

# Interactive Identification of Seismic Faults with the Segment Anything Model

Mengchao Zhao<sup>1</sup>, Zhonghua Ma<sup>1,2\*</sup>

<sup>1</sup>School of Science, Tianjin University of Technology and Education, Tianjin, China

<sup>2</sup>School of Big Data, Lvliang Vocational and Technical College, Lvliang, China

Email: \*mazh@tute.edu.cn

**How to cite this paper:** Zhao, M.C. and Ma, Z.H. (2025) Interactive Identification of Seismic Faults with the Segment Anything Model. *Journal of Applied Mathematics and Physics*, **13**, 4224-4233. <https://doi.org/10.4236/jamp.2025.1312234>

**Received:** November 25, 2025

**Accepted:** December 9, 2025

**Published:** December 12, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

Accurate identification of seismic faults is critical to seismology, geological modeling, and disaster assessment. Traditional manual interpretation entails heavy reliance on domain expertise and suffers from inefficiency and subjectivity. Although deep learning-based segmentation has advanced the field, these methods remain limited by a dependence on scarce high-quality labeled data and a lack of generalization flexibility. The Segment Anything Model (SAM), a powerful vision foundation model, offers a breakthrough with its robust zero-shot transfer capabilities and prompt-driven mechanism. This study investigates the application of SAM to seismic fault interpretation. We propose an interactive prompting strategy tailored to the linear and discontinuous nature of faults, enabling users to guide the model in accurately segmenting complex fault structures via sparse prompts, such as points or bounding boxes. Experimental results indicate that this approach significantly lowers the technical barrier and reduces the time required for interpretation. By enhancing both flexibility and efficiency without compromising accuracy, this method presents a promising new solution for intelligent seismic interpretation.

## Keywords

SAM, Earthquake Faults, Interactive Identification

## 1. Research Background

Accurate identification of seismic faults is fundamental to seismic interpretation and geological modeling. The technological trajectory of this field has evolved from manual interpretation to intelligent, interactive systems. Historically, fault identification relied exclusively on expert visual interpretation, analyzing features such as phase axis misalignment, torsion, and abrupt amplitude changes, or trac-

ing linear discontinuities within coherence and curvature volumes [1]. While direct, this approach is plagued by low efficiency, subjectivity, and poor repeatability, failing to meet the processing demands of massive 3D seismic datasets. To enhance objectivity and efficiency, researchers adopted traditional image processing and machine learning techniques, employing Canny and Sobel operators for edge detection, or utilizing ant colony algorithms and Hough transforms to extract linear features [2]. Subsequently, machine learning models such as Support Vector Machines (SVM) and Random Forests enabled semi-automatic identification by classifying image blocks based on handcrafted features like texture and grayscale [3]. However, the performance of these methods is constrained by the quality of handcrafted features; they remain sensitive to noise and lack the generalization capability required to characterize complex subsurface structures.

The advent of deep Convolutional Neural Networks (CNNs) revolutionized fault identification. Specifically, semantic segmentation models with encoder-decoder architectures, such as U-Net and SegNet, enable end-to-end learning of deep features from raw seismic data, achieving pixel-level prediction and significantly advancing automation [4] [5]. Recent studies have further refined these models through attention mechanisms and multi-scale feature fusion [6]. Despite these advances, deep learning approaches face inherent limitations. First, they rely heavily on extensive high-quality, pixel-level labeled data, the generation of which is prohibitively expensive and labor-intensive. Second, these models typically function as “black boxes” with static behaviors. They often lack flexibility when applied to unseen survey data or novel fault patterns, and they offer limited interactivity, preventing users from incorporating expert knowledge to correct segmentation results in real-time.

Recently, breakthroughs in vision foundation models have introduced a new paradigm to address these challenges. Most notably, the Segment Anything Model (SAM) released by Meta offers powerful zero-shot generalization and a flexible, prompt-driven segmentation mechanism, derived from training on a dataset exceeding 1 billion masks [7]. Users can intuitively guide the model using sparse prompts, such as points or bounding boxes. SAM facilitates the development of solutions that minimize reliance on task-specific labeled data while combining the high precision of deep learning with expert prior knowledge. In this context, this study systematically investigates the application of SAM to seismic fault identification. By designing an interaction strategy tailored to the specific characteristics of seismic data, we propose a novel, efficient, and interactive method that overcomes the critical bottlenecks of data dependency and insufficient flexibility.

