. The dataset originates from the offshore Taranaki Basin in New Zealand, a structurally complex and hydrocarbon-rich region characterized by deepwater depositional systems. The seismic volume spans a spatial extent of approximately 15km × 20km with a vertical sampling interval of 10ms, with dimensions of 1000 × 780 × 590 samples in depth, crossline, and inline directions. This dataset includes a wide variety of structural and stratigraphic features, providing a technically challenging benchmark for seismic facies segmentation. Pixel-level annotations provide six seismic facies classes, including basement, slope mudstones, mass transport deposits, channelized bodies, and submarine canyon structures, enabling detailed characterization of subsurface heterogeneity. Additionally, we compute an RGT volume using an automatic horizon-tracking method[55], which propagates seed horizons through the seismic volume using structure-guided similarity measures and dip-coherence constraints. The resulting RGT attribute is then incorporated into our framework as a structural prior in the form of prompt embeddings, guiding the model toward stratigraphically consistent facies segmentation. The 180 sections of shape 1000 × 780 are extracted at varying azimuthal orientations and positions from these volumes. Sections are rotated at 10° increments and sampled at intervals of 16 grid units, excluding slices aligned along the inline or crossline directions. The training set comprises 163 samples, while the remaining 17 are reserved for validation.

Qualitative comparisons (Fig. 7) exhibit the ability of our framework in resolving complex stratigraphic boundaries. FM-LoRA1 (Fig. 7c) achieves superior detection of thin and discontinuous facies layers (highlighted by red arrows), owing to the geometric priors embedded during pretraining on large-scale image datasets. In contrast, FM-Scratch (Fig. 7d), which is trained from random initialization, struggles to delineate these features accurately. Quantitative metrics in Fig. 8a corroborate these observations. LoRA-based adaptations (LoRA1 and LoRA2) consistently yield high scores across all metrics, which underscores the efficiency of lightweight fine-tuning strategies. Conversely, heavyweight backbones and FM-Scratch suffer from overfitting due to the limited availability of labeled data, resulting in degraded performance on unseen sections. Although full fine-tuning (FM-Finetune) achieves marginally better results, the improvement is limited and comes at the cost of higher computational demand. Notably, LoRA achieves comparable performance with a significantly smaller model footprint.

Additionally, we train the model without structural prompts and then activate the prompting system to assess the contribution of geological guidance. Performance consistently improves across the metrics when the RGT-encoded structural embeddings are integrated alongside seismic representations. To quantify this effect, we modulate the relative contribution of structural prompts and seismic embeddings using a softmax-weighted scheme (Fig. 8b). The metrics exhibit an upward trend as the proportion of structural embeddings increases, reaching an optimum when prompts constitute approximately 70% of the total encoding. Beyond this point, further emphasis on RGT geometry leads to a performance decline. The degradation arises from the suppression of amplitude-based facies distinctions in cases with thin layers or interbedded disruptions. These underscore the value of structural priors as guiding signals in facies segmentation, but their integration needs to be balanced to preserve lithological contrast.

Although the RGT attributes incorporated as prompts encode structural cues, they remain approximations of the subsurface and may contain uncertainties. The prompting mechanism operates as a soft prior rather than a hard geological constraint. The learnable weighting enables the model to down-weight unreliable structural information and rely more strongly on the seismic image and the pretrained visual priors when inconsistencies arise. The framework exhibits stable performance and retains its ability to recover fine-scale stratigraphic details even when the structural priors are imperfect. The prompt embeddings provide beneficial contexts without compromising flexibility to allow the network to balance amplitude-driven inference with structure-aware guidance.

Foundation model-guided seismic impedance inversion

To demonstrate the applicability of our framework beyond facies segmentation, we evaluate it on the publicly available Teapot Dome dataset[51] for seismic structural interpretation and impedance inversion. Fig. 9a, b show

