electronics

MDPI

Article

Advancing 3D Seismic Fault Identification with SwiftSeis-AWNet: A Lightweight Architecture Featuring Attention-Weighted Multi-Scale Semantics and Detail Infusion

Ang Li, Rui Li, Yuhao Zhang, Shanyi Li, Yali Guo, Liyan Zhang * and Yuqing Shi

Petroleum Institute, China University of Petroleum-Beijing at Karamay, Karamay 834000, China; 2023592101@cupk.edu.cn (A.L.); 2022015530@st.cupk.edu.cn (R.L.); 2022015524@st.cupk.edu.cn (Y.Z.); 2022015177@st.cupk.edu.cn (S.L.); 2022015529@st.cupk.edu.cn (Y.G.); 2022015401@st.cupk.edu.cn (Y.S.)
* Correspondence: 2023592102@cupk.edu.cn

Abstract

The accurate identification of seismic faults, which serve as crucial fluid migration pathways in hydrocarbon reservoirs, is of paramount importance for reservoir characterization. Traditional interpretation is inefficient. It also struggles with complex geometries, failing to meet the current exploration demands. Deep learning boosts fault identification significantly but struggles with edge accuracy and noise robustness. To overcome these limitations, this research introduces SwiftSeis-AWNet, a novel lightweight and high-precision network. The network is based on an optimized MedNeXt architecture for better fault edge detection. To address the noise from simple feature fusion, a Semantics and Detail Infusion (SDI) module is integrated. Since the Hadamard product in SDI can cause information loss, we engineer an Attention-Weighted Semantics and Detail Infusion (AWSDI) module that uses dynamic multi-scale feature fusion to preserve details. Validation on field seismic datasets from the Netherlands F3 and New Zealand Kerry blocks shows that SwiftSeis-AWNet mitigates challenges like the loss of small-scale fault features and misidentification of fault intersection zones, enhancing the accuracy and geological reliability of automated fault identification.

Keywords: fault identification; deep learning; MedNeXt; edge detection; SwiftSeis-AWNet

check for updates

Academic Editor: Ping-Feng Pai

Received: 6 July 2025
Revised: 24 July 2025
Accepted: 27 July 2025
Published: 31 July 2025

Citation: Li, A.; Li, R.; Zhang, Y.; Li, S.; Guo, Y.; Zhang, L.; Shi, Y. Advancing 3D Seismic Fault Identification with SwiftSeis-AWNet: A Lightweight Architecture Featuring Attention-Weighted Multi-Scale Semantics and Detail Infusion. *Electronics* **2025**, *14*, 3078. https://doi.org/10.3390/electronics14153078

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Traditional methods identify faults by computing specific attributes designed to measure the continuity of seismic reflectors. First-generation coherence techniques, for instance, infer potential fault locations by identifying data discontinuities [1]. Second-generation coherence algorithms, such as D2, calculate seismic trace similarity using covariance matrices [2], while third-generation algorithms derive coherence attributes from eigenvalue analysis for seismic fault identification [3]. However, coherence-based algorithms are susceptible to noise and residual stratigraphic responses. Methods predicated on fault-induced discontinuities, such as variance [4,5] and gradient magnitude [6], also exhibit limitations stemming from their sensitivity to noise and stratigraphic features. Early fault enhancement methods employed computation windows or smoothing directions premised on simplified geometric assumptions [7–9]. Given the complex morphology of actual faults, subsequent methods evolved to perform smoothing along the true dip and azimuth of faults, either by directly computing attributes [10,11] or by enhancing existing ones [12–14]. Nonetheless, these approaches are often computationally demanding. To overcome this challenge, more

sophisticated fault enhancement techniques emerged, [15,16] as well as optimal surface voting [17]. Nevertheless, their performance is critically dependent on the quality of initial fault attributes and meticulous parameter tuning, rendering them computationally intensive and prone to generating inaccurate or spurious fault connections in complex geological settings. Dorn and James et al. [18] integrated geological prior knowledge with digital signal processing techniques to develop an automated fault extraction (AFE) method, which proved effective for identifying large, discontinuous faults. Admasu et al. [19] proposed the principal contour line technique, which achieved semi-automated fault tracking by combining manual and automated interpretation. Although existing fault identification methods exhibit a degree of intelligence and automation, their efficacy often remains contingent upon the initial selection of attributes and the chosen computational strategies. More fundamentally, these traditional approaches are rooted in a "model-driven" or "rule-based" paradigm. They rely on human experts to pre-define a simplified geophysical or geometric proxy for a fault—such as a surface of discontinuity, a zone of low coherence, or a specific dip and azimuth. The algorithms are then designed to search for patterns that conform to these explicit, pre-defined rules. As exploration targets deeper and more challenging environments, we increasingly encounter faults that do not adhere to these simple models: subtle micro-faults, complex intersecting networks, and geologically significant fractures with no discernible displacement. In such scenarios, rule-based methods are not just less accurate; they are conceptually inadequate, as the phenomena being sought lie outside their descriptive capacity. Their reliance on a limited set of handcrafted attributes represents a low-dimensional projection of a high-dimensional reality, inherently incapable of capturing the complex, non-linear relationships hidden within the raw seismic data. This limitation necessitates an epistemological transition from a model-driven to a "data-driven" approach, for which deep learning serves as the core enabling technology. Instead of prescribing what a fault should look like, deep learning models, particularly Convolutional Neural Networks, learn to recognize faults by discovering intricate patterns directly from the data itself. This represents a fundamental paradigm shift, empowering us to tackle a level of geological complexity that was previously intractable.

The remarkable success of deep learning in interdisciplinary fields such as computer vision has propelled automated fault identification into the mainstream. In 2018, Di et al. [20] conducted a comparative analysis by applying both Multilayer Perceptrons (MLPs) and Convolutional Neural Networks (CNNs) to the same fault dataset, further demonstrating the superiority of CNNs for fault identification. In the same year, Wu et al. [21] also utilized CNNs for fault identification. Subsequently, in 2019, Wu et al. [22] introduced the 3D U-Net architecture to the domain of fault interpretation and proposed the FaultSeg3D network. By leveraging synthetically generated fault data samples, this networld achieved precise fault identification in field data. Yang et al. [23] applied U-Net++ to seismic fault identification, enhancing the clarity of fault edge delineation. Recognizing the limitations of standard skip connections in U-Net, researchers have explored various architectural enhancements. For instance, Yan et al. [24] proposed W-Net, which introduces a second expansive path to the U-Net architecture to enrich context information and improve information flow across scales. Cui et al. [25] introduced MS-Unet, which focuses on fusing multi-scale feature maps from different encoder levels at the decoder side to better capture both local details and global context. Hu et al. [26] employed a slice-based processing approach for 3D seismic data, applying the VGG16 network for fault identification. They further integrated Atrous Spatial Pyramid Pooling (ASPP) to effectively improve both the efficiency and accuracy of automated fault recognition.

AI-driven methods have brought a leap in efficiency for seismic fault identification, enabling faster and more automated processing of massive datasets. Nevertheless, de-

spite this significant improvement in efficiency, limitations persist, primarily manifesting in (1) insufficient precision in fine-grained fault edge identification, often hindering the accurate delineation of complex fault geometries, and (2) low robustness of networks when dealing with 3D seismic data characterized by noise, low signal-to-noise ratio (SNR), or weak fault responses. To address the issue of insufficient precision in fault edge identification mentioned in (1), MedNeXt [27] was introduced and adaptively simplified and adjusted to construct a backbone network suitable for 3D seismic data, owing to its superior performance in 3D edge detection. However, given the vast scale and diverse fault morphologies of 3D seismic data, the feature fusion mechanisms within MedNeXt may not effectively distinguish and utilize information from different hierarchical levels, potentially even introducing noise. Concurrently, to mitigate the noise issues mentioned in (2) and enhance feature discriminability, we observed that the introduction of the Semantic and Detail Infusion (SDI) module in U-Net v2 [28], by improving traditional feature fusion strategies, has to some extent overcome the problem of simple feature concatenation easily introducing noise. Therefore, we integrated the SDI module into the adjusted MedNeXt architecture to leverage the strengths of both. However, we further discovered that, particularly under geological conditions with low SNR, weak fault responses, or complex structural styles, the original Hadamard product-based feature fusion method within the SDI module inherently suffers from limitations, readily causing the drowning or loss of critical fault information, thereby directly restricting its performance on complex 3D seismic data. To overcome this critical bottleneck, inspired by adaptive feature recalibration mechanisms proposed by Woo et al. [29], we designed an Attention-Weighted Semantics and Detail Infusion (AWSDI) module. The core innovation of this module lies in its utilization of an attention mechanism, enabling it to dynamically and adaptively weight and fuse feature information from different encoder depths, based on the prominence of multi-scale fault responses within the seismic data. This refined feature fusion strategy significantly enhances the network's ability to capture multi-scale features, from regional major faults to local minor faults, and more effectively addresses the information loss issue of the original SDI module under complex geological conditions, consequently improving the network's accuracy and robustness against complex faults.

In summary, by optimizing the backbone network and innovatively introducing the AWSDI module, our proposed SwiftSeis-AWNet directly addresses the limitations of existing advanced AI methods in achieving precise edge identification and robust performance in complex noisy environments for 3D seismic fault recognition. This work aims to achieve high-precision and high-robustness automatic identification of complex fault systems in 3D seismic data, providing more accurate fault data support for subsequent tasks such as structural analysis, prospect evaluation, and well placement.

2. Fault Fundamentals and Traditional Methods

2.1. Seismic Fault Overview

Seismic faults, representing fracture surfaces or zones within the Earth's crust along which significant relative displacement of rock masses has occurred, are central objects of investigation in structural geology and seismology. Based on the sense of relative motion between the rock blocks on either side of the fault plane, seismic faults are classified into various types. Dip-slip faults constitute a major category, wherein displacement occurs predominantly parallel to the dip of the fault plane. Within the category of dip-slip faults, a normal fault ((Figure 1 left) is characterized by the downward movement of the hanging wall relative to the footwall. Such motion typically signifies that the Earth's crust is undergoing extension and is commonly observed in tectonic settings such as continental rifts, passive continental margins, and oceanic spreading centers. Conversely, a reverse

fault (Figure 1 right) exhibits the upward movement of the hanging wall relative to the footwall, indicative of the crust being subjected to intense compressional stress. These faults commonly develop along convergent plate margins, such as in subduction zones and collisional orogens. In reality, seismic faults seldom manifest as idealized, singular planar surfaces. Instead, they typically present as fault zones (Fault Zone) possessing finite thickness and complex internal architecture. Such zones may encompass principal slip surfaces, subsidiary fractures, fault breccia, fault gouge, and mylonites, among other fault rocks. This complexity reflects the cumulative effects of fault activity and the diverse deformation mechanisms involved.

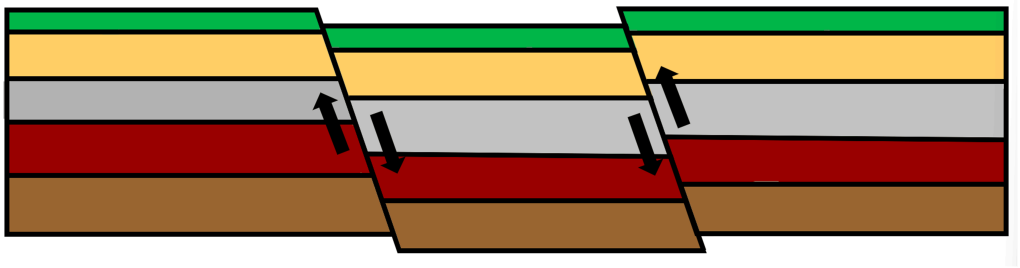


Figure 1. Geometrical representation of dip-slip faults. (**left**) Normal fault (**right**) Reverse fault.

2.2. *Ant Tracking Technique*

The core concept underpinning the Ant Tracking algorithm [15,16] originates from simulating the foraging behavior observed in natural ant colonies. During their search for food sources, ants release a chemical substance known as a pheromone, leaving a trail along their traversed paths. Subsequent ants, when selecting a path, exhibit a preference for routes with higher pheromone concentrations, simultaneously reinforcing these paths by depositing their own pheromones. This establishes a positive feedback mechanism. Over a period of iterations, pheromones accumulate significantly along the shortest or optimal paths connecting the nest to the food source, thereby guiding the colony to efficiently locate the target.

Applying this principle to seismic fault identification, the fundamental workflow is as follows: (1) The 3D seismic data volume is discretized into a grid of nodes or voxels, which constitute the search space for artificial ants. (2) Local seismic attributes capable of characterizing fault features are defined. These may include, but are not limited to, coherence (or similarity), dip, azimuth, curvature, or the response of edge detection operators (e.g., Canny and Sobel). The values of these attributes guide the movement decisions of the artificial ants, meaning ants tend to move in directions where seismic attributes indicate a higher likelihood of a fault. Multiple artificial ants deployed within the data volume conduct path searches either independently or cooperatively. When an ant traverses a path, if the sequence of nodes along that path is geometrically continuous and aligns with a priori geological knowledge of fault development, the ant "releases" pheromones along that path. Concurrently, to prevent premature convergence to local optima and to encourage the exploration of new potential paths, the pheromones gradually "evaporate" over time (i.e., with increasing iteration count).