Between 2021 and 2024, the field witnessed further advancements addressing the limitations of standard CNNs, particularly their inability to capture long-range dependencies. Researchers began integrating attention mechanisms and Transformer architectures into seismic interpretation. For instance, modified U-Net architectures incorporating residual attention modules and Vision Transformers (ViT) have shown improved performance in distinguishing continuous

fault geometries from stratigraphic noise by modeling global context [8] [9]. Furthermore, to mitigate the chronic bottleneck of label scarcity, recent studies have pivoted toward semi-supervised and transfer learning approaches. These methods typically leverage models pre-trained on massive synthetic geological datasets, which are then adapted to real-world seismic volumes with minimal supervision [10] [11].

## **2. Method**

### **2.1. Seismic Data Characteristics and Fault Interpretation**

Seismic exploration delineates subsurface geology by acquiring reflected wave signals from subsurface media, typically presenting the data as 2D profiles or 3D volumes. Morphologically, continuous seismic reflection events correspond to stratigraphic interfaces, whereas faults are manifested as discontinuities, distortions, abrupt terminations, or sudden changes in dip. To enhance the visibility of these structural features, various seismic attributes are derived from the raw data. The coherence (or similarity) attribute is particularly critical; by quantifying the discontinuity between adjacent traces, it renders faults as clear linear low-coherence anomalies on time slices, significantly aiding interpretation. Fundamentally, fault interpretation is a semantic segmentation task aimed at isolating fault structures from the background. However, this process faces significant challenges, including varying signal-to-noise ratios (SNR) that blur fault signatures, the geometric complexity of faults (varying scales and dips), and the difficulty of differentiating faults from other linear stratigraphic features like channels or coherent noise.

### **2.2. Deep Learning for Semantic Segmentation**

Image segmentation, a fundamental task in computer vision, involves partitioning an image into semantically meaningful regions. This study focuses on semantic segmentation, specifically the pixel-wise classification of seismic images into categories such as “fault” and “non-fault.” Convolutional Neural Networks (CNNs) have established themselves as the dominant architecture for processing such data. Through layers of convolution and pooling, CNNs automatically extract hierarchical representations, transitioning from low-level edges to high-level semantic features. The Encoder-Decoder architecture, epitomized by U-Net, is the standard for such tasks. The encoder serves as a downsampling path to capture global contextual information by compressing features, while the decoder functions as an upsampling path to recover spatial resolution, enabling precise pixel-level localization and classification.

### **2.3. Vision Foundation Models and the SAM Architecture**

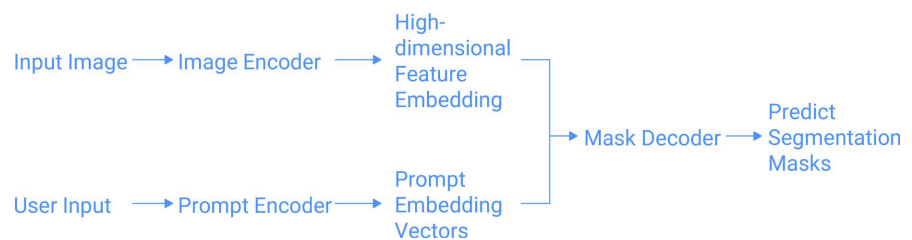
A “foundation model” is defined as a model trained on broad data at scale that can be adapted to a wide range of downstream tasks via prompting, often without the need for fine-tuning. While the GPT series exemplifies this in natural language

processing, the Segment Anything Model (SAM) marks a parallel paradigm shift in computer vision.