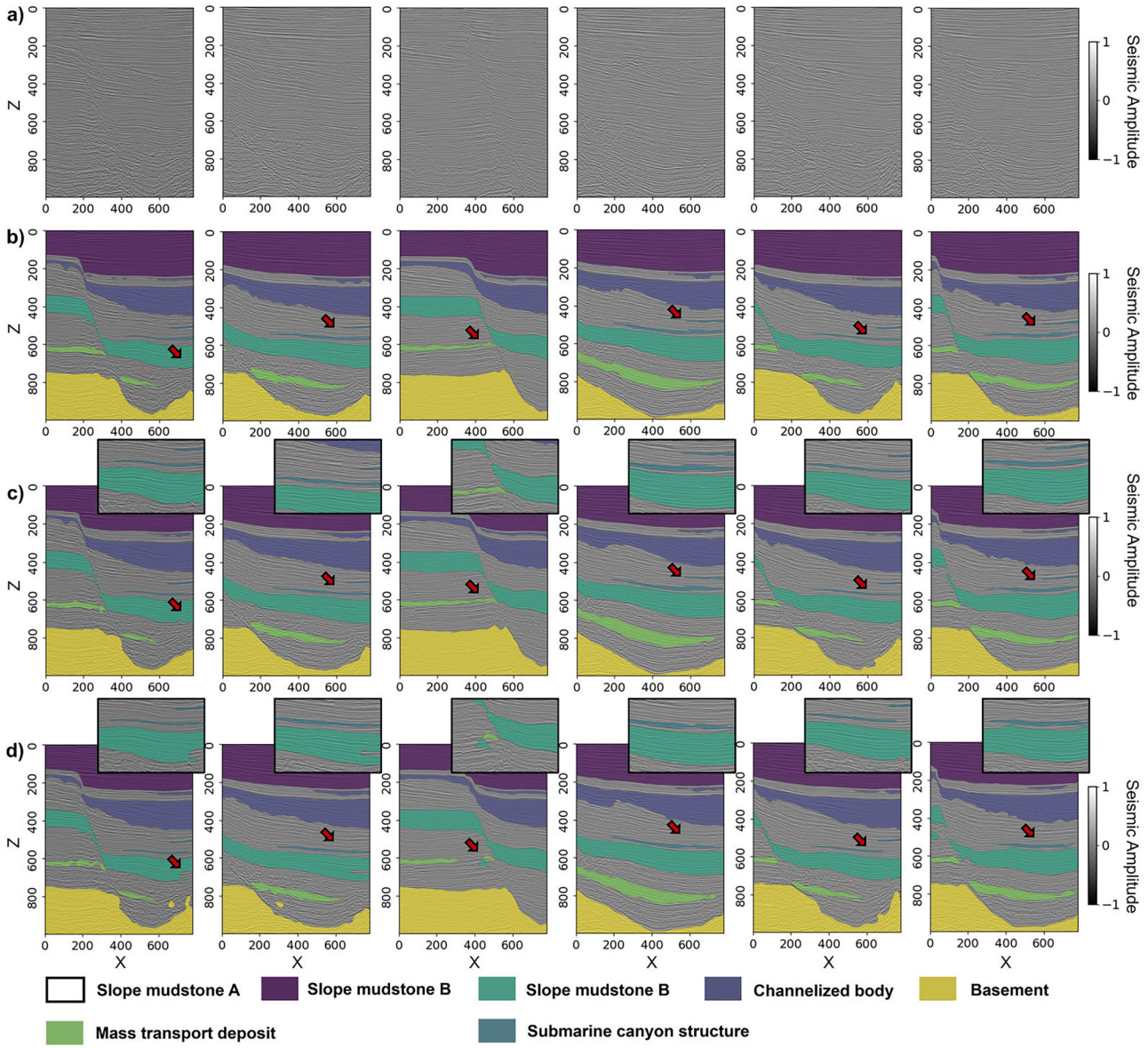

Fig. 7 | Qualitative comparisons of seismic facies segmentation on the Parihaka dataset. a Seismic section from the validation set. **b** Ground-truth facies labels with seven facies classes. **c** Prediction results from FM-LoRA1. **d** Prediction results from FM-Scratch. The black insets show zoomed-in views of the regions indicated by the arrows.

the spatial distribution of the 27 selected wells overlain on a horizontal slice of the seismic volume. Although the dataset includes hundreds of wells, only 27 are retained after discarding those lacking density or velocity logs. Among them, 23 wells (orange triangles) are used for training, while the remaining four wells (red triangles) are reserved exclusively for validation.

The 2-D training and validation datasets are constructed by randomly generating paths that intersect the well locations and then extracting the corresponding vertical seismic profiles. For training, we generated 702 random paths that traverse at least four training wells, thereby ensuring sufficient geological diversity across the survey (Fig. 9a). For validation, 98 paths are generated, each intersecting at least one validation well to enable fair comparison of impedance values. Additionally, we constructed an initial impedance model from the 23 training wells using a seismic-structure-guided interpolation method. This interpolated model is generally consistent with the measured well logs near the wells and laterally aligned with seismic reflectors. As interpolation accuracy decreases with increasing distance from the wells, we employ a distance-weighted scheme in which the interpolated values are assigned weights decaying smoothly from one at the well locations to zero in regions far away. Instead of feeding this interpolated model into the network, we use it as weak supervision, allowing the model to learn spatial dependencies and fill subsurface properties in a geologically consistent manner.

Figure 9c–e present four input seismic sections, interpolated impedance profiles, and predictions from FM-LoRA1, each corresponding to paths that intersect at least one validation well. Despite not being seen during training, the model predicts impedance variations that align with seismic structures across the entire section.

The predictions preserve thin layers where impedance values change rapidly with depth, which indicates the ability of our model to recover fine-scale heterogeneity. Figure 9f–i compare the predicted impedance traces at the four validation wells. The FM-LoRA1 (red curve) and FM-Finetune (blue curve) predictions exhibit close agreement with the measured logs (black curve) and outperform both ResNet (brown curve) and EfficientNet (purple curve) backbones. Moreover, FM-LoRA1 and FM-Finetune better capture rapid impedance fluctuations within thin stratigraphic units, which conventional backbones tend to oversmooth. Quantitative results (Table 2) also corroborate these observations. While FM-Finetune yields the highest similarity to the interpolated profiles, FM-LoRA achieves the lowest MSE and MAE when compared to the measured well-log values.