Despite the numerous advantages demonstrated by the Ant Tracking algorithm in seismic fault identification, its application to large-scale 3D seismic data volumes imposes considerable demands on computational resources and time, owing to the extensive iterations and the large number of artificial ants required. Furthermore, the Ant Tracking algorithm is susceptible to becoming trapped in local optima. This risk is particularly pronounced in areas with low signal-to-noise ratios in the seismic data or in regions with exceedingly complex geological structures, where the algorithm might prematurely con-

verge on locally strong anomalies or noise, thereby failing to comprehensively identify the true fault system.

3. Methods

3.1. Overall Architecture

The SwiftSeis-AWNet network proposed in this paper adopts an encoder–bottleneck–decoder architecture adapted from MedNeXt [27], incorporating our refined AWSDI module to reconfigure the interaction of cross-scale features. The network receives a 3D seismic volume as input, typically with spatial dimensions of (L, L, L) and a single channel.

Table 1 intuitively elucidates the primary distinctions between SwiftSeis-AWNet and the original MedNeXt concerning network depth (number of downsampling layers) and feature fusion mechanisms, alongside the rationale underpinning these modifications.

Table 1. Comparison of original MedNeXt and SwiftSeis-AWNet architectures.

Feature Component	Original MedNeXt	SwiftSeis-AWNet	Rationale for Modification
Domain	Medical Image Segmentation	3D Seismic Fault Identification	Optimization for seismic data characteristics.
Encoder/Decoder Depth	4 upsampling and downsampling operations	3 upsampling and downsampling operations	Balancing performance and efficiency, referencing established architectures in fault identification like FaultSeg3D, and validated by experiments.
Skip Connection Feature Fusion	Simple element-wise addition	AWSDI Module	Overcoming noise and information loss from simple addition; enabling adaptive, dynamic fusion of multi-scale features via attention to improve complex fault identification.

The SwiftSeis-AWNet encoder path is responsible for hierarchical feature extraction from the input data, progressively reducing spatial resolution to capture multi-scale contextual information. This path comprises three consecutive stages. In each encoder stage, features are first extracted via a Feature Fusion Block (as indicated by the green box in Figure 2). The generated feature maps then undergo downsampling through a Downsampling Block (as indicated by the purple box in Figure 2). After a total of three such "feature extraction-downsampling" operations, the spatial resolution of the feature maps at the end of the encoder is reduced to $(L/8)^3$, and the number of channels is expanded to 8 times the initial channel count C. The bottleneck layer serves as the transition point between the encoder and decoder, consisting of two Feature Fusion Blocks, and processes the feature maps from the deepest layer of the encoder.

The decoder path is designed to progressively restore the spatial resolution of the feature maps, integrating fine spatial details provided by the encoder with deep semantic information propagated through the decoder. The decoder also comprises three layers, structured symmetrically to its encoder counterpart. Feature upsampling is performed by an Upsampling Block (as indicated by the yellow box in Figure 2). Following this upsampling operation, the pivotal AWSDI module is then incorporated. This module receives two sets of inputs: feature maps from the corresponding encoder level, and the upsampled feature maps from the current decoder layer. Leveraging its internal attention mechanism, the AWSDI module computes importance weights for disparate features and subsequently performs an adaptive weighted summation to yield a fused feature map. This dynamic fusion strategy aims to intelligently combine multi-scale information based

on the local relevance of features. The fused feature map is then passed to the Feature Fusion Block of that decoder stage for further refinement, thereby completing the primary processing for that decoder stage.

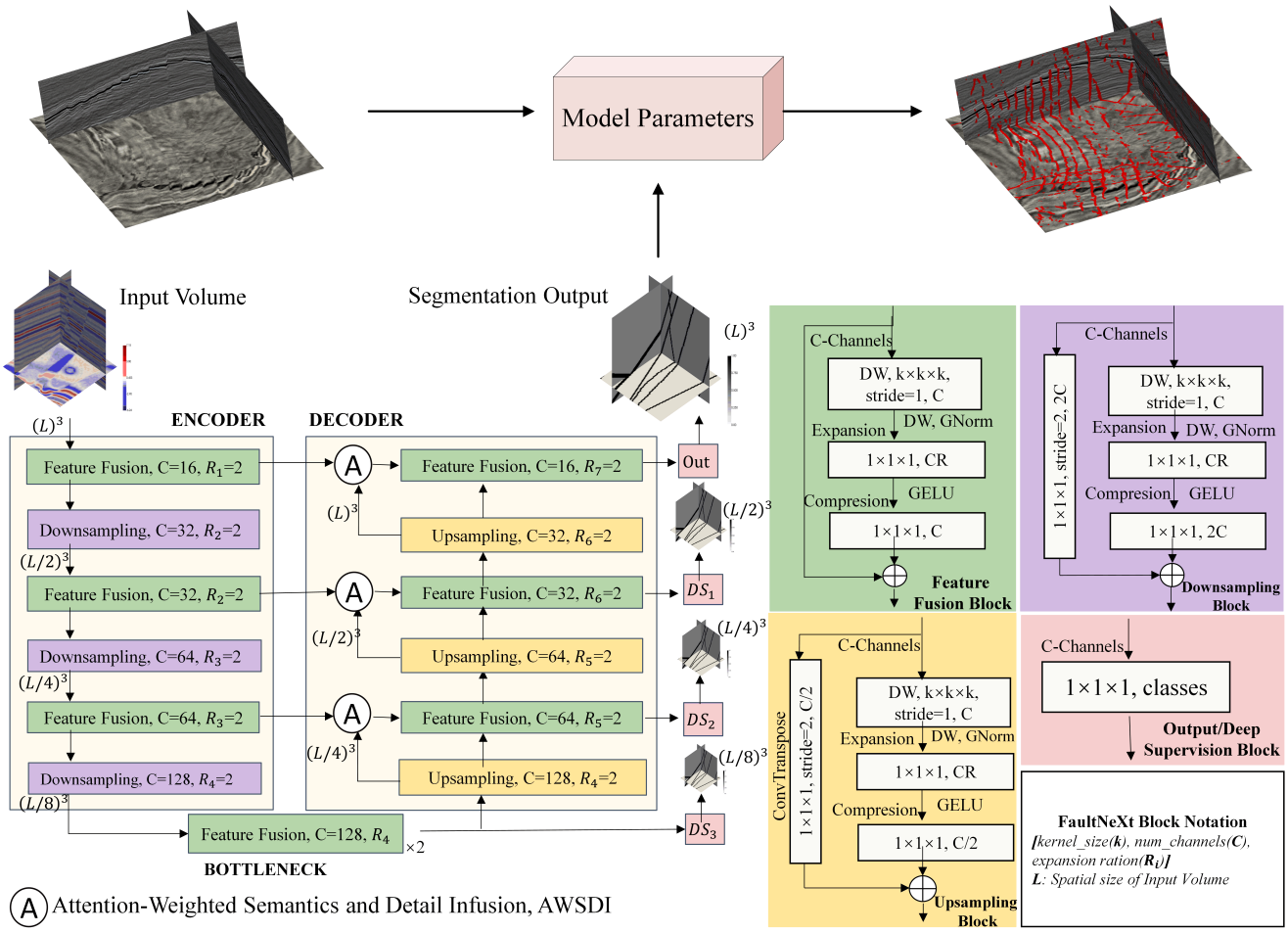


Figure 2. Overall architecture of the proposed SwiftSeis AWNet network for 3D seismic fault identification.

The final feature map output from the last decoder stage is fed into the ultimate Output Module (Out Block, as indicated by the pink box in Figure 2). This module is typically a simple $1 \times 1 \times 1$ convolutional layer, which maps the feature channels to the final number of output classes, generating a fault probability map. To optimize the training process and enhance gradient propagation, the network employs a Deep Supervision (DS) strategy, where the total loss is computed as a weighted sum of the main output loss and several auxiliary losses.

3.2. Depthwise Separable Convolution

Depthwise separable convolution [30] consists of two main components: depthwise convolution (DW) and pointwise convolution (PW) (Figure 3). This architecture is employed to achieve efficient feature extraction for a given receptive field. The process also includes an additional pointwise convolution for channel adjustment, elaborated as follows:

Depthwise Convolution This stage focuses on extracting spatial features. For the input feature volume $X \in \mathbb{R}^{C_{\text{in}} \times D \times H \times W}$ (where C_{in} denotes the number of input channels, and D, H, W represent the depth, height, and width of X), C_{in} independent $k \times k \times k$ (where $k = 3$ in this research) single-channel convolutional kernels $W_{dw}^{(c)}$ are

applied, performing convolution on each input channel separately. By symmetric zero padding, the spatial dimensions are kept unchanged, yielding the depthwise convolution output feature volume $F_{dw} \in \mathbb{R}^{C_{\text{in}} \times D \times H \times W}$. Its calculation process (for the c-th channel) can be expressed as

$$F_{dw}^{c} = W_{dw}^{c} * X_{c}, \tag{1}$$

where $*$ denotes the 3D convolution operation, and X_c is the c-th channel of the feature volume X.

Group Normalization To alleviate potential channel-wise distribution shifts caused by depthwise convolution and enhance training stability, the depthwise convolution output F_{dw} is subjected to Group Normalization. With the number of groups set equal to the number of input channels C_{in}, independent standardization of each channel's features is performed, yielding the normalized feature volume $F_{norm} \in \mathbb{R}^{C_{\text{in}} \times D \times H \times W}$:

$$F_{norm}^{c} = \gamma_c \cdot \frac{F_{dw}^{c} - \mu_c}{\sqrt{\sigma_c^2 + \epsilon}} + \beta_c, \tag{2}$$

where μ_c and σ_c are the mean and standard deviation of the c-th channel features, γ_c and β_c are learnable affine transformation parameters, and ϵ is a small constant to prevent division by zero.

Pointwise Convolution—Channel Expansion Drawing inspiration from Transformer architectures, a channel expansion ratio R (set to 2 in this research) is introduced. This stage uses a $1 \times 1 \times 1$ pointwise convolution $W_{pw1} \in \mathbb{R}^{(R \cdot C_{\text{in}}) \times C_{\text{in}} \times \mathbb{1} \times \mathbb{1} \times \mathbb{1}}$ to expand the channel dimension of the normalized feature volume F_{norm} by a factor of R. The expanded features then pass through the GELU activation function to introduce non-linearity, enhancing the network's capacity to network complex patterns. This results in the expanded feature volume $F_{expand} \in \mathbb{R}^{(R \cdot C_{\text{in}}) \times D \times H \times W}$:

$$F_{expand} = GELU(W_{pw1} * F_{norm}), \tag{3}$$

Pointwise Convolution—Channel Compression Finally, through another $1 \times 1 \times 1$ pointwise convolution $W_{pw2} \in \mathbb{R}^{C_{\text{out}} \times (R \cdot C_{\text{in}}) \times \mathbb{1} \times \mathbb{1} \times \mathbb{1}}$, the channel dimension of the expanded feature volume F_{expand} is compressed to the target output dimension C_{out}. This completes the cross-channel information fusion, yielding the final output feature volume $F_{out} \in \mathbb{R}^{C_{\text{out}} \times D \times H \times W}$:

$$F_{out} = W_{pw2} * F_{expand}, \tag{4}$$

The depthwise separable convolution strategy significantly reduces the network's parameter count and computational cost. Compared to a standard 3D convolution using a $k \times k \times k$ kernel and C_{out} output channels (with parameter cost of $O(k^3 \cdot C_{\text{in}} \cdot C_{\text{out}})$), the parameter cost of depthwise separable convolution is the sum of the depthwise convolution $O(k^3 C_{\text{in}})$ and the pointwise convolutions $O(C_{\text{in}}(RC_{\text{in}}) + (RC_{\text{in}})C_{\text{out}})$. The efficiency gain is particularly significant when k is large, highlighting its advantage in constructing efficient deep networks.