Trained on the SA-1B dataset comprising over 1 billion masks, SAM possesses robust prior knowledge for general object segmentation. Its core mechanism is prompt-driven interactive segmentation. Users guide the model to generate masks via sparse prompts, such as positive/negative points, bounding boxes, or coarse masks. A positive point signals the target object, while a negative point excludes specific regions (**Figure 1**). Architecturally, SAM comprises three components: 1) An Image Encoder, based on a Vision Transformer (ViT), which maps the image into a high-dimensional feature embedding. Unlike traditional CNNs, ViT leverages the self-attention mechanism to capture global contextual information, formulated as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)$$

where  $Q$ ,  $K$ , and  $V$  are the query, key, and value matrices, and  $d_k$  is the dimension of the keys. 2) A Prompt Encoder, which converts user inputs (points, boxes) into embedding vectors; and 3) A lightweight Mask Decoder, which efficiently fuses the image and prompt embeddings to predict final segmentation masks in real-time.



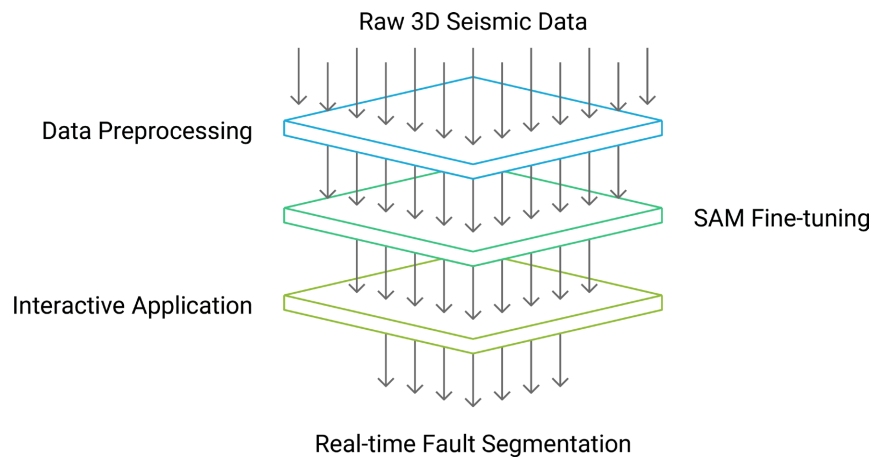
**Figure 1.** SAM model flowchart.

### 3. Experimental Procedure

As illustrated in **Figure 2**, the workflow consists of three stages: (A) Data Preprocessing: Seismic slices and U-Net derived labels are processed via separate pipelines involving resizing, normalization, and morphological filtering to generate standard training pairs ( $1024 \times 1024$ ). (B) Model Fine-tuning: The SAM architecture is adapted for seismic data. The Image Encoder and Prompt Encoder are frozen to retain pre-trained knowledge, while the Mask Decoder is fine-tuned using box prompts with coordinate perturbations, optimized by a combination of Dice and BCE loss. (C) Interactive Application: A PyQt5-based tool integrates the fine-tuned model, allowing users to extract faults in real-time by drawing bounding boxes.

#### 3.1. Dataset Acquisition and Preprocessing

The dataset is selected from public datasets. Preprocessing was conducted separately for seismic images and annotation masks to ensure data quality. For image



**Figure 2.** Seismic data processing workflow.

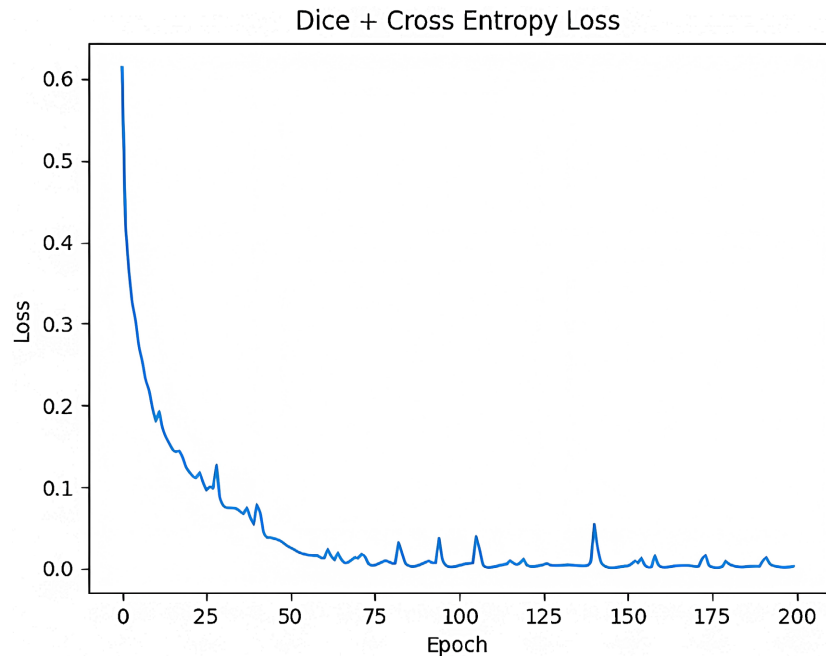
data, raw inputs are first converted to floating-point format. We apply bicubic interpolation to standardize the resolution to  $1024 \times 1024$  and employ stacking or truncation to ensure a three-channel RGB format. To improve feature visibility, histogram equalization is applied to low-contrast images, followed by intensity normalization to the  $[0, 1]$  range and outlier removal. The processed images are finally stored as float 32 tensors.