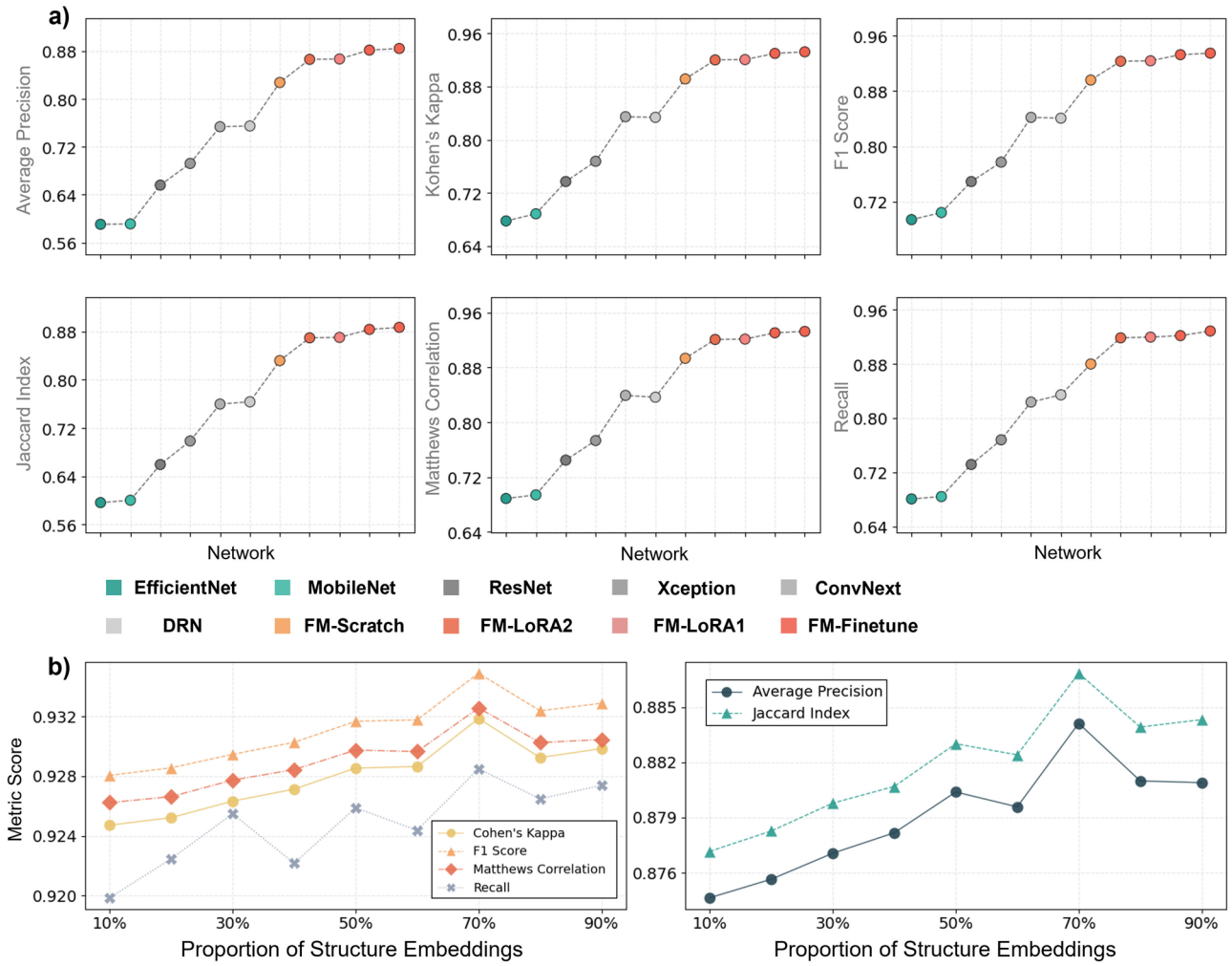

Fig. 8 | Quantitative evaluation of the proposed framework on the Parihaka seismic dataset for segmentation tasks. a Comparison of segmentation performance across adaptation strategies, including full fine-tuning and lightweight tuning, which demonstrates the effectiveness of parameter-efficient transfer. **b** Ablation analysis of stratigraphic prompting using Relative Geologic Time (RGT), showing model performance as a function of the prompt-to-seismic embedding ratio, which controls the strength of geological conditioning.

The superior performance of FM-LoRA1 relative to both scratch-trained and fully fine-tuned models suggests that our tuning strategies preserve the representational vision priors while constraining overfitting. The foundation model adaptation enables geologically consistent impedance inversion, even when only a limited number of wells are available. This highlights the potential of parameter-efficient foundation model tuning as a unifying paradigm for subsurface characterization. Additionally, because the network learns general mappings from seismic observations to subsurface properties, it is not limited to impedance. Any attribute with available well-log measurements can be inverted from a seismic image within the same framework.

Discussion

In this section, we discuss the internal representations learned by foundation models, the contributions of structural priors, and the potential for extending this paradigm toward 3-D seismic interpretation. These perspectives offer insights into the mechanisms driving the model's generalization capacity but also point to promising research directions.

Vision foundation model encodings of seismic image

Recent progress in vision foundation models has transformed our ability to extract generalized spatial features across diverse downstream tasks. Extending these models to seismic interpretation offers a promising avenue in domains where high-quality labeled data are scarce and costly to obtain. Instead of training models from scratch on seismic data, we employ the priors embedded in pretrained foundation models to extract transferable representations. The models learn to recognize textures, edges, and spatial organization, all of which are key characteristics captured in seismic images. This shared spatial inductive bias allows them to generalize beyond their original training domain. To evaluate the internal representation capacity of the vision foundation model from a feature level perspective, we apply FM-LoRA1 to seismic sections from the Parihaka dataset (Fig. 10a). Intermediate encoder features are projected into a spatial subspace using Principal Component Analysis. Mapping the first three principal components to RGB channels yields a composite visualization of the latent representations.