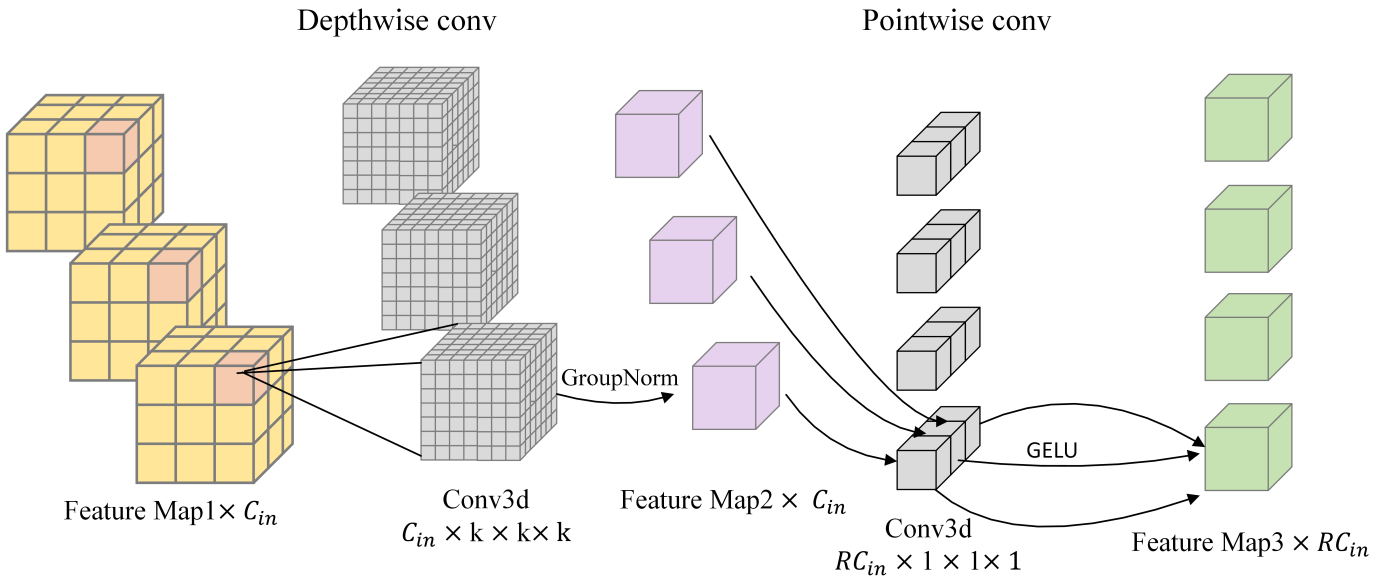


Figure 3. Architecture of depthwise separable convolution.

3.3. Inverted Residual Structure

Traditional Residual Networks [31] (as shown in Figure 4a) typically employ a 'bottleneck' channel configuration. This involves performing channel dimensionality reduction on input features via $1 \times 1 \times 1$ pointwise convolutions, executing subsequent convolutional operations in a low-dimensional space to extract spatial features, and finally expanding the dimensionality again using $1 \times 1 \times 1$ pointwise convolutions. While this design effectively controls the computational overhead, the significant channel compression may lead to the loss of high-frequency details or critical information.

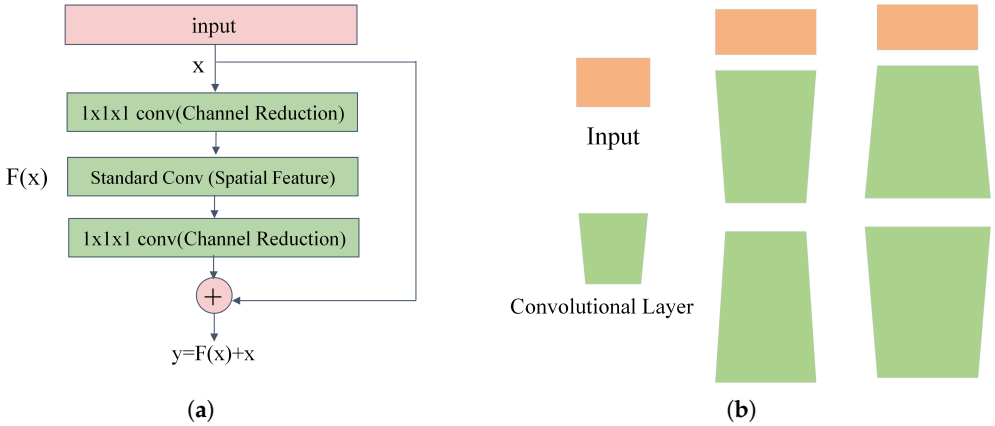


Figure 4. Residual block designs and channel flow. (**a**) Structure of a traditional residual block. (**b**) Channel width comparison for traditional vs. inverted residual modules.

To address this, this research draws inspiration from the design philosophy of the inverted residual module [32]. The comparison of channel configurations between the inverted residual module and the traditional residual structure is illustrated in Figure 4a. The inverted residual module (illustrated on the right side of Figure 4b) is fundamentally characterized by its unique channel transformation strategy: it employs higher-dimensional feature representations in the intermediate stage compared to the input and output layers, thereby forming an inverted bottleneck structure. This unique 'depthwise convolution–expansion–compression' sequential design significantly reduces the risk of information loss by performing primary feature processing in a high-dimensional space.

All residual modules incorporate skip connections to facilitate gradient propagation and enhance training stability. Specifically, in the Feature Fusion Block, input features are directly added element-wise to the output of the main path, exemplifying a classic residual connection pattern. For the Downsampling Block, its skip connection is implemented by spatially downsampling the input features via a convolutional block and doubling their channel count relative to the input, with the adjusted result then added element-wise to the output of the main path. Similarly, in the Upsampling Block, the skip connection upsamples the input features via a transposed convolution operation and compresses their channel count to C/2, before adding this result element-wise to the output of the main path.

This residual connection mechanism effectively alleviates the vanishing gradient problem during backpropagation. This is evident in its gradient term calculation formula:

$$\frac{\partial \mathbf{L}}{\partial \mathbf{X}} = \frac{\partial \mathbf{L}}{\partial \mathbf{Y}} \cdot \left(\mathbf{a} + \frac{\partial \mathbf{F}(\mathbf{X})}{\partial \mathbf{X}} \right), \tag{5}$$

As a is a constant, it is typically set to 1 to ensure the stability of the gradient magnitude. Even when the gradient of the main path, $\frac{\partial \mathbf{F}(\mathbf{X})}{\partial \mathbf{X}}$, approaches zero, it guarantees effective backpropagation of the gradients, thus greatly facilitating the training of deep networks.

3.4. Attention-Weighted Semantics and Detail Infusion (AWSDI) Module

In the tasks of fine interpretation and automatic identification of fault systems within 3D seismic data volumes, fault responses typically exhibit complex multi-scale and multi-morphology characteristics, ranging from large-scale regional fault zones to intricate local fracture networks. Their manifestations on seismic sections—such as waveform displacement, seismic horizon distortion, diminished continuity, and distinct dip and azimuth responses—as well as their signal-to-noise ratios, exhibit significant variability. The original SDI module [28] utilizes a Hadamard product for multi-scale feature fusion. While capable of integrating information, its fixed multiplicative fusion method struggles with the pervasive strong noise interference and the inherent diversity and discontinuity of fault responses across varying geological settings. Consequently, simple fusion may fail to effectively differentiate genuine fault discontinuities from stratum- or noise-induced artifacts that mimic fault-like features. To overcome these limitations, we propose the AWSDI module.

As depicted in Figure 5, this module processes multi-level 3D feature volumes $\{F_i \in \mathbb{R}^{B_i \times C_i \times D_i \times H_i \times W_i}\}_{i=1}^{N}$. Here, C_i denotes the channel count for the i-th feature volume, while D_i, H_i, W_i represent its spatial dimensions. The module first employs a dynamic spatial interpolation strategy to align these multi-scale features: if the spatial dimensions $D_i \times H_i \times W_i$ of an input feature volume exceed the target size $D_{\text{target}} \times H_{\text{target}} \times W_{\text{target}}$ (defined by the first feature volume, F_1), `AdaptiveAvgPool3d` is applied; otherwise, trilinear interpolation is used for upsampling.

Subsequently, independent convolutional paths are applied, each utilizing a 1-stride, 1-padding kernel, to project each feature volume to a consistent channel dimension, matching that of F_1. With both spatial and channel alignments complete, the module concatenates the processed feature volumes along the channel dimension into $F_{\text{concat}} \in \mathbb{R}^{B \times (N \times C_1) \times D_{\text{target}} \times H_{\text{target}} \times W_{\text{target}}}$. This F_{concat} is then fed into a lightweight attention network, Equation (7). The processing flow of this network is presented as Attention_conv in Figure 5. The primary objective of this network is to learn spatially correlated attention weights for each input scale. More precisely, the attention network generates a set of attention weight maps $\{A_i\}_{i=1}^{N}$ by applying the following function:

$$A_i = \sigma(\text{Conv}1 \times 1 \times 1_{\text{out}}(\text{ReLU}(\text{GroupNorm}(\text{Conv}1 \times 1 \times 1_{\text{mid}}(F_{\text{concat}})))))_i, \tag{6}$$

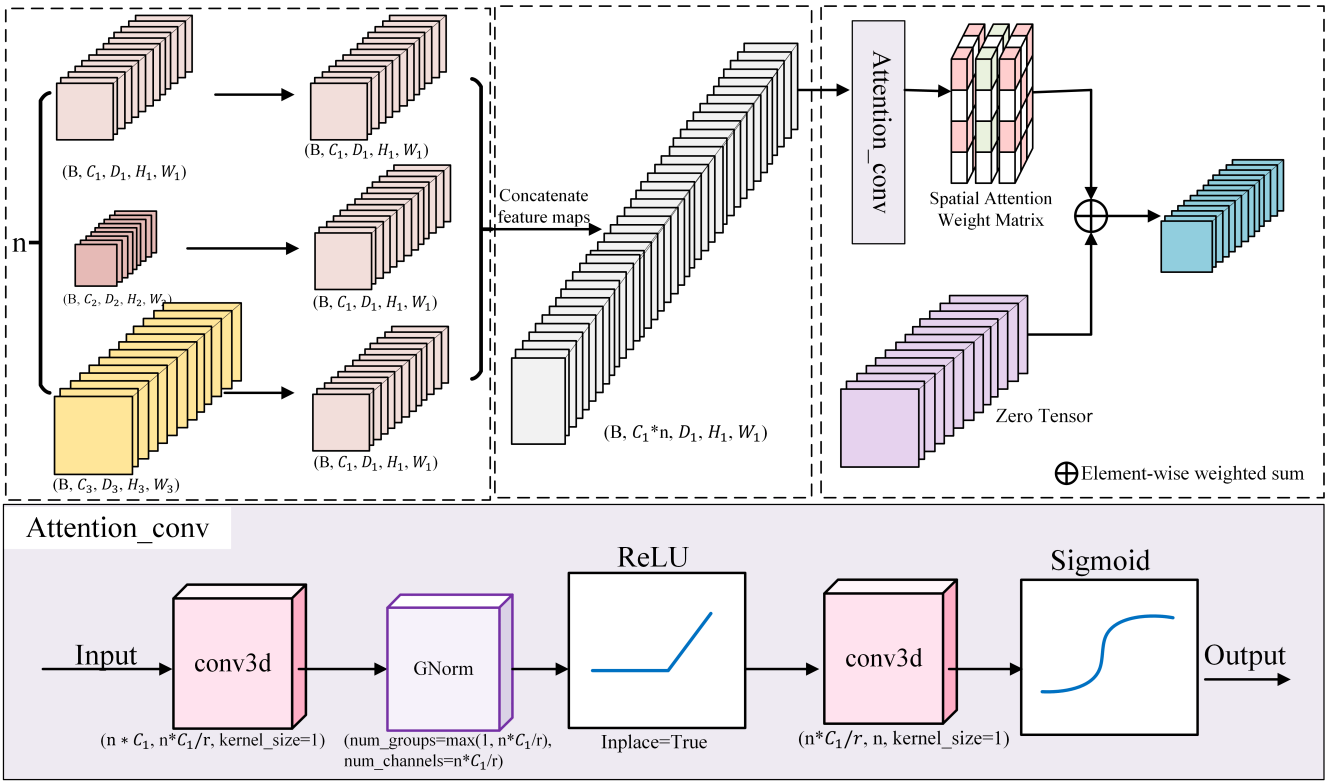


Figure 5. Overview of the AWSDI module. Multi-scale inputs (n scales) are processed to uniform channel (C_i) and spatial dimensions, then concatenated.

In this formulation, σ denotes the Sigmoid function, and the notation $(\cdot)_i$ signifies the feature map extracted from the i-th channel. The resulting spatial attention weight matrix, $A_i \in \mathbb{R}^{B\times N\times D_{\text{target}}\times H_{\text{target}}\times W_{\text{target}}}$, contains values within the range of $[0, 1]$. These values indicate the relative importance of the i-th scale feature at specific spatial locations (d, h, w). Subsequently, the fused features are computed through an element-wise weighted summation as defined by Equation (7):

$$F_{\text{fused}} = \sum_{i=1}^{N} A_i \odot \Phi_i(F_i), \tag{7}$$

Here, Φ_i denotes the i-th convolutional path, and $\odot$ represents element-wise multiplication. To ensure training stability, the bias terms of the convolutional layers whose outputs directly contribute to the fusion (e.g., those producing A or $\Phi_i(F_i)$) are zero-initialized during module initialization. This promotes a more uniform initial contribution from each path during fusion.