For annotation masks, we implemented a standardized pipeline to adapt various label formats for deep learning. An automated parsing mechanism handles diverse inputs: binary masks are standardized to foreground/background; multi-class labels preserve class IDs; and RGB pseudo-color labels are converted to grayscale to extract semantic indices. The pipeline consists of four key steps: 1) Resizing to  $1024 \times 1024$  using nearest-neighbor interpolation to preserve boundary integrity; 2) Morphological filtering to eliminate noise (regions  $<100$  pixels) and fill holes ( $<200$  pixels); 3) Casting to uint8 for storage efficiency; and 4) Rigorous quality assurance to verify dimensions, data types, and label validity.

### 3.2. Training Strategy

The training framework employs a fine-tuning strategy for the SAM model based on bounding box prompts. The model is initialized with pre-trained weights from the ViT-B architecture. The training process was conducted over 200 epochs, with a batch size of 2 and a learning rate of 0.0001. Methodologically, a custom data loader was constructed to randomly select individual fault instances from .npy files containing  $1024 \times 1024$  images and  $256 \times 256$  annotations to generate binary masks. To enhance data diversity, bounding box coordinates were randomly perturbed. Regarding the model architecture, the prompt encoder was frozen, while only the image encoder and mask decoder were trained. The optimization objective combined Dice loss and binary cross-entropy loss. Mixed-precision training was utilized to accelerate computation, accompanied by a comprehensive checkpoint saving mechanism. **Figure 3** illustrates the evolution of Dice loss and cross-entropy loss during the training process, which were monitored to ensure conver-

gence. The entire process focused on enabling the model to accurately segment specific fault structures in seismic images based on box prompts, thereby adapting it to downstream interpretation tasks.



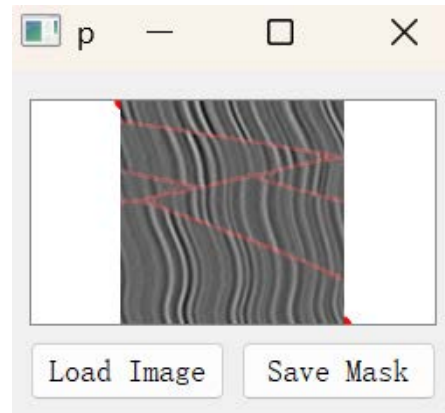
**Figure 3.** Dice and cross entropy loss.

### 3.3. Implementation of the Interactive Interpretation Tool

The proposed interactive tool facilitates fault identification by leveraging the weights optimized during the training phase. Built upon the PyQt5 framework, the system integrates the fine-tuned SAM model to enable precise, bounding box-guided segmentation. By bridging state-of-the-art segmentation algorithms with a user-friendly Graphical User Interface (GUI), the tool significantly streamlines the annotation workflow and enhances accuracy. The system architecture is organized into three layers: an interaction layer for visual feedback, a model inference layer encapsulating the SAM network, and a data preprocessing layer for standardization. Upon initialization, the system loads the ViT-B based SAM model, which has been fine-tuned on seismic datasets to robustly characterize complex fault structures.