As shown in Fig. 10b, the FM-LoRA1 produces coherent clusters in the latent space that align with lithologic transitions or structural boundaries (marked by red arrows). These patterns emerging without explicit supervision suggest that the model has internalized a form of geologically consistent encoding through cross-domain pretraining. In contrast, FM-Scratch trained without pretraining fails to exhibit comparable structural coherence in its latent space (Fig. 10c). This demonstrates the advantage of using vision priors for seismic interpretation and reveals the limitation of models trained from scratch under scarce labeled data. The emergence of structure-aware representations in the pretrained model has implications for downstream seismic interpretation tasks. These tasks often rely on subtle stratigraphic cues embedded within seismic reflections, in which the pretrained model appears sensitive.

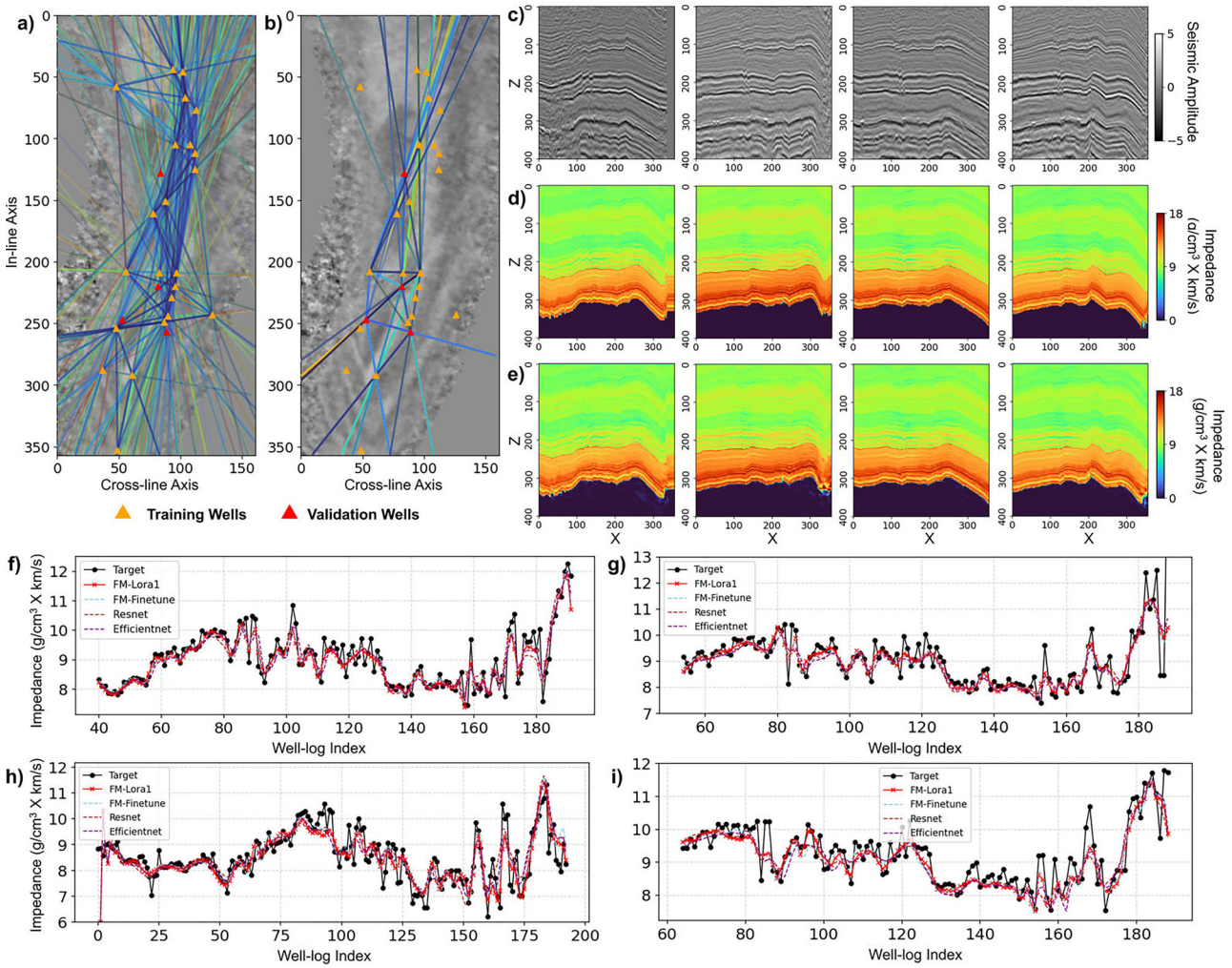

Fig. 9 | Demonstration of foundation model-based impedance inversion on the Teapot Dome dataset. a, **b** Spatial distribution of 27 wells overlaid on a horizontal slice of the seismic volume, where training wells are shown in orange and validation wells in red. **c**–**e** Example seismic sections along random paths intersecting validation wells, with corresponding interpolated models and predictions from FM-LoRA1. The model captures stratigraphic continuity and preserves fine-scale impedance contrasts that align with seismic reflectors. **f**–**i** Comparison of predicted impedance traces against measured logs at four validation wells.