To qualitatively evaluate the operational mechanism of the proposed AWSDI module, particularly its stability and noise sensitivity under low signal-to-noise ratio (SNR) conditions, we conducted a visualization analysis of its internal complete data processing pipeline (as illustrated in Figure 6). At the input stage, the module receives two feature maps: a high-resolution feature map originating from the encoder's skip connection (Figure 6b) and a low-resolution feature map from the decoder's upsampling path (Figure 6e). As depicted, the high-resolution input is susceptible to significant high-frequency noise and geological artifacts. In contrast, the low-resolution input features, derived from deeper network layers, are smoother, have preliminarily captured the main structures of faults, and contain stronger semantic information. During the preprocessing stage, these input features undergo alignment and convolutional operations, resulting in Processed Feature Maps (Figure 6b,f). In the attention generation stage, the module generates spatially adaptive

attention weight maps for the preprocessed features. Despite the high-resolution features (Figure 6b) being affected by noise and geological artifacts, their corresponding attention maps (Figure 6c) accurately assign high weights to the linear fault structures while assigning near-zero weights to the background noise regions. This phenomenon indicates that the attention allocation of the AWSDI module is not dominated by local high-frequency noise but successfully leverages the global contextual information provided by the low-resolution features (Figure 6f) to guide a stable focus on effective signals. Concurrently, the attention map for the low-resolution features (Figure 6g) also maintains a high degree of focus on fault regions, ensuring the preservation of critical semantic information. The combined weight map (Figure 6d) visually reveals this synergistic mechanism: the bright yellow and green areas distributed along the fault lines suggest that the fusion process is primarily guided by the low-resolution features (green) and is complemented by the high-resolution features (red) for detail refinement. The final output feature map (Figure 6h), obtained after attention weighting, demonstrates a significant quality improvement. Compared to the original inputs, the fused features not only effectively suppress noise and geological artifacts but also exhibit a substantial enhancement of the fault structures, manifesting as clearer boundaries. This series of visualization results provides strong empirical evidence for the intrinsic stability and noise resilience of the AWSDI module in complex geological environments, effectively alleviating concerns about potential artifact introduction by the attention mechanism and confirming its efficacy in dynamically fusing features from different network depths.

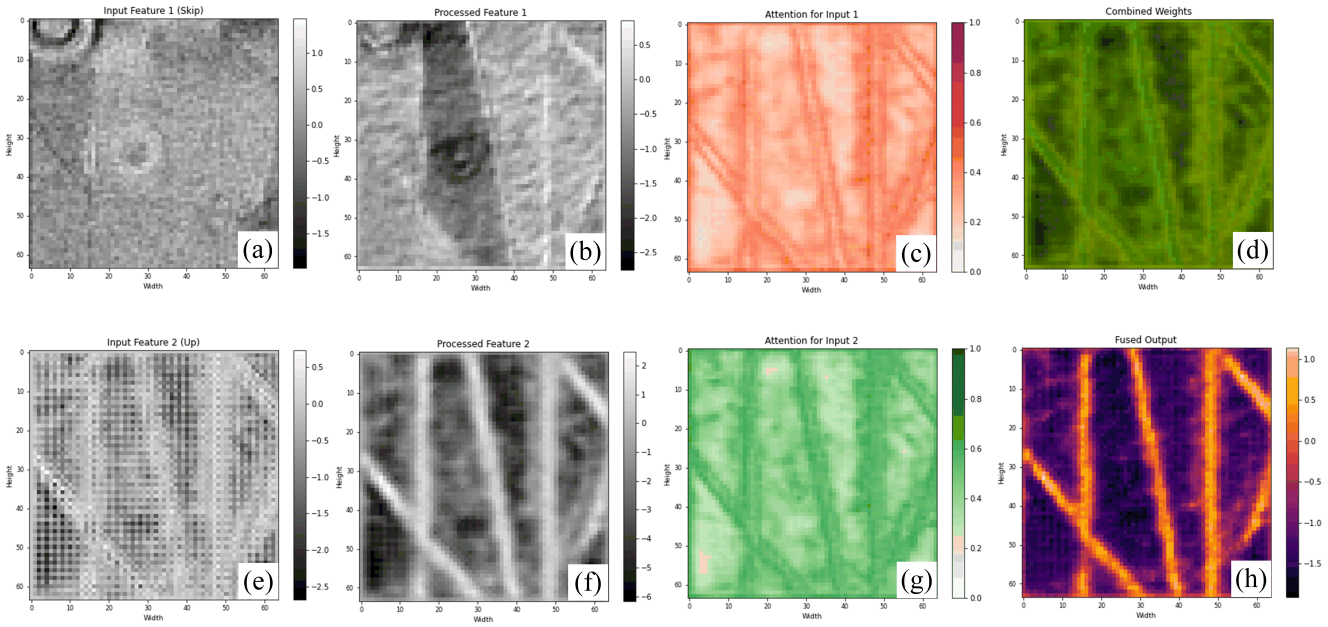


Figure 6. Visualization of the AWSDI module' attention mechanism. (**a**,**e**) are input feature maps from high-resolution skip connection and low-resolution upsampling path, respectively. (**b**,**f**) are their corresponding preprocessed features. (**c**,**g**) are spatial adaptive attention maps, respectively. The composite map (**d**) shows the fusion strategy, leading to a clean, enhanced output (**h**).

3.5. Loss Function

For the voxel-level binary classification task of 3D seismic fault detection, this research selects the Binary Cross-Entropy with Logits Loss (BCEWithLogitsLoss) as the loss function for network training.

For any voxel i in the data volume, its loss L_i is fundamentally defined as

$$L_i = -y_i \log(\sigma(x_i)) - (1 - y_i) \log(1 - \sigma(x_i)), \tag{8}$$

where x_i is the raw logit value output by the network for voxel i, y_i is the true binary label for that voxel, and σ represents the Sigmoid activation function. The total loss for the entire data volume is calculated as the mean over all voxel losses.

3.6. Evaluation Indices

For the quantitative assessment of the proposed method, this research introduced two types of evaluation indices, specifically comprising voxel-level segmentation accuracy and fault morphology and structural assessment.

3.6.1. Voxel-Level Segmentation Accuracy Indices

These indices quantify the pixel-wise prediction performance. They include Accuracy, Precision, Recall, Intersection over Union (IoU), and the Dice coefficient. Detailed calculation formulas for these indices are provided in Appendix A.1.

3.6.2. Fault Morphology and Structure Evaluation Indices

- Fault Connectivity Score (FCS): This metric is introduced to directly assess the continuity and integrity of the predicted fault system, which is a key indicator of a model's robustness against noise. The FCS quantifies the fragmentation by comparing the number of connected components in the predicted fault skeleton to that of the ground truth. A score closer to 1 indicates that the predicted fault system has a connectivity and completeness level highly consistent with the ground truth, suggesting superior noise resilience.
- Average Symmetric Surface Distance (ASSD): To precisely evaluate the edge accuracy, we employ surface distance metrics. The ASSD measures the average geometric deviation between the predicted fault surface (S_{pred}) and the ground truth surface (S_{gt}). It calculates the average of two distances: the average distance from every point on S_{pred} to its closest point on S_{gt}, and vice versa. A lower ASSD value (in voxels) signifies a better overall alignment and a higher degree of geometric fidelity between the two surfaces, providing a stable and comprehensive measure of boundary accuracy.
- Hausdorff Distance 95% (HD95): While ASSD reflects the average fit, the HD95 metric captures the maximum localized boundary deviation, making it highly sensitive to edge inaccuracies. It identifies the greatest distance between the predicted and ground truth surfaces but crucially uses the 95th percentile of distances to ensure robustness to a small number of outliers. A lower HD95 value (in voxels) indicates that even the worst-case boundary errors are small, confirming high precision in delineating fine fault edges and complex geometries.

By jointly analyzing these structural indices alongside traditional voxel-level metrics, we can conduct a more comprehensive and geologically meaningful evaluation of a model's performance, moving beyond simple pixel-wise accuracy to assess the true utility of the generated fault interpretations. (Detailed calculation formulas for these indices are provided in Appendix A.2.)

4. Experiments

4.1. Datasets

To ensure the rigor and reproducibility of our study, we meticulously designed our dataset partitioning and utilization strategy to rigorously evaluate the model's performance and generalization capabilities. The datasets are categorized into three distinct roles: a training set, a validation set, and a generalization test set.

4.1.1. Synthetic Data for Training and Validation

The training and validation data are sourced from the publicly available 3D synthetic seismic dataset constructed by Wu et al. [22] (Figure 7). This dataset was chosen not only for its accessibility but, more importantly, for the rich and diverse geological features encoded within its generation process, which is crucial for training a robust fault identification model. The generation process involves the following:

(1) An initial 1D horizontal random reflectivity network is constructed.
(2) Complex fold structures are introduced through vertical shearing (defined by randomized 2D Gaussian functions and linear scaling) and planar shearing.
(3) Multiple planar faults with randomized orientations and spatially varying displacements are embedded within this folded network.
(4) The reflectivity network, incorporating both structures and faults, is then convolved with Ricker wavelets (with randomized peak frequencies) after deformation to generate synthetic seismic traces.
(5) Random noise is added, and the final training images of size $128 \times 128 \times 128$ and their corresponding binary fault labels (faults as 1, non-faults as 0) are cropped from these noisy traces.

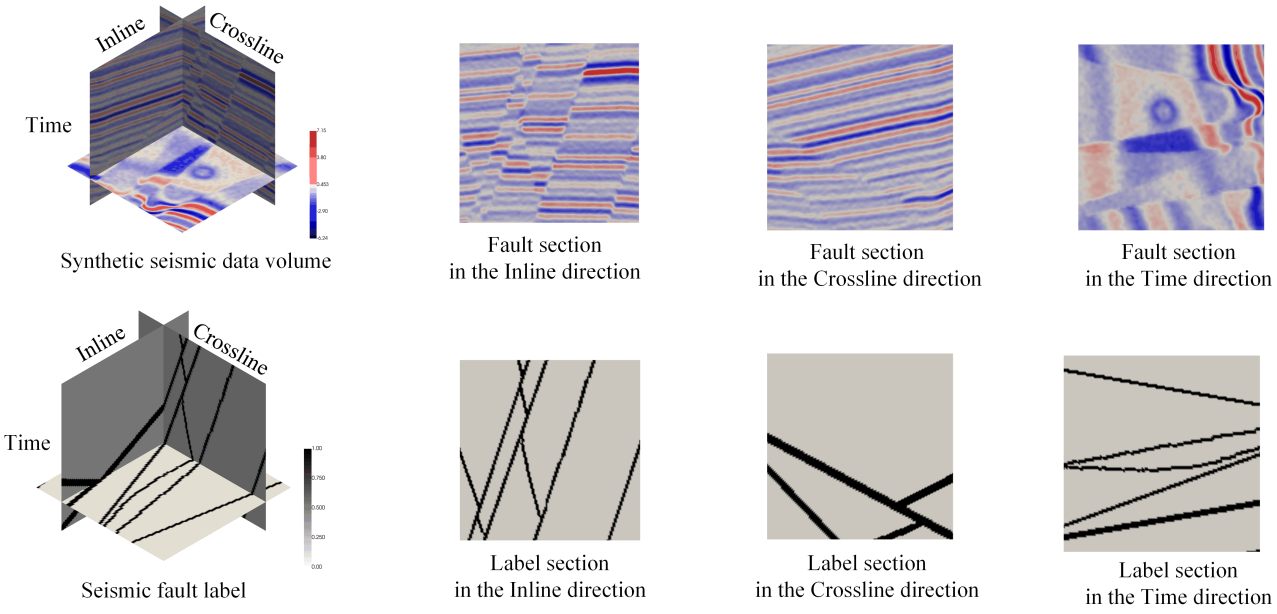


Figure 7. Synthetic seismic dataset used for model training and validation.

The systematic randomization of all key parameters ensures the diversity and uniqueness of the training dataset. The synthetic dataset comprises 220 independent 3D seismic data volumes and their corresponding fault masks. Of these, 200 volumes and their labels are used for model training, while the remaining 20 volumes and their labels are designated for evaluating the network's generalization performance.

The 3D seismic data typically consist of three orthogonal directional dimensions: Inline, Crossline (Xline), and Time. Inline and Crossline represent the spatial planar coordinates of the seismic exploration area, while Time indicates the two-way travel time of seismic waves perpendicular to this plane or the corresponding subsurface depth. In this research, the complete $128 \times 128 \times 128$ 3D data volumes are directly fed as input to the network for end-to-end fault segmentation.

4.1.2. Field Data for Generalization Testing

To strictly evaluate the model's ability to generalize from the synthetic to the real world, we employed two publicly available field datasets as our generalization test sets. To ensure consistency between the training and testing data domains, Z-score standardization was applied to the field datasets.

(1) Netherlands F3 Block Dataset: This dataset, from the Dutch North Sea, is a well-known benchmark for fault interpretation. The volume has dimensions of 128(vertical) × 384(inline) × 512(crossline).

(2) New Zealand Kerry 3D Dataset: This dataset is provided by the New Zealand Crown Minerals. The sub-volume utilized for our assessment measures 287(vertical) × 735(crossline) × 1252(inline).

4.2. Experimental Setup

All experiments were conducted on a workstation equipped with an NVIDIA Quadro RTX 4080 GPU, utilizing the PyTorch 2.1.2 deep learning framework. For the network, the AdamW optimizer was chosen, with the initial learning rate set to 1×10^{-4} and a weight decay coefficient of 0.001. The training set comprised 200 samples of size $128 \times 128 \times 128$, and the validation set included 20 samples of the same dimensions. A batch size of 1 was employed to accommodate GPU memory constraints.