The workflow begins with loading a seismic profile via the file dialog. The system automatically initiates the feature extraction pipeline: the image is resized to a resolution of  $1024 \times 1024$ , normalized, and processed by the Image Encoder to generate high-dimensional feature embeddings. Crucially, this computationally intensive step is performed only once, enabling real-time responsiveness for subsequent interactions. Users define segmentation targets by drawing bounding boxes (prompts). The system renders these boxes in red for immediate visual confirmation. Upon mouse release, the coordinates are normalized to the model's input scale, triggering the inference engine (**Figure 4**). The Prompt Encoder maps

the bounding box to sparse embeddings, which are then fused with the pre-computed image embeddings by the Mask Decoder to predict pixel-level fault masks instantly. the initial segmentation result. Finally, the  $256 \times 256$  low-resolution mask is upsampled to the original image size using bilinear interpolation and binarized with a threshold of 0.5. For visualization, a semi-transparent overlay technique is used, blending the color mask with the original image at 20% transparency, clearly displaying the segmented region while preserving the details of the underlying image.



**Figure 4.** Screenshot of the interactive fault identification tool. Users guide the segmentation by drawing a box around the target area. The red bounding box indicates the user's prompt, and the red lines within the image show the real-time SAM segmentation results.



**Figure 5.** Seismic profile, mask and SAM segmentation results.

As shown in **Figure 5**, the first image represents the input seismic profile. Through interactive bounding box prompting, the SAM model produces the fault mask displayed in the third image, which serves as the prediction result. For comparison, the second image shows the ground truth label. It can be visually observed that the segmented image is highly consistent with the label.

To rigorously quantify this performance, we performed a Seismic Profile Similarity Analysis between the SAM-generated mask and the ground truth label, evaluating metrics such as Structural Similarity (SSIM) and Pixel Accuracy (**Table 1**). The Structural Similarity (SSIM) score of 0.9808 is exceptional, approaching a perfect match. This demonstrates that the annotated map maintains high consistency with the original seismic data in terms of spatial layout and fault mor

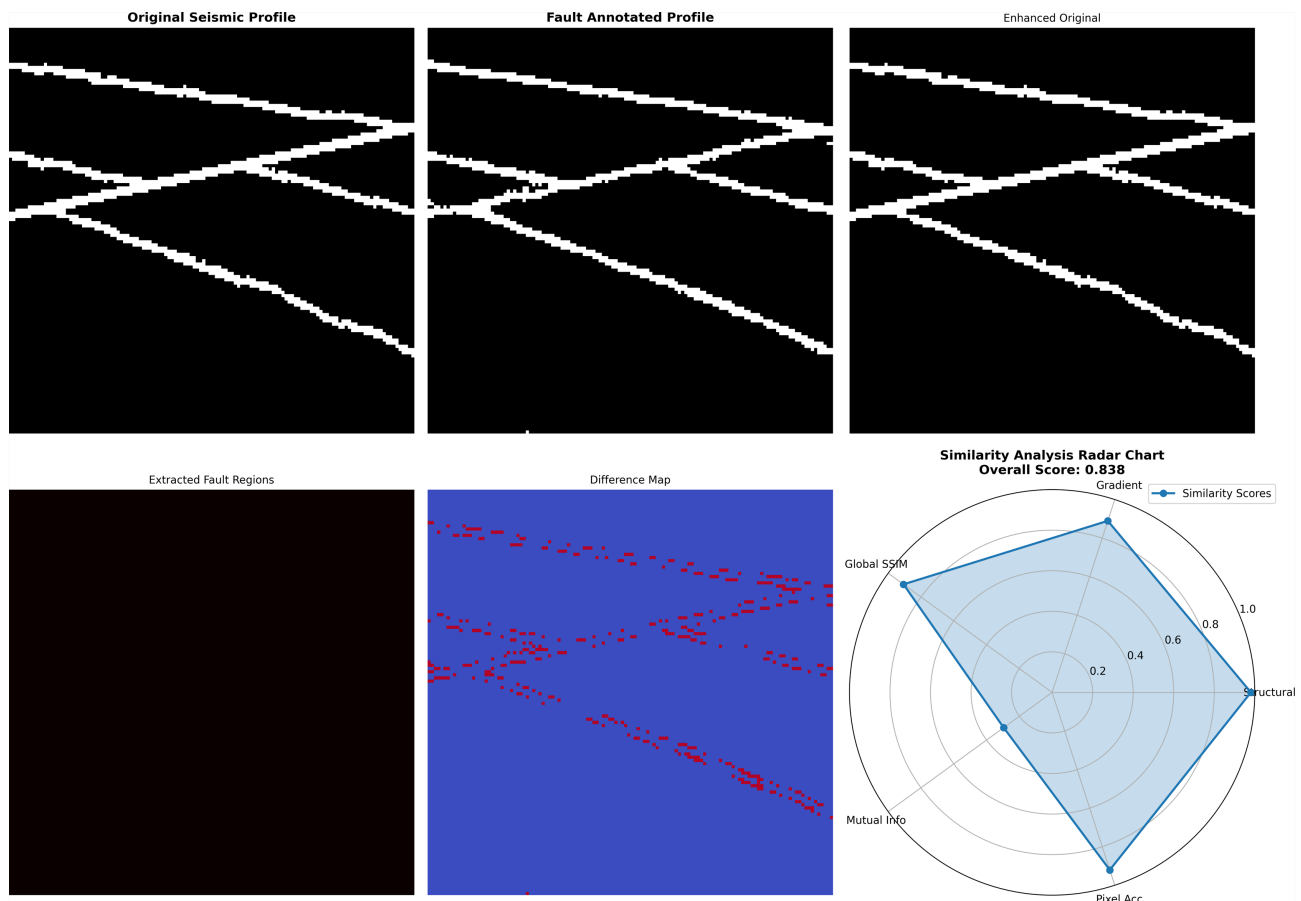
**Table 1.** Seismic profile similarity analysis report.