Table 2 | Quantitative evaluation of impedance inversion performance across different network backbones

Backbone	Impedance Profiles (from Training Wells)					Validation Wells	
	MAE ↓	MSE ↓	PSNR (dB) ↑	RMSE ↓	SSIM ↑	MAE ↓	MSE ↓
FM-LoRA1	0.1789	0.2213	26.58	0.4591	0.7614	4.7455	**54.86**
FM-LoRA2	0.1739	0.1956	25.46	0.4337	0.7588	**4.7446**	54.89
FM-Finetune	**0.1623**	**0.1862**	**27.50**	**0.4235**	**0.7729**	4.7451	54.91
FM-Scratch	0.1893	0.2192	25.66	0.4583	0.7621	4.7519	55.06
ResNet	0.2132	0.2834	24.52	0.5218	0.7315	4.7570	54.88
DRN	0.1793	0.2426	24.83	0.4798	0.7550	4.7491	55.00
ConvNext	0.1908	0.2495	24.06	0.4880	0.7372	4.7685	55.18
Xception	0.2098	0.3270	22.02	0.5576	0.7264	4.7500	54.77
EfficientNet	0.2254	0.3822	23.57	0.6077	0.7165	4.7606	55.00
MobileNet	0.2396	0.3822	23.35	0.6102	0.7134	4.7459	55.44

Bold values indicate the best performance for each metric.
Metrics are reported for both interpolated impedance profiles derived from training wells (left) and predictions at held-out validation wells (right). Lower values indicate better performance for mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE), whereas higher values indicate better performance for peak signal-to-noise ratio (PSNR, dB) and structural similarity (SSIM).

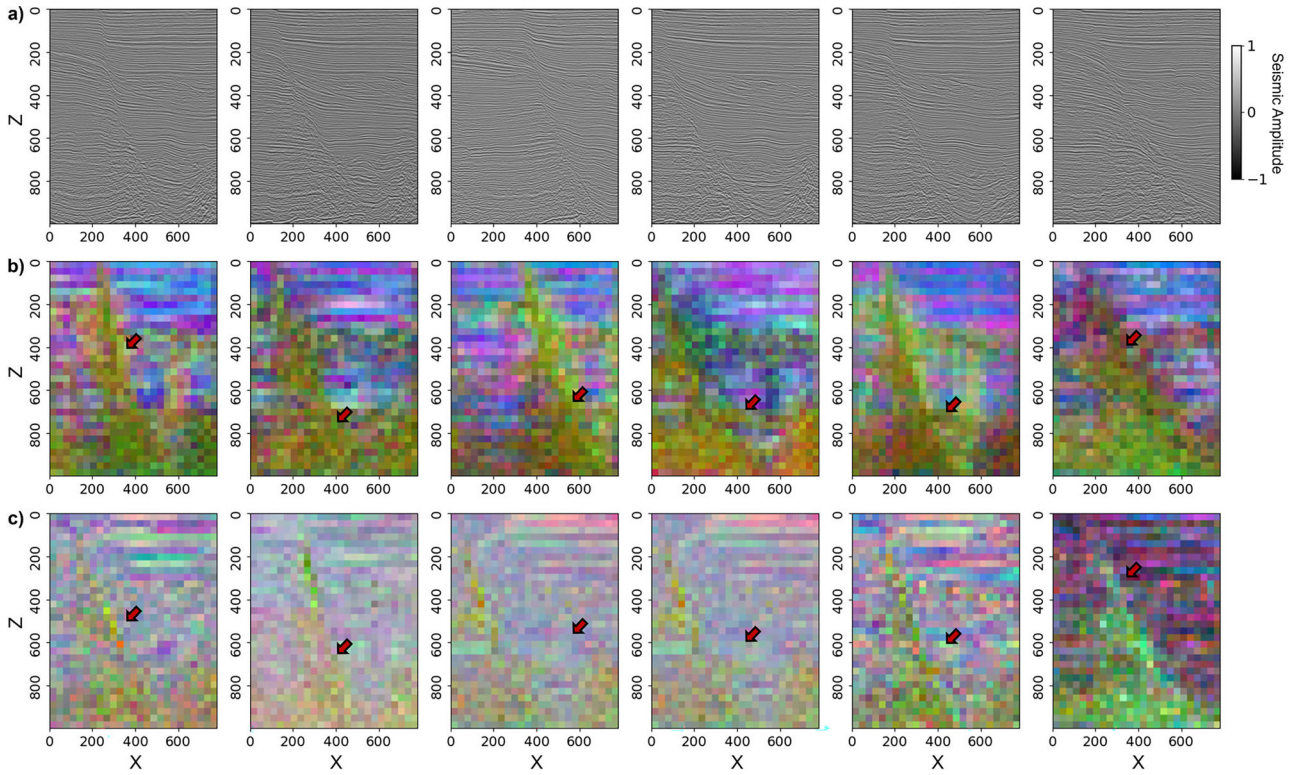

Fig. 10 | Latent representations of seismic images extracted by a vision foundation model. a Seismic sections from the Parihaka dataset. **b** Color composite visualization of principal components derived from FM-LoRA1 encoder embeddings, revealing coherent patterns aligned with seismic structures (highlighted by red arrows). **c** Latent representation from FM-Scratch, showing reduced sensitivity to seismic features.