To evaluate the training dynamics and performance of the SwiftSeis-AWNet network, Figure X illustrates the evolution of key indices on both the training and validation sets with respect to training epochs. As depicted, the loss function (Figure 8a) exhibits a rapid decrease during the initial training phase, ultimately stabilizing at a low value of approximately 0.05. The close concordance between the training and validation loss curves indicates that the network did not suffer from significant overfitting. Core segmentation performance indices, such as the Dice coefficient (Figure 8b) and IoU (Figure 8c), demonstrate a swift increase from their initial values, subsequently approaching saturation, with the validation set Dice coefficient ultimately stabilizing around 0.85. Furthermore, other evaluation indices, including Recall (Figure 8d), Precision (Figure 8e), and Accuracy (Figure 8f), also display similarly favorable convergence characteristics. The evolutionary trajectories of all indices on both the training and validation sets manifest a high degree of consistency, with a minimal gap maintained between them. This robustly substantiates that the proposed network possesses good learning capability, a stable convergence process, and excellent generalization performance.

4.3. Noise Experiment

Synthetic seismic datasets are capable of clearly exhibiting fault features, primarily owing to their distinct amplitude variations and prominent dislocation of seismic horizons. However, field seismic data are often subjected to various noise and environmental interferences, which can obscure fault characteristics to some extent, thereby increasing the difficulty of accurate fault identification. To assess the anti-noise capability of the proposed SwiftSeis-AWNet network under simulated field noisy conditions, Gaussian noise with a signal-to-noise ratio (SNR) of 40 was introduced into the original fault images. As shown in Figure 9, even after the addition of noise, the network's segmentation results on randomly selected slices maintained high accuracy and showed strong agreement with the ground truth labels, powerfully demonstrating the excellent robustness of the network.

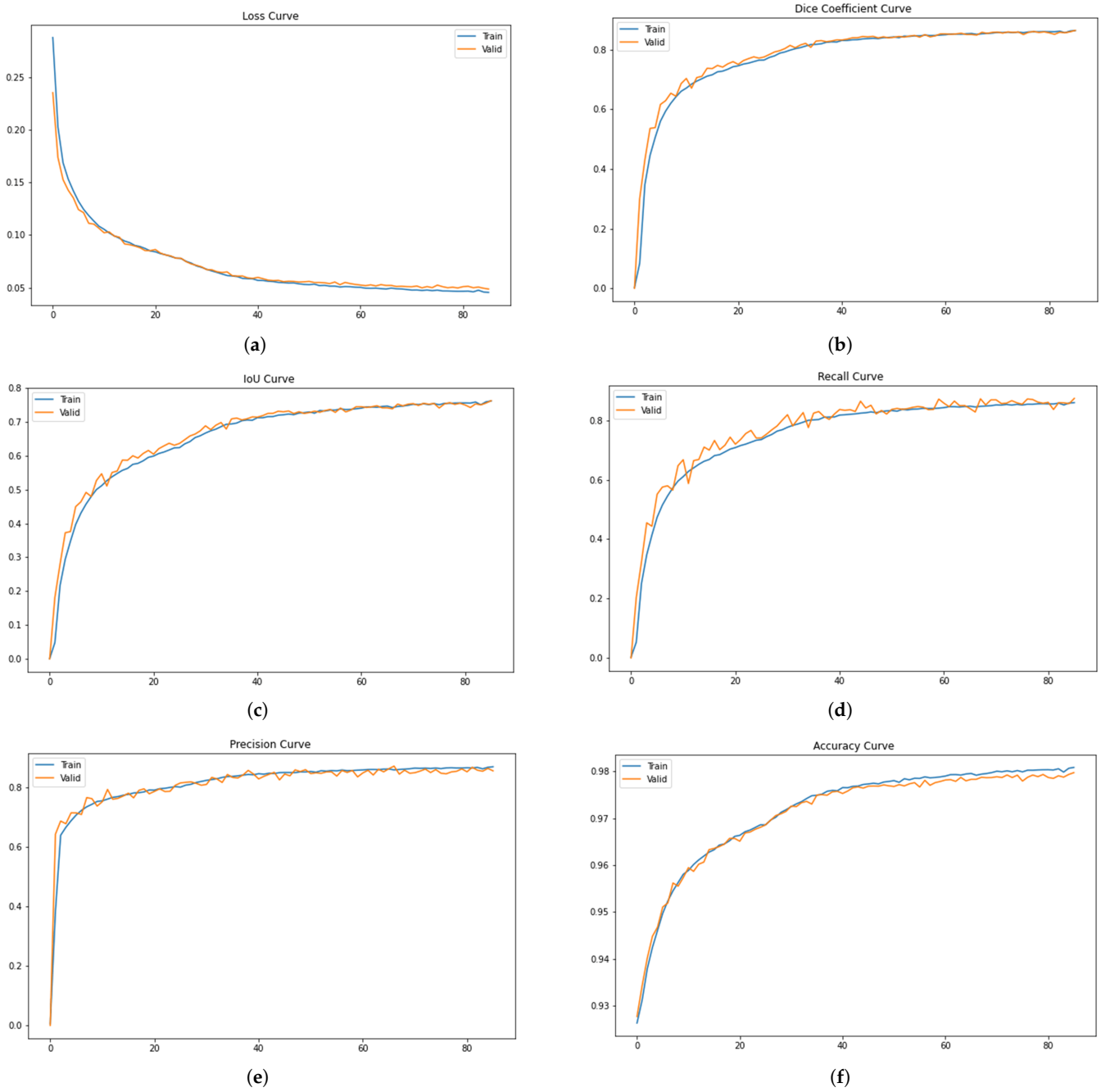


Figure 8. Performance metric curves of the SwiftSeis-AWNet network during the training process. The figure illustrates the evolution of (**a**) Loss, (**b**) Dice Coefficient, (**c**) Intersection over Union (IoU), (**d**) Recall, (**e**) Precision, and (**f**) Accuracy with respect to the training epochs.

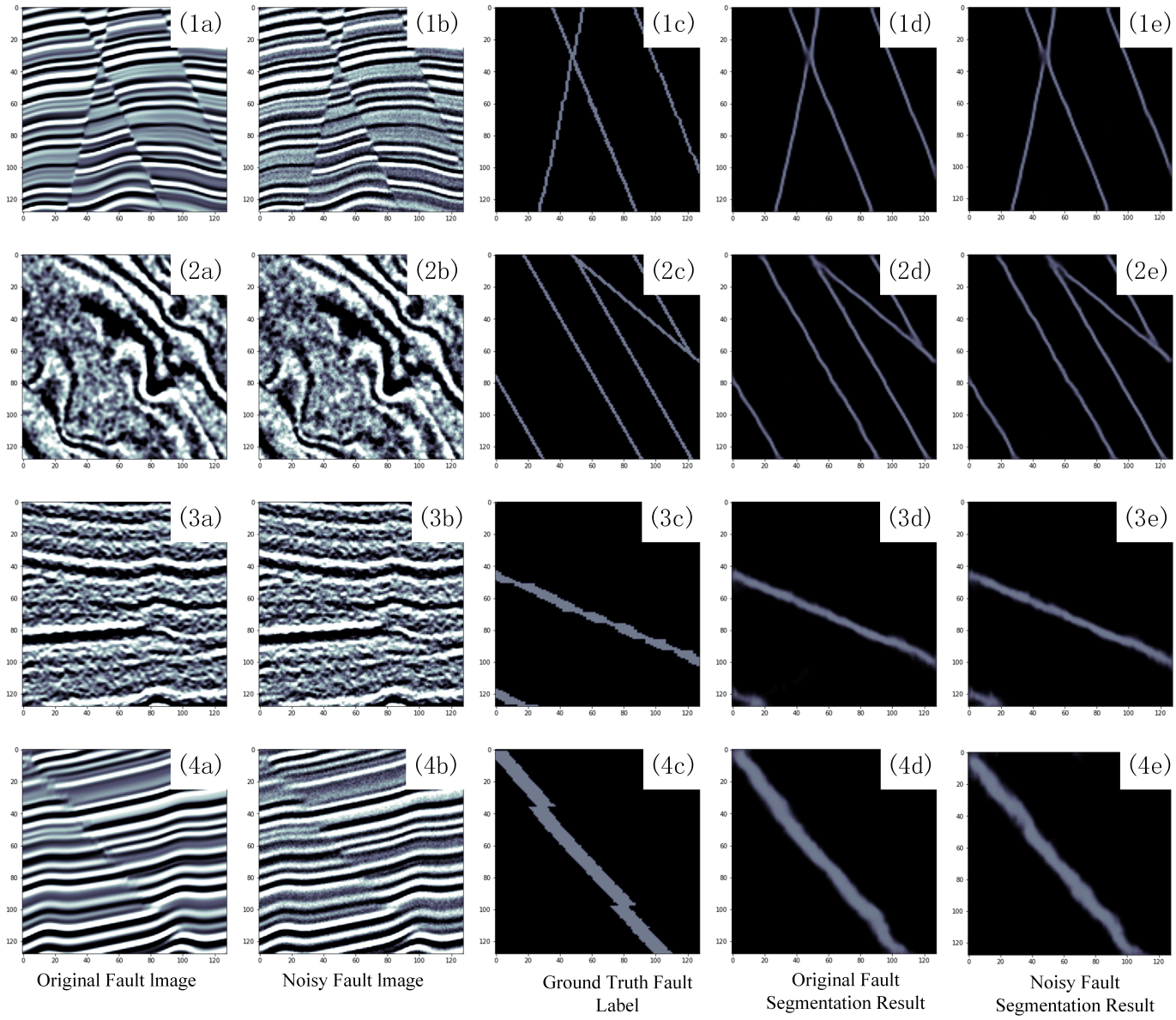


Figure 9. Comparison of fault segmentation results under noisy conditions.

4.4. Comparative Experiments

For the task of fault identification in 3D seismic data, this research employs a multi-dimensional evaluation framework to systematically compare the proposed SwiftSeis-AWNet with mainstream networks (ResUNet, FaultSeg3D, and TransUNet). This comparative analysis was conducted on both a synthetic validation dataset and field seismic datasets.

4.4.1. Synthetic Validation Dataset

This research evaluated the segmentation performance of the proposed SwiftSeis-AWNet network against other comparative networks on the validation set. As presented in Table 2, the SwiftSeis-AWNet network demonstrates comprehensive advantages in the task of 3D fault segmentation: its lowest loss function value not only reflects superior convergence characteristics and a stronger capability to fit the training data, but it also significantly outperforms other networks across all evaluated segmentation indices. Furthermore, the proposed network boasts the fewest parameters and the lowest computational complexity (time complexity), fully showcasing its integrated superiority in terms of both segmentation accuracy and efficiency.

Table 2. Quantitative comparison of different networks for 3D seismic fault identification.

Method	Accuracy	Precision	Recall	IoU	Dice	FCS	ASSD (voxels)	HD95 (voxels)	Loss	Parameters (M)	FLOPs (G)
ResUNet	0.9635	0.7850	0.7200	0.6020	0.7516	0.8405	2.8012	10.7382	0.1109	1.42	257.52
FaultSeg3D	0.9650	0.8020	0.7350	0.6210	0.7672	0.8573	2.1034	9.5445	0.0903	1.40	252.52
TransUNet	0.9730	0.8380	0.7900	0.6820	0.8133	0.9195	1.3486	7.7600	0.0875	33.32	18,438.21
SwiftSeis-AWNet	0.9794	0.8615	0.8730	0.7546	0.8607	0.9326	0.9371	5.3781	0.0492	0.621	162.57

Building upon the comprehensive multi-dimensional quantitative evaluation presented above, this research further conducted qualitative visualization analysis to more intuitively elucidate the performance disparities among different networks in terms of fault identification details. Four 3D seismic data samples were randomly selected from the validation set. Their predicted results by the proposed SwiftSeis-AWNet network (Figure 10(A2–D2)) and three comparative networks—TransUNet (Figure 10(A3–D3)), FaultSeg3D (Figure 10(A4–D4)), and ResUNet (Figure 10(A5–D5))—were visualized in 3D. The comprehensive visualization is presented in Figure 10, with the first row (Figure 10(A1–D1)) showing the corresponding ground truth fault labels.

Overall, the visual comparison in Figure 10 strongly corroborates the quantitative findings in Table 2. The SwiftSeis-AWNet network consistently demonstrates exceptional identification performance across all four samples. Its predicted fault systems not only exhibit high fidelity to the ground truth fault labels at a macroscopic structural framework level but also excel in preserving the geometric continuity of fault surfaces, delineating intricate fracture networks, and identifying subtle, small-scale discontinuities. Notably, the network exhibits superior noise robustness.