| Evaluation Dimensions    | Indicator Name        | Score  | Weight | Weighted score | Rating          |
|--------------------------|-----------------------|--------|--------|----------------|-----------------|
| Comprehensive assessment | Overall similarity    | 0.8380 | 1      | 0.8380         | High similarity |
| Structural features      | Structural Similarity | 0.9808 | 0.35   | 0.3433         | excellent       |
| Gradient features        | Gradient Similarity   | 0.8894 | 0.25   | 0.2224         | good            |
| global structure         | Global SSIM           | 0.9066 | 0.15   | 0.1360         | excellent       |
| Pixel precision          | Pixel Accuracy0.      | 0.9206 | 0.10   | 0.0921         | excellent       |

phology. Crucially, key geological elements—such as fault strikes and propagation patterns—are accurately preserved, representing the most vital performance metric in this evaluation. Furthermore, the Gradient Similarity (0.8894) and Global Structural Similarity (0.9066) scores are excellent, reflecting high fidelity in preserving texture details, stratigraphic continuity, and subtle local features. Essential seismic attributes, including reflection events and amplitude variations, remain intact. The high Pixel Accuracy (0.9206) confirms precise pixel-level correspondence in most regions, verifying the effectiveness of the annotation quality control. The only outlier is the relatively low Mutual Information (0.2956) score (Figure 6). This suggests that despite the structural and visual alignment, there are divergences in statistical distribution. This is attributed to the introduction of discrete segmentation labels (e.g., color masks and boundary lines), which inevitably alter the continuous grayscale distribution of the original seismic data. Such statistical shifts are an inherent consequence of the annotation process and are considered acceptable for the purpose of fault interpretation.

The SAM model demonstrates substantial technical superiority in seismic fault identification. By leveraging interactive prompts—such as points and bounding boxes—the model rapidly and accurately localizes fault zones, drastically enhancing efficiency. In contrast to traditional methods, SAM exhibits robust generalization capabilities; its zero-shot segmentation ability transfers effectively to seismic data, suggesting the model has encoded universal representations of edges and geometric shapes. Experimental results validate its adaptability, achieving high-precision segmentation that accurately captures linear fault features. Notably, it delivers pixel-level precision in critical regions characterized by the discontinuity and distortion of seismic reflection events, maintaining high fidelity even amidst complex geological backgrounds.