Balancing structural priors and seismic data diversity

Incorporating structural priors through RGT-based positional encoding improves stratigraphic consistency and promotes geologically coherent predictions. However, not all facies strictly follow horizon-conformable patterns in settings where heterogeneous lithologies or depositional overprinting introduce complexity. For example, channelized deposits may crosscut or disrupt stratigraphic continuity, reducing the effectiveness of rigid horizon alignment. In such cases, enforcing strict structural conformity risks obscuring lithological variability within the same stratigraphic unit. To address this challenge, future work may further enhance the flexibility by developing attention mechanisms that dynamically modulate the influence of structural priors based on local amplitude variability and global geological consistency.

Extending foundation models to volumetric seismic data

While our framework demonstrates robust performance on 2-D seismic sections, many interpretation tasks require 3-D geometrical reasoning to capture the geometrical complexity of subsurface structures. Operating on individual sections limits the model's ability to enforce global 3-D coherence, even though the LoRA-based adaptation and decoder fusion help stabilize slice-wise predictions. To partially address this limitation, our prompting module allows volumetric attributes to be injected as spatially aligned guidance. These prompts provide stratigraphic trajectories across adjacent slices and help preserve geological consistency.

The development of 3-D foundation models remains in an early stage. Existing volumetric transformers from video understanding[56,57] or medical imaging[58] are trained on domains that differ from seismic data, and their inductive biases might not directly transfer to seismic volumes. Moreover, the datasets used to pretrain these architectures are several orders of magnitude smaller than those used to pretrain modern 2-D foundation models, which limits their ability to encode broadly generalizable spatial priors. Future research may focus on developing foundation models capable of volumetric seismic understanding through 3-D extensions of vision transformers, including spatial attention mechanisms and volumetric patch embeddings. The architectures that couple 2-D and 3-D representations also hold promise, which allows the method to employ the knowledge of 2-D pretraining while including explicit volumetric context when necessary.

Conclusions

We have introduced a generalizable transfer learning framework that adapts vision foundation models for seismic interpretation through parameter-efficient fine-tuning and geologically informed prompting. By embedding LoRA and prefix tuning into a frozen pretrained backbone, our method achieves cross-domain transfer without the prohibitive cost of full retraining, while preserving rich representational priors. A task-adaptive decoder and hierarchical structural prompting based on seismic attributes derived positional encodings enable the model to produce predictions that are stratigraphically consistent and sensitive to local seismic variability. Evaluations across multiple benchmark datasets demonstrate that the framework consistently outperforms scratch-trained and fully fine-tuned baselines while requiring fewer trainable parameters in different seismic interpretation tasks. These results highlight the potential of parameter-efficient foundation model adaptation as a geologically informed paradigm for next-generation seismic interpretation, which bridges data-driven learning with physical consistency in complex subsurface settings.

More broadly, our study emphasizes that vision foundation models can provide scalable backbones for seismic analysis even without exposure to domain data during pretraining. This ability arises from the shared spatial characteristics between natural images and seismic profiles, such as edge continuity, texture organization, and multiscale patterns. By injecting appropriate domain adaptation and structural priors, such models offer a promising path toward automating interpretation at scale, enabling improved subsurface characterization across geological settings. With continued advancements in foundation models, we envision an emerging class of models that can reason over large-scale geoscientific data with minimal supervision.

Data availability

The study uses publicly available seismic datasets and newly generated example data. The Netherlands F3 block dataset for facies segmentation benchmarking is available via Zenodo (https://zenodo.org/records/1471548)[47,48]. The Teapot Dome 3-D seismic survey is provided by the Society of Exploration Geophysicists (SEG) and can be accessed at https://wiki.seg.org/wiki/Teapot_dome_3D_survey[50]. The Parihaka 3-D seismic dataset is available from the SEG Open Data repository at https://wiki.seg.org/wiki/Parihaka-3D. The example dataset generated in this study is publicly available on Figshare with https://doi.org/10.6084/m9.figshare.30702569[59]. All data are stored in publicly accessible repositories as indicated above. Additional data and materials related to this study are available from the corresponding author upon reasonable request.

Code availability

The custom code used to reproduce the results in this study is publicly available on GitHub at https://github.com/jqfzfb/cross-domain-seismic-fm-master. No restrictions apply to the access of the code.