In contrast, predictions from FaultSeg3D and ResUNet commonly suffer from more pronounced fault fragmentation, blurred boundaries, and a significant presence of high-frequency noise voxels—either surrounding the fault zones or scattered throughout the background. These visually manifest as isolated high-amplitude points or irregular patchy artifacts, severely compromising the structural integrity of the interpreted fault system and the reliability of geological interpretation. While the TransUNet network demonstrates relatively robust performance and a more complete main fault structure in samples A to C, it reveals insufficient adaptability to specific complex geological conditions in Sample D (Figure 10(D3)). This manifests as numerous fine, geologically meaningless, discrete predicted points, creating artifacts that impede accurate interpretation of the primary fault structures.

Specific highlighted regions within Figure 10 further underscore SwiftSeis-AWNet's precise identification capabilities. For instance, in the red bounding boxes, our proposed network accurately captures subtle geological features such as fault tips, branching intersections, and minute geometric variations along fault surface edges. This ensures both the sharpness of fault boundaries and the fidelity of their geometric morphology, effectively mitigating common issues observed in other networks, such as eroded fault boundaries, sudden displacement variations due to excessive smoothing, or the loss of small-scale en-echelon array details. Crucially, in the ground truth label of Sample D (Figure 10(D1)), the green bounding box highlights two spatially very proximate fault surfaces. SwiftSeis-AWNet (Figure 10(D2)) successfully delineates these two independent structural units with clear separation, maintaining a distinct and identifiable inter-fault block. Furthermore, each fault surface maintains its complete morphology and excellent continuity, without exhibiting unreasonable fault fusion or artificial bridging phenomena.

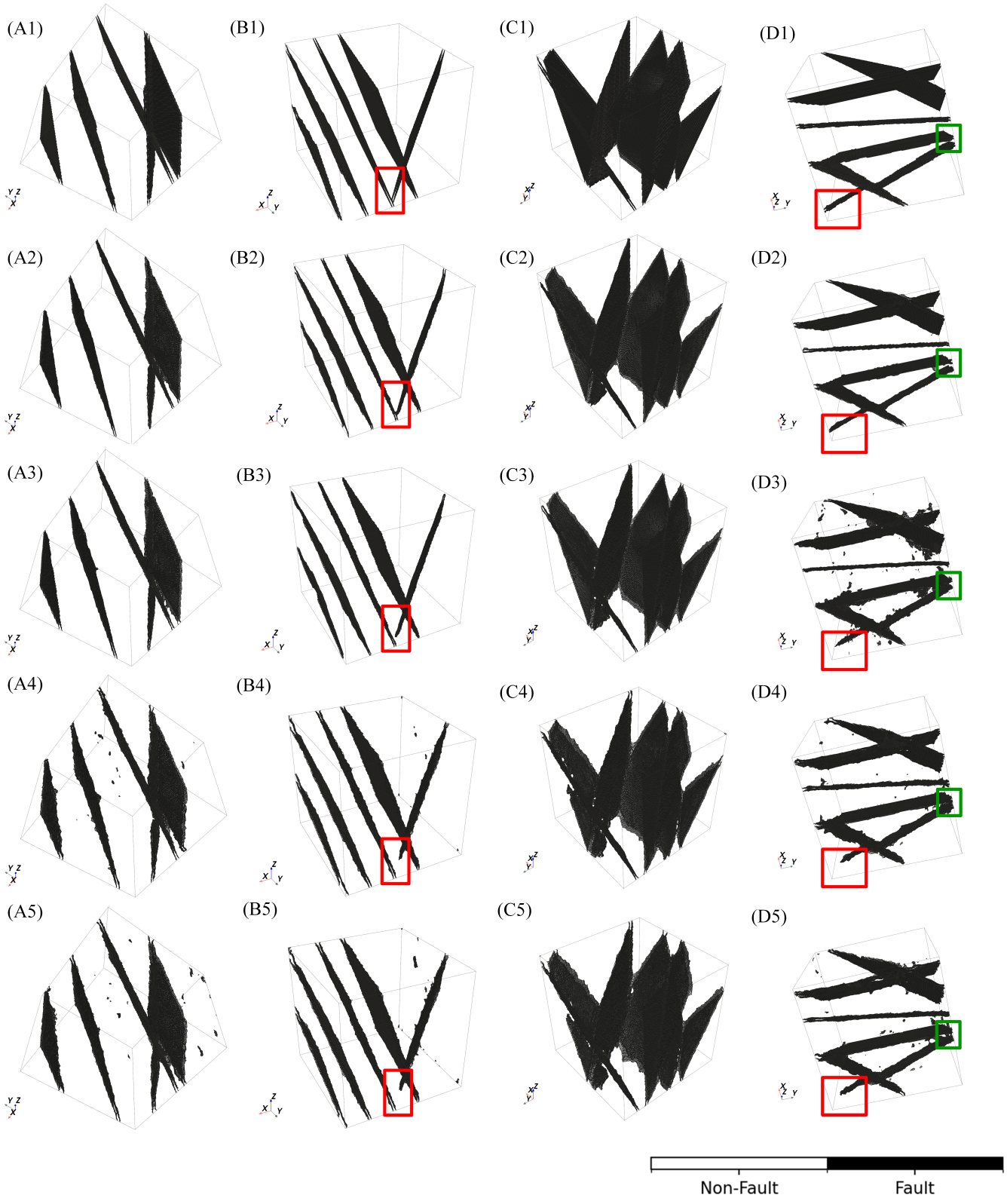


Figure 10. Visual comparison of fault identification results for different networks on four randomly selected 3D seismic data samples from the synthetic validation set. Each column (labeled A, B, C, D from left to right) represents an individual 3D seismic data sample from the validation set. The rows sequentially display the ground truth (**A1**–**D1**), identification results from our proposed SwiftSeis-AWNet network (**A2**–**D2**), the TransUNet network (**A3**–**D3**), the FaultSeg3D network (**A4**–**D4**), and the ResUnet network (**A5**–**D5**). Highlighted regions (red and green boxes) show areas of notable difference.

4.4.2. F3 Dataset

For a more intuitive evaluation of the proposed SwiftSeis-AWNet network's performance in the 3D seismic fault detection task, comparative experiments were conducted using the aforementioned baseline networks on seismic data from a specific area within the F3 block of the Dutch North Sea. The results are visually presented in Figure 11. Figure 11 illustrates (a) a slice of the original seismic volume and (b) the fault response obtained using the traditional Ant Tracking algorithm. As observed in Figure 11b, traditional methods, such as Ant Tracking, prove ineffective in capturing seismic faults. Particularly in low signal-to-noise ratio (SNR) areas, the identified fault information appears chaotic, making it challenging to effectively distinguish true fault structures. Moreover, this method is susceptible to interference from strong-amplitude, continuous formation reflections, often leading to the misidentification of formation boundaries as faults and consequently, low reliability in the identified fault interpretations.

Focusing on the area highlighted by the yellow box, which exhibits a series of nearly vertical, densely developed faults, the proposed network (Figure 11c) successfully identified these faults, presenting clear, sharp fault surface responses with excellent vertical continuity. These results accurately reflect the offset relationships of the seismic reflections (horizons) visible in the original data (Figure 11a). In contrast, FaultSeg3D (Figure 11d) exhibits significant discontinuity in fault tracing within this area, with prediction results appearing as numerous short, fragmented segments that fail to effectively connect and form complete fault surfaces. The identification result of TransUNet (Figure 11e), while capturing the main fault trends, reveals relatively blurred fault surfaces. Furthermore, some adjacent faults exhibit a 'sticking' or 'merging' phenomenon, indicating insufficient spatial resolution. The ResUNet-based network (Figure 11f) demonstrates the poorest performance in this area, with its fault identification results appearing extremely fragmented, rendering effective structural interpretations almost impossible.

The area highlighted by the green box represents a more complex geological structure, characterized by low-dip, relatively weak-response faults or structural deformation zones. In this area, the original data (Figure 11a) clearly shows the bending and discontinuity of seismic reflections (horizons). The proposed network (Figure 11c) continues to perform excellently, reliably tracing these dipping, low-SNR fracture features. The identified fault network structure is coherent and highly consistent with the geological expectations. Conversely, FaultSeg3D (Figure 11d) exhibits poor fault continuity in this area, often missing many subtle fractures. The identification results from TransUNet (Figure 11e) and ResUNet (Figure 11f) in this area are similarly unsatisfactory, presenting numerous scattered, geologically insignificant isolated points or short segments that fail to effectively delineate the structural framework.

4.4.3. Kerry Dataset

To evaluate the network's capability in identifying more elusive and intricate fracture systems within field data, our approach was tested on the New Zealand Kerry 3D seismic dataset. Specifically, a sub-volume characterized by abundant faults, covering inline range 100–196, crossline range 200–712, and time samples 80–400 ms, was utilized for this assessment.

This dataset is particularly challenging, as it comprises a unique class of geologically present faults with atypical geophysical responses—specifically, fracture surfaces that have formed without significant vertical or horizontal displacement of the strata on either side (as highlighted by the green circle in Figure 12). These "no-displacement faults" pose a formidable challenge to traditional identification methods that primarily rely on the principle of seismic horizon offset.

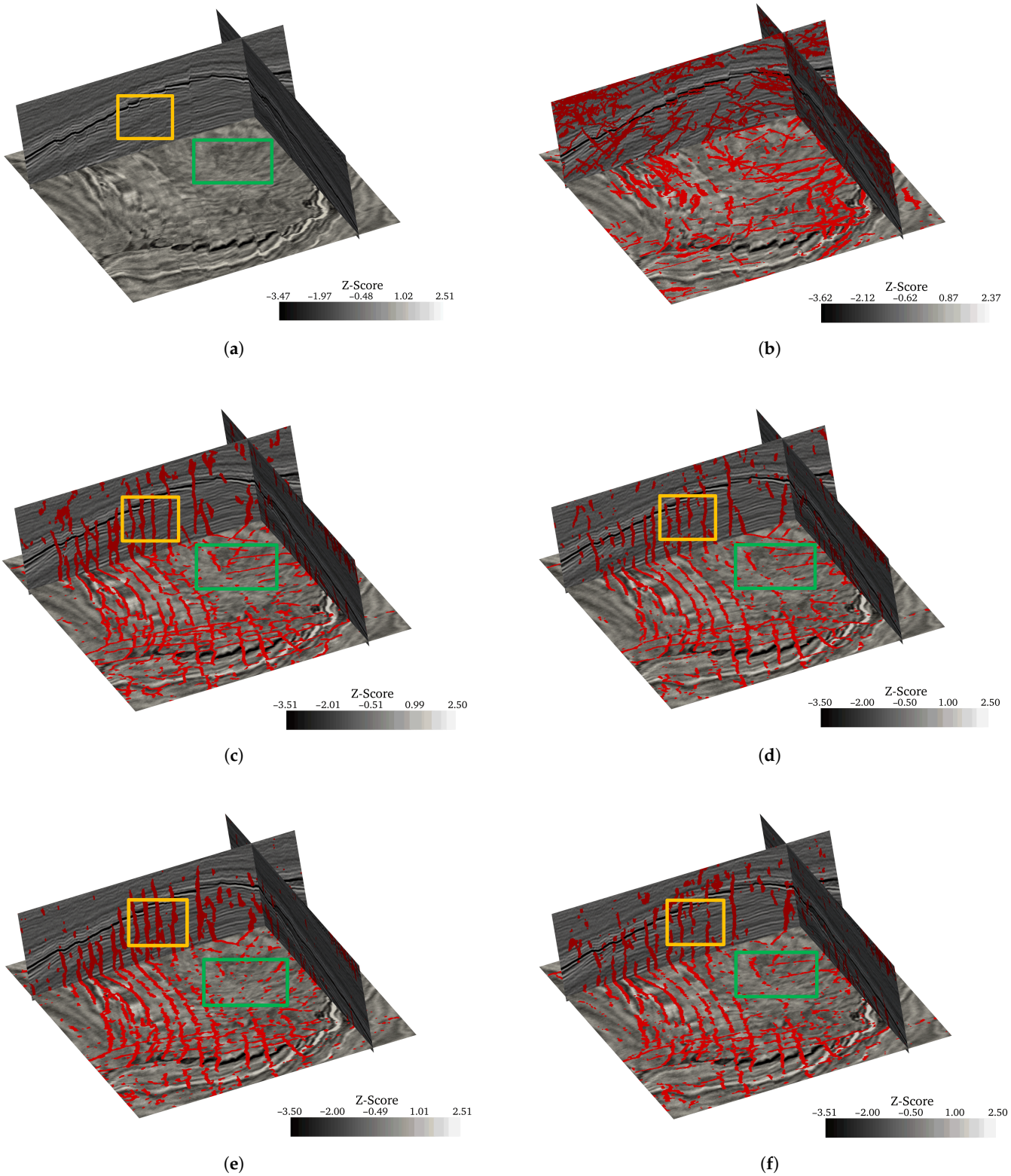


Figure 11. Comparison of fault identification results on the F3 dataset. (**a**) Original seismic section. Fault identification results (overlaid in red) are shown for (**b**) Ant Tracking algorithm, (**c**) SwiftSeisAWNet (Ours), (**d**) FaultSeg3D, (**e**) TransUNet, and (**f**) ResUNet. Highlighted regions (yellow and green boxes) show areas of notable difference.