In practical application, SAM fundamentally alleviates the burden of manual interpretation. Whereas traditional workflows necessitate the laborious tracing of faults, SAM's interactive paradigm enables the automatic delineation of complete fault structures via sparse prompts. This efficiency offers profound value for structural analysis in geophysical exploration. Nevertheless, certain limitations persist in complex scenarios. The model's sensitivity to subtle or concealed fault



**Figure 6.** Fault detection validation.

signatures requires further refinement, and boundary definition remains challenging within intricate fault intersection zones.

#### 4. Conclusion and Outlook

In conclusion, SAM presents a novel AI paradigm for seismic fault identification that effectively synergizes deep learning capabilities with human expertise. To maximize practical utility, we advocate a workflow that combines “AI-driven preliminary screening with manual verification”, in a typical 3D seismic volume interpretation scenario, the interpreter would first allow the model to automatically process slices with high confidence. For complex or ambiguous regions, the interpreter would step in to provide sparse bounding box prompts on key 2D profiles. The model then propagates these refined features to adjacent slices, effectively mapping the entire 3D fault system. This “human-in-the-loop” approach ensures that expert geological insight guides the AI, balancing operational efficiency with interpretative precision. Nevertheless, certain limitations persist. A primary challenge in seismic interpretation is the presence of high-amplitude coherent noise and low Signal-to-Noise Ratio (SNR) zones, which can obscure fault evidence. Future efforts focusing on domain-adaptive training and robust noise-suppression pre-training promise to further refine the model’s sensitivity to these subtle

seismic features, extending its potential within geophysical exploration.

## Funding

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

## Conflicts of Interest

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

## References

- [1] Brown, A.R. (2011) Interpretation of Three-Dimensional Seismic Data. SEG Books.
- [2] Aqrabi, A.A. and Boe, T.H. (2011). Improved Fault Segmentation Using a Dip Guided and Modified 3D Sobel Filter. *SEG Technical Program Expanded Abstracts*, **30**, 999-1003. <https://doi.org/10.1190/1.3628241>
- [3] Di, H., Gao, D. and AlRegib, G. (2019) Developing a Seismic Texture Analysis Neural Network for Machine-Aided Seismic Interpretation. *Leading Edge*.
- [4] Wu, X., Liang, L., Shi, Y. and Fomel, S. (2019) FaultSeg3D: Using Synthetic Data Sets to Train an End-to-End Convolutional Neural Network for 3D Seismic Fault Segmentation. *Geophysics*, **84**, IM35-IM45. <https://doi.org/10.1190/geo2018-0646.1>
- [5] Zhao, T. and Zhang, J. (2020) Seismic Fault Detection with a Deep Convolutional Neural Network. *Geophysics*, **85**, O67-O77.
- [6] Huang, L., Dong, X. and Clee, T.E. (2017) A Scalable Deep Learning Platform for Modern Geologic Interpretation. *SEG Technical Program Expanded Abstracts*, No. 1, 2037-2040.
- [7] Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., *et al.* (2023) Segment Anything. 2023 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Paris, 1-6 October 2023, 4015-4026. <https://doi.org/10.1109/iccv51070.2023.00371>
- [8] Liu, K., Li, J. and Chen, X. (2022) Seismic Fault Detection Using U-Net with Attention Mechanism. *IEEE Geoscience and Remote Sensing Letters*, **19**, 1-5.
- [9] Yu, B., Wang, Y. and Liu, H. (2023) Seismic Fault Detection Based on Swin Transformer U-Net. *IEEE Transactions on Geoscience and Remote Sensing*, **61**, 1-12.
- [10] Dou, Y., Li, S. and Song, H. (2022) Semi-Supervised Learning for Seismic Fault Detection with Consistent Regularization. *Geophysics*, **87**, IM67-IM78.
- [11] An, S., Liu, L. and Chen, Y. (2023) Unsupervised Domain Adaptation for Seismic Fault Detection from Synthetic to Real Data. *Computers & Geosciences*, **171**, Article ID: 105268.