Received: 24 August 2025; Accepted: 17 April 2026;
Published online: 12 May 2026

References

1. Chopra, S. & Marfurt, K. J. Seismic attributes-a historical perspective. *Geophysics* **70**, 3SO–28SO (2005).
2. Posamentier, H. W., Davies, R. J., Cartwright, J. A. & Wood, L. J. in Seis- mic geomorphology – an overview (eds Davies, R. J., Posamentier, H. W., Wood, L. J. & Cartwright, J. A.) Seismic Geomorphology: Applications to Hydrocar- bon Exploration and Production, Vol. 277 of Geological Society, London, Special Publications 1–14 (The Geological Society of London, 2007).
3. Zhao, T., Jayaram, V., Roy, A. & Marfurt, K. J. A comparison of classification techniques for seismic facies recognition. *Interpretation* **3**, SAE29–SAE58 (2015).
4. Wrona, T., Pan, I., Gawthorpe, R. L. & Fossen, H. Seismic facies analysis using machine learning. *Geophysics* **83**, O83–O95 (2018).
5. Gunderson, K. L. et al. Machine learning applications to seismic structural interpretation: Philosophy, progress, pitfalls, and potential. *AAPG Bull.* **106**, 2187–2202 (2022).
6. Bergen, K. J., Johnson, P. A., de Hoop, M. V. & Beroza, G. C. Machine learning for data-driven discovery in solid earth geoscience. *Science* **363**, eaau0323 (2019).
7. Reichstein, M. et al. Deep learning and process understanding for data-driven earth system science. *Nature* **566**, 195–204 (2019).
8. Yu, S. & Ma, J. Deep learning for geophysics: current and future trends. *Rev. Geophys.* **59**, e2021RG000742 (2021).
9. Sun, Z. et al. A review of Earth artificial intelligence. *Comput. Geosci.* **159**, 105034 (2022).
10. Wu, X. et al. Sensing prior constraints in deep neural networks for solving exploration geophysical problems. *Proc. Natl. Acad. Sci.* **120**, e2219573120 (2023).
11. Perol, T., Gharbi, M. & Denolle, M. Convolutional neural network for earthquake detection and location. *Sci. Adv.* **4**, e1700578 (2018).
12. Zhu, W. & Beroza, G. C. Phasenet: a deep-neural-network-based seismic arrival-time picking method. *Geophys. J. Int.* **216**, 261–273 (2019).
13. Mousavi, S. M., Ellsworth, W. L., Zhu, W., Chuang, L. Y. & Beroza, G. C. Earthquake transformer-an attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nat. Commun.* **11**, 3952 (2020).
14. Ham, Y.-G., Kim, J.-H. & Luo, J.-J. Deep learning for multi-year ENSO forecasts. *Nature* **573**, 568–572 (2019).
15. Kadow, C., Hall, D. M. & Ulbrich, U. Artificial intelligence reconstructs missing climate information. *Nat. Geosci.* **13**, 408–413 (2020).
16. Kim, T. et al. Solar farside magnetograms from deep learning analysis of stereo/EUVI data. *Nat. Astron.* **3**, 397–400 (2019).
17. Meier, F. et al. Deep learning the collisional cross sections of the peptide universe from a million experimental values. *Nat. Commun.* **12**, 1185 (2021).
18. Wu, X., Yan, S., Bi, Z., Zhang, S. & Si, H. Deep learning for multidimensional seismic impedance inversion. *Geophysics* **86**, R735–R745 (2021).
19. Bi, Z., Wu, X., Li, Z., Chang, D. & Yong, X. Deepismnet: Three-dimensional implicit structural modeling with convolutional neural network. *Geosci. Model Dev. Discuss.* **2022**, 1–28 (2022).
20. Bi, Z. et al. Geologic-time-based interpolation of borehole data for building high-resolution models: Methods and applications. *Geophysics* **87**, IM67–IM80 (2022).
21. Wu, X. et al. Building realistic structure models to train convolutional neural networks for seismic structural interpretation. *Geophysics* **85**, WA27–WA39 (2020).
22. Bi, Z., Wu, X., Geng, Z. & Li, H. Deep relative geologic time: A deep learning method for simultaneously interpreting 3-d seismic horizons and faults. *J. Geophys. Res.: Solid Earth* **126**, e2021JB021882 (2021).
23. Geng, Z., Wu, X., Shi, Y. & Fomel, S. Relative geologic time estimation using a deep convolutional neural network. In *SEG International Exposition and Annual Meeting* D033S038R001 (SEG, 2019).
24. Zhuang, F. et al. A comprehensive survey on transfer learning. *Proc. IEEE* **109**, 43–76 (2020).
25. Alayrac, J.-B. et al. Flamingo: a visual language model for few-shot learning. *Adv. neural Inf. Process. Syst.* **35**, 23716–23736 (2022).
26. Gallifant, J. et al. Peer review of gpt-4 technical report and systems card. *PLOS digital health* **3**, e0000417 (2024).
27. Singhal, K. et al. Large language models encode clinical knowledge. *Nature* **620**, 172–180 (2023).
28. Oquab, M. et al. Dinov2: learning robust visual features without supervision. *arXiv preprint* https://doi.org/10.48550/arXiv.2304.07193 (2023).
29. Kirillov, A. et al. Segment anything, 4015–4026 (IEEE, 2023).
30. Bommasani, R. et al. On the opportunities and risks of foundation models. *arXiv preprint* https://doi.org/10.48550/arXiv.2108.07258 (2021).
31. Zhou, C. et al. A comprehensive survey on pretrained foundation models: A history from bert to ChatGPT. *arXiv preprint* https://doi.org/10.48550/arXiv.2302.09419 (2023).
32. Nguyen, T., Brandstetter, J., Kapoor, A., Gupta, J. K. & Grover, A. ClimaX: a foundation model for weather and climate. *arXiv preprint* https://doi.org/10.48550/arXiv.2301.10343 (2023).
33. Chen, K. et al. Rsprompter: Learning to prompt for remote sensing instance seg- mentation based on visual foundation model. *IEEE Transactions on Geoscience and Remote Sensing* **62**, 1–17 (2024).
34. Bi, Z., Wu, X., Xu, X. & Nakata, N. Global detection and morphological characterization of seamounts with weakly supervised deep learning. *J. Geophys. Res. Mach. Learn. Comput*. **3**, e2025JH000848 (2025).
35. Chen, R., Zhang, Z. & Ma, J. Seismic fault sam: adapting sam with lightweight modules and 2.5 d strategy for fault detection. In *Proc. IEEE 17th International Conference on Signal Processing (ICSP)* 436-441 (IEEE, 2024).
36. Atolagbe, J. & Koeshidayatullah, A. Towards user-guided seismic facies interpretation with a pre-trained large vision model. *IEEE Access* **13**, 42965–42976 (2025).
37. Hu, E. J. et al. Lora: Low-rank adaptation of large language models. *ICLR* **1**, 3 (2022).
38. Li, X. L. & Liang, P. Prefix-tuning: Optimizing continuous prompts for generation. *arXiv preprint arXiv* https://doi.org/10.48550/arXiv.2101.00190 (2021).
39. Cheng, B., Schwing, A. & Kirillov, A. Per-pixel classification is not all you need for semantic segmentation. *Adv. neural Inf. Process. Syst.* **34**, 17864–17875 (2021).

40. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition, 770–778 (IEEE, 2016).
41. Xie, S., Girshick, R., Dollár, P., Tu, Z. & He, K. Aggregated residual transformations for deep neural networks. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 1492–1500 (2017).
42. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 1251–1258 (2017).
43. Liu, Z. et al. A convnet for the 2020s. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition* 11976–11986 (2022).
44. Zheng, G. et al. DRN: A deep reinforcement learning framework for news recommendation, 167–176 (ACM, 2018).
45. Tan, M. & Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning* 6105–6114 (PMLR, 2019).
46. Sinha, D. & El-Sharkawy, M. Thin mobilenet: An enhanced mobilenet architecture. In *Proc. IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)* 0280–0285 (IEEE, 2019).
47. Alaudah, Y., Michałowicz, P., Alfarraj, M. & AlRegib, G. A machine-learning benchmark for facies classification. *Interpretation* **7**, SE175–SE187 (2019).
48. Silva, R. M. et al. Netherlands dataset: a new public dataset for machine learning in seismic interpretation. *arXiv preprint* https://doi.org/10.48550/arXiv.1904.00770 (2019).
49. Tolstaya, E. & Egorov, A. Deep learning for automated seismic facies classification. *Interpretation* **10**, SC31–SC40 (2022).
50. Chiaramonte, L., Zoback, M. D., Friedmann, J. & Stamp, V. Seal integrity and feasibility of co 2 sequestration in the Teapot Dome EOR pilot: geomechanical site characterization. *Environ. Geol.* **54**, 1667–1675 (2008).
51. Anderson, T. AV history of geologic investigations and oil operations at Teapot Dome, Wyoming. *AAPG Search and Discovery Article* 20215 (2009).
52. Stark, T. J. Relative geologic time (age) volumes-relating every seismic sample to a geologically reasonable horizon. *Lead. Edge* **23**, 928–932 (2004).
53. Stark, T. J. System for multi-dimensional data analysis. US Patent 6,850,845 (2005).
54. Stark, T. *Visualization Techniques for Enhancing Stratigraphic Inferences from 3D Seismic Data Volumes*. First Break 24 (European Association of Geoscientists & Engineers, 2006).
55. Wu, X. & Fomel, S. Least-squares horizons with local slopes and multigrid correlations. *Geophysics* **83**, IM29–IM40 (2018).
56. Li, K. et al. Unmasked teacher: Towards training-efficient video foundation models, 19948–19960 (IEEE, 2023).
57. Wang, Y. et al. Internvideo2: Scaling foundation models for multimodal video understanding. In *European Conference on Computer Vision* 396–416 (Springer, 2024).
58. Shen, Y. et al. Fastsam3d: An efficient segment anything model for 3d volumetric medical images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* 542–552 (Springer, 2024).
59. Bi, Z., Nakata, N. & Wu, X. Example Data Supporting a Cross-Domain Transfer Learning Framework for Seismic Interpretation. https://figshare.com/articles/dataset/Example_Data_Supporting_a_Cross-Domain_Transfer_Learning_Framework_for_Seismic_Interpretation/30702569 (2025).

Acknowledgements

This work was supported by the National Key R&D Program of China (Grant No. 2024YFC3012803-05), the National Science Foundation of China under Grant 42374127 and the U.S. Department of Energy under Award Number DE-AC02-5CH11231.

Author contributions

Z.B. led and conceived the research, developed the methodology, implemented the framework, conducted the experiments, and wrote the original manuscript. X.W. supervised the research, contributed to conceptual development, revised the manuscript, and served as the corresponding author. N.C. assisted with data processing and experimental evaluation. N.N. contributed to interpretation of results and manuscript revision. All authors discussed the results and contributed to the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s44172-026-00674-9.

Correspondence and requests for materials should be addressed to Xinming Wu.

Peer review information *Communications Engineering* thanks Siwei Yu and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: [Philip Coatsworth]. A peer review file is available.