Mainstream deep learning networks, including FaultSeg3D (Figure 12d), TransUNet (Figure 12e), and ResUNet (Figure 12f), alongside traditional ant-tracking algorithms (Figure 12b), uniformly failed to effectively delineate these crucial geological structures.

This is primarily because the feature extraction and fusion mechanisms within these networks or algorithms tend to prioritize strong amplitude contrasts and pronounced geometric offsets associated with macroscopic fractures. Consequently, they often interpret these subtle discontinuity signals as mere background noise or stratigraphic variations, leading to significant information omission.

It is important to note that the field seismic datasets utilized in this section, namely the F3 and Kerry block, do not have publicly available, ground-truth fault annotations. Consequently, a direct quantitative evaluation using metrics such as the Dice coefficient is not feasible. While new manual annotations could be created, they would introduce a significant degree of interpreter-dependent subjectivity. Therefore, to ensure a robust and objective assessment, we adopted a comparative qualitative analysis framework, a standard and accepted practice in seismic interpretation under such circumstances. The performance of all models, including our proposed method and the baselines, was systematically evaluated against a consistent set of geologically driven criteria: (1) fault continuity and coherence, (2) boundary sharpness, and (3) overall structural plausibility against the seismic reflectors. This approach allows for a rigorous relative assessment of performance, establishing which model produces the most geologically meaningful and interpretable results.

In contrast, the proposed SwiftSeis-AWNet network, leveraging its attention mechanism, dynamically assesses and weights feature information across diverse scales. This capability enables it to acutely capture subtle seismic waveform response changes directly attributable to the rock fracturing itself, independent of any accompanying stratigraphic displacement. Even in the absence of macroscopic displacement, this module can discern spatially continuous, subtle patterns associated with fracture surfaces within multi-scale feature maps and assign them higher weights during the feature fusion process. Consequently, the network not only successfully identified this challenging fault but also delineated its surface with remarkable clarity and continuity, distinctly differentiating it from the surrounding geological background. This comparative experiment powerfully validates SwiftSeis-AWNet's unique advantages in fine structural analysis, particularly its substantial potential in identifying early-stage, subtle, or atypical fracture architectures.

4.5. Ablation Experiments

To precisely quantify the contribution of the proposed AWSDI module and evaluate its effectiveness within the SwiftSeis-AWNet architecture for 3D seismic fault identification, an extensive ablation research was designed and conducted in this section. This research aims to isolate and validate the crucial role of the AWSDI module by comparatively analyzing the impact of different feature fusion strategies on overall network performance.

We systematically compared the performance of three network configurations on the same test dataset using common quantitative indices, including Accuracy, Precision, Recall, IoU, Dice Coefficient, and Loss Function Value, alongside three structural integrity indices. The detailed quantitative evaluation results are presented in Table 3.

Table 3. Quantitative evaluation of different configurations in the ablation research for SwiftSeis-AWNet.

Method	Accuracy	Precision	Recall	IoU	Dice	FCS	ASSD (Voxels)	HD95 (Voxels)	Loss
MedNeXt	0.9711	0.8359	0.8289	0.7135	0.8340	0.9213	1.1128	6.0267	0.0762
Simplified MedNeXt (Baseline)	0.9685	0.8299	0.8235	0.7092	0.8223	0.9194	1.1368	6.2940	0.0829
Simplified MedNeXt + SDI	0.9702	0.8527	0.8493	0.7337	0.8491	0.9257	1.0719	5.8279	0.0610
Simplified MedNeXt + AWSDI(SwiftSeis-AWNet)	0.9794	0.8615	0.8730	0.7546	0.8607	0.9326	0.9371	5.3781	0.0492

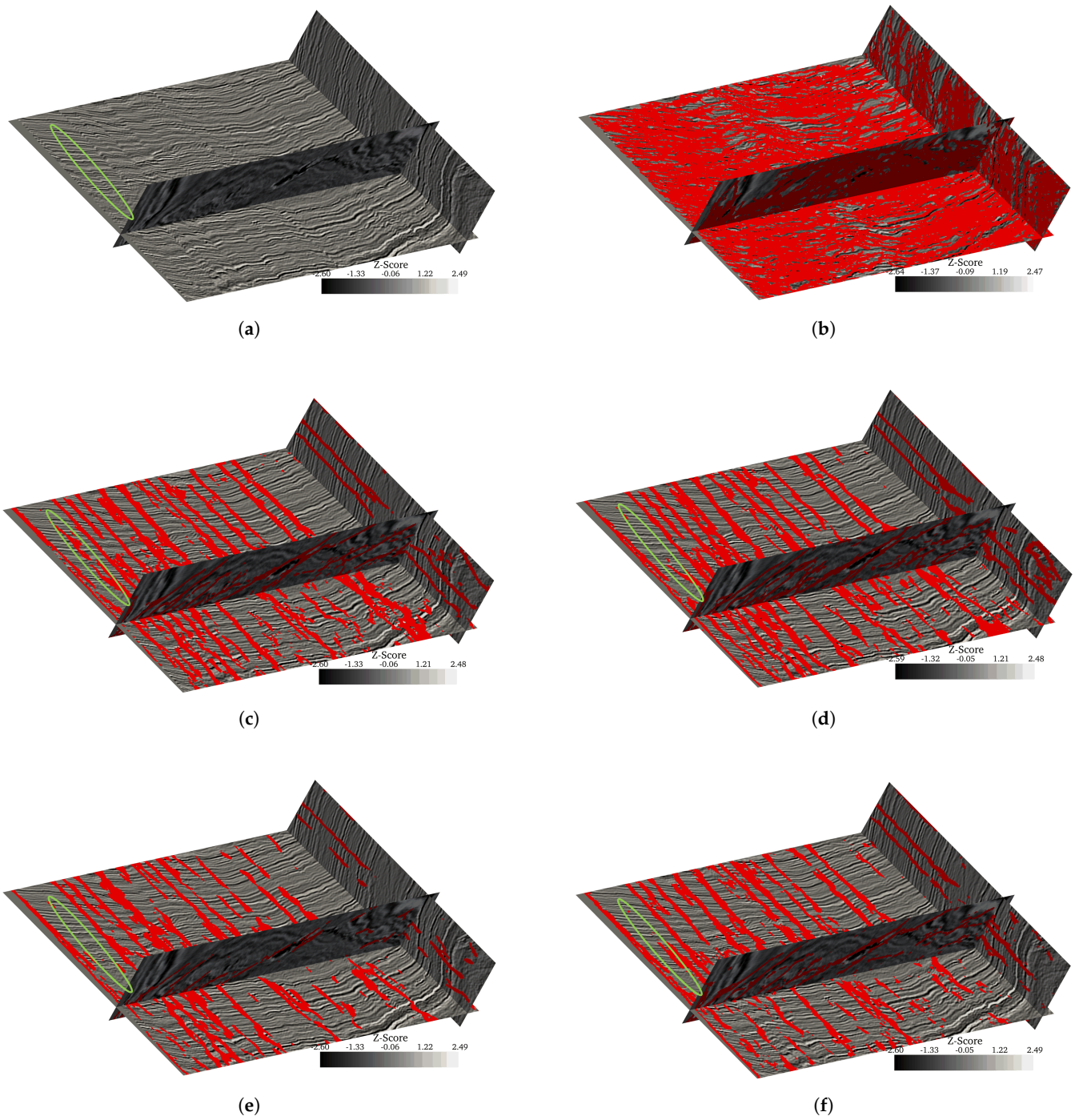


Figure 12. Comparison of fault identification results on the Kerry dataset. (**a**) Original seismic section. Fault identification results (overlaid in red) are shown for (**b**) the Ant Tracking algorithm, (**c**) SwiftSeis-AWNet (Ours), (**d**) FaultSeg3D, (**e**) TransUNet, and (**f**) ResUNet. Highlighted regions (green circle) show areas of notable difference.

A comprehensive analysis of the experimental results leads to the following key conclusions: The negligible performance difference between the original and our simplified MedNeXt empirically validates that our architectural streamlining effectively reduces complexity without a significant trade-off in representational power. The SwiftSeis-AWNet network consistently demonstrated optimal performance across all evaluation indices. It not only achieved the highest values in traditional indices measuring voxel-level segmentation accuracy and possessed the lowest loss function value but also led in structural

indices reflecting geological structure fidelity: securing the highest FCS and the lowest ASSD and 95th Percentile HD95.The results from SwiftSeis-AWNet and Simplified MedNeXt + SDI underscore the pivotal role of the attention-weighted mechanism. A clear performance degradation across all indices was observed when AWSDI was replaced with the attention-free SDI module. This unequivocally confirms that the attention mechanism effectively optimizes the multi-scale feature fusion process, not only enhancing pixel-level prediction accuracy but also critically improving the continuity and geometric precision of the predicted fault structures. Furthermore, the comparison between Simplified MedNeXt + SDI and the baseline Simplified MedNeXt highlights the necessity of a dedicated multi-scale fusion module over simple feature merging. Simplified MedNeXt + SDI consistently outperformed the baseline network, which relies on element-wise addition for fusion, across all evaluated indices. This demonstrates that, compared to basic skip-connection fusion methods, designing specialized multi-scale feature injection and fusion mechanisms is essential for enhancing both voxel-level accuracy and structural preservation in fault identification.

To provide visual corroboration for the quantitative ablation results, this study presents a qualitative comparison on representative slices from the F3 and Kerry datasets (Figures 13 and 14). Comparing the original MedNeXt (c) with our Simplified MedNeXt (d), this study observed no significant degradation in fault identification, confirming that our backbone simplification effectively reduces complexity without compromising core performance. The introduction of the SDI module (e) substantially suppresses the background noise and artifacts present in the baseline result (d). Finally, the addition of our AWSDI module in SwiftSeis-AWNet (f) not only retains this noise robustness but also markedly enhances the sharpness and continuity of fine fault details, which appear smoother and less defined in the SDI-only version (e).

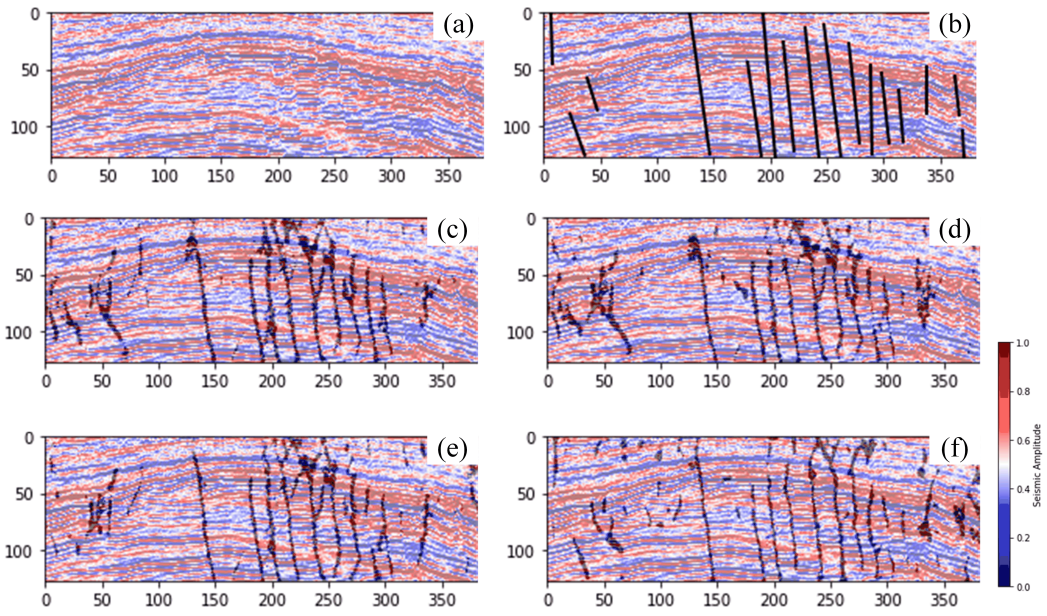


Figure 13. Ablation study results for fault identification on the F3 field seismic dataset. (**a**) Original seismic image. (**b**) Ground truth fault labels annotated by geologists. (**c**) Prediction results from the MedNeXt network. (**d**) Prediction results from the Simplified MedNeXt network. (**e**) Prediction results from the Simplified MedNeXt + SDI network. (**f**) Prediction results from the Simplified MedNeXt + AWSDI (SwiftSeis-AWNet) model.

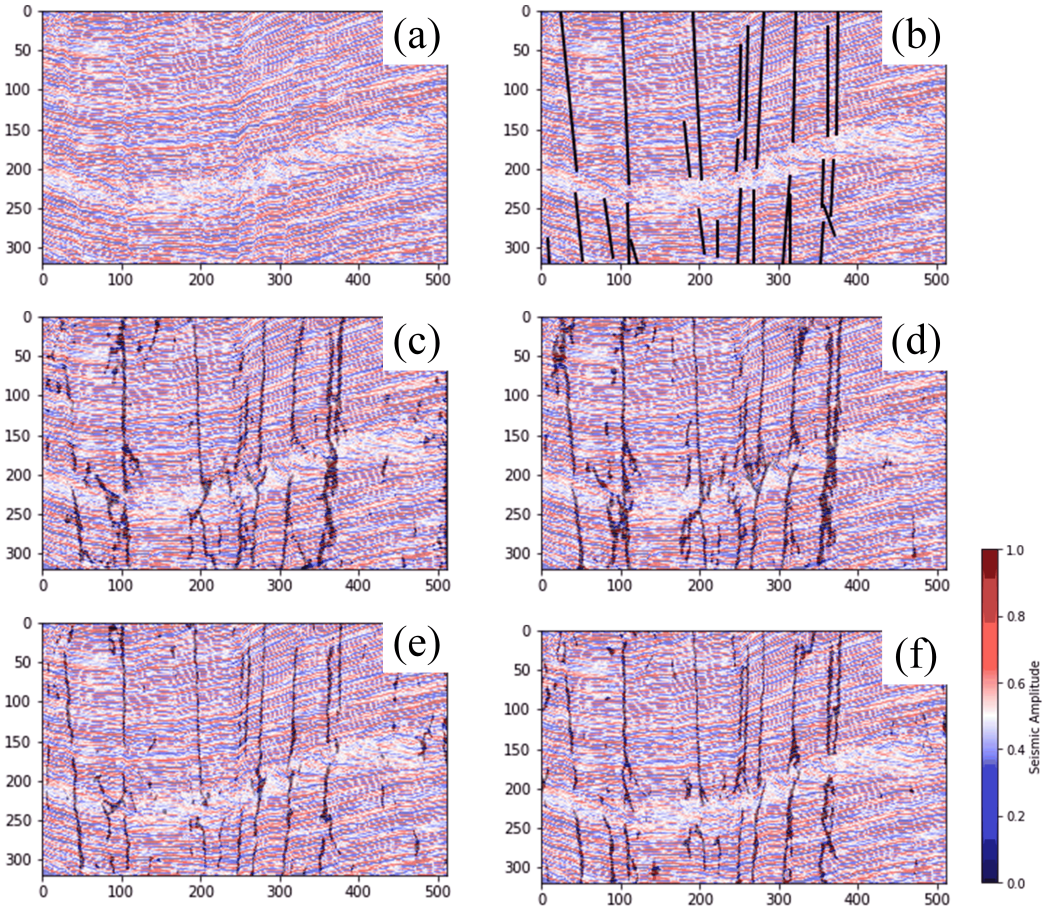


Figure 14. Ablation study results for fault identification on the Kerry field seismic dataset. (**a**) Original seismic image. (**b**) Ground truth fault labels annotated by geologists. (**c**) Prediction results from the MedNeXt network. (**d**) Prediction results from the Simplified MedNeXt network. (**e**) Prediction results from the Simplified MedNeXt + SDI network. (**f**) Prediction results from the Simplified MedNeXt + AWSDI (SwiftSeis-AWNet) model.

5. Discussion

The experimental results presented in this study demonstrate that our proposed SwiftSeis-AWNet achieves state-of-the-art performance in 3D seismic fault identification, consistently outperforming established baseline models across both quantitative metrics and qualitative geological plausibility. This success stems from a synergistic combination of a lightweight, domain-adapted backbone architecture and a novel, attention-driven feature fusion mechanism. In the following sections, we will delve deeper into the key aspects of our methodology, discussing the rationale behind our architectural design choices, the implications of our findings for industrial-scale applications, and the inherent limitations of the current study that pave the way for future research.

5.1. Rationale for Architectural Design Choices

The selection of the expansion ratio R within our network's inverted residual blocks is a critical design choice that directly mediates the balance between model expressiveness and computational complexity. In this study, R was uniformly set to two, a decision informed by a comprehensive consideration of model efficiency, established design paradigms, and task-specific characteristics. Primarily, adhering to our core objective of constructing a "swift" and lightweight network, a conservative expansion ratio is a key strategy for managing the parameter count and computational load (FLOPs), which is crucial when processing large-scale 3D seismic volumes. Furthermore, this choice draws inspiration from the suc-

cess of canonical efficient architectures like MobileNetV2 [27], which have demonstrated that modest expansion factors can achieve an excellent trade-off between performance and efficiency. More importantly, for the specific task of fault identification, which involves recognizing geometrically structured features, an excessively high expansion ratio risks creating an "over-parameterized" model. Such a model may be prone to learning noise or irrelevant stratigraphic details from the data, rather than the essential structural characteristics of the faults themselves. Consequently, a moderate expansion ratio helps to instill an appropriate inductive bias in the model, mitigating overfitting and ultimately optimizing overall performance.

5.2. Basis of Model's Generalization to Complex Faults

The model's ability to learn and identify complex and subtle faults (such as the no-displacement faults discussed in Section 4.4.3) is attributed to the following geological diversity contained within the synthetic volumes:

(1) Spatially Varying Displacements and "Quasi-Zero-Displacement" Zones: A key feature, as detailed by Wu et al. [22], is that fault displacements are not constant but vary spatially, decreasing from a maximum at the fault's center to zero at its edges and tips. This design is paramount, as it naturally creates a vast number of samples representing low-to-zero displacement regions. The seismic response in these zones mimics that of subtle, no-displacement fractures, characterized by faint waveform distortions, amplitude dimming, or phase discontinuities, rather than obvious reflector offsets. This provides the model with critical training examples for learning non-typical fault responses.

(2) Diversity in Fault Geometry and Morphology: The synthesis process incorporates a high degree of randomization in fault dip, strike, and morphology. This generates a wide array of fault types, from high-angle normal faults to complex networks involving intersections, splays, and en-echelon patterns. The fault thickness is also implicitly varied through the labeling process, where fault labels span two pixels adjacent to the fault plane. This comprehensive geometric diversity compels the model to learn a more generalized and robust representation of fault-related seismic textures, preventing it from relying on any single, idealized fault pattern.

5.3. Implications for Industrial-Scale Applications

Beyond academic metrics, the industrial viability of a deep learning model is determined by a holistic balance of accuracy, speed, and resource requirements. It is important to contextualize the "efficiency" of SwiftSeis-AWNet in this light. While its single-patch inference time may not be the absolute fastest among all tested models, its computational profile, characterized by a low parameter count and FLOPs, offers a distinct advantage in resource efficiency. A comparison with TransUNet, a high-performing baseline, illustrates this point: SwiftSeis-AWNet achieves comparable or superior accuracy while achieving a 22.4% reduction in inference time (predicting a 128^3 volume in 93.8 ms compared to TransUNet's 120.9 ms), with a reduction of approximately 98% in parameter count and FLOPs. This substantial improvement in resource efficiency translates into tangible industrial benefits. The significantly lower memory and computational footprint can reduce hardware costs and broaden access to advanced AI interpretation. For basin-scale surveys, this efficiency can enhance scalability and potentially accelerate overall project turnaround times. Furthermore, it makes interactive workflows more feasible, where an interpreter can rapidly iterate on the model. In essence, SwiftSeis-AWNet demonstrates a strong balance between achieving high accuracy and maintaining a high degree of computational

and resource economy, positioning it as a practical and scalable solution for real-world industrial applications.

A pertinent consideration for any new method is its practical feasibility on large-scale industrial surveys, where data volumes often exceed the dimensions used during training. The architecture of SwiftSeis-AWNet allows for a degree of flexibility in input size. This was demonstrated in our application to the field datasets, where both the F3 and Kerry data volumes, which have dimensions different from the 128^3 training cubes, were processed directly. However, for industrial-scale surveys measured in terabytes, processing the entire volume at once is generally infeasible due to universal GPU memory constraints. In such contexts, a standard patch-based processing workflow is employed. This procedure involves systematically dividing the large volume into smaller, overlapping 3D patches of a manageable size (e.g., 128^3 voxels). Each patch is then processed individually by the trained network. Finally, the resulting output patches are reassembled into a single, full-sized volume, with predictions in the overlapping regions blended to ensure a seamless result. The lightweight nature of SwiftSeis-AWNet, characterized by its low resource requirements per patch, makes this workflow highly efficient and scalable, positioning it as a practical solution for real-world industrial implementation.

The preceding discussions highlight that the strength of SwiftSeis-AWNet lies not in a single metric but in its synergistic performance across multiple dimensions. Our architectural choices have yielded a model that is not only accurate at the voxel level but also robust in preserving geological structures, efficient in its use of computational resources, and flexible enough for practical large-scale deployment. This holistic balance between precision, structural fidelity, and practical efficiency is what ultimately defines its value for real-world seismic interpretation challenges.

6. Conclusions

This research introduced SwiftSeis-AWNet, a novel deep learning framework specifically engineered to address the critical shortfalls of existing state-of-the-art AI methods in 3D seismic fault identification—namely, their insufficient edge precision and poor noise robustness. The framework is built upon an optimized MedNeXt backbone, enhanced by integrating a SDI module. Critically, to overcome information loss in the SDI module, we designed the innovative AWSDI module. This core contribution uses a dynamic weighting mechanism to adaptively fuse multi-scale features, directly confronting the aforementioned challenges.

Comprehensive evaluations confirmed the superiority of our approach. On a public synthetic dataset, SwiftSeis-AWNet outperformed mainstream models like ResUNet, FaultSeg3D, and TransUNet, not only in traditional voxel-level metrics but also in advanced geological structure indices (FCS, ASSD, HD95), which validated its enhanced ability to preserve fault continuity and geometric fidelity. Furthermore, generalization tests on the F3 and Kerry field datasets demonstrated its practical utility, yielding interpretations that were demonstrably sharper and more structurally coherent than those from both traditional algorithms and competing deep learning models. Ablation studies decisively verified that the AWSDI module is the key driver of these improvements, enhancing sensitivity to subtle fractures and overall result clarity.

However, we acknowledge the limitations of this study. While SwiftSeis-AWNet shows strong resilience to Gaussian noise, its performance against other noise types remains to be quantified. It is also important to clarify the implication of our model's lightweight nature. While its low parameter count and FLOPs do not necessarily guarantee the absolute fastest single-patch inference time, they signify a crucial advantage in deployment agility and resource efficiency. Our model achieves a superior balance of accuracy and

macro-level efficiency with significantly lower resource consumption, making advanced AI-powered interpretation more accessible. Moreover, the model's generalization capabilities, while promising, require further validation across a wider spectrum of complex geological settings. Future work will focus on incorporating multi-modal seismic attributes to improve geological interpretability and exploring self-supervised learning to mitigate the dependency on large annotated datasets.

Author Contributions: Conceptualization, A.L., R.L., Y.Z. and L.Z.; Methodology, A.L., R.L. and Y.G.; Software, A.L., Y.Z. and S.L.; Validation, Y.Z. and S.L.; Formal analysis, A.L., R.L. and L.Z.; Investigation, R.L., Y.Z., S.L. and Y.S.; Resources, A.L.; Data curation, R.L. and Y.S.; Writing—original draft, R.L., Y.Z. and S.L.; Writing—review & editing, A.L., R.L., Y.Z. and L.Z.; Visualization, S.L. and Y.S.; Supervision, A.L.; Project administration, L.Z.; Funding acquisition, A.L. and L.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research is jointly supported by the following projects: the Scientific Research Startup Fund Project of China University of Petroleum (Beijing) at Karamay Campus, "Evaluation of Shale Formation Fracturability Based on Seismic Rock Physics (XQZX20240015)"; the Distinguished Young Scientists Fund of the Natural Science Foundation of Xinjiang Uygur Autonomous Region, "Mechanism of Seismic Imaging and Omnidirectional Velocity Modeling Methods for Ultra-Deep Layers in the Central Tarim Basin (2024D01E08)"; the Scientific Research Startup Fund Project of China University of Petroleum (Beijing) at Karamay Campus, "Anisotropy Analysis and Correction Methods for Wide-Azimuth Seismic Data in Shale Oil Exploration (XQZX20240029)"; the Provincial Key Research and Development Plan of Xinjiang Uygur Autonomous Region (2024B01016, 2024B01016-2, 2024B01016-3); and the "Tianchi Talent" Innovation Leadership Program of Xinjiang Uygur Autonomous Region.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The synthetic 3D seismic dataset used in this research is publicly available. It was originally created [22] and can be accessed at: https://drive.google.com/open?id=1I-kBAfc_ag68xQsYgAHbqWYdddk4XHHd, accessed on 12 October 2024.

Acknowledgments: During the preparation of this manuscript, the authors used Gemini 2.5 pro for the purposes of language polishing and grammar checking. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

SDI	Semantics and Detail Infusion
AWSDI	Attention-Weighted Semantics and Detail Infusion
FCS	Fault Connectivity Score
DW	Depthwise Convolution
GNorm	Group Normalization
DS	Deep Supervision
PW	Pointwise Convolution
BCE	Binary Cross-Entropy
IoU	Intersection over Union
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
SNR	Signal-to-noise Ratios
ASSD	Average Symmetric Surface Distance

HD95 Hausdorff Distance 95%

Appendix A. Evaluation Indices

Appendix A.1. Voxel-Level Segmentation Accuracy Indices

- **Accuracy :** Measures the overall proportion of correctly classified voxels:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{A1}$$

- **Precision:** Quantifies the proportion of true fault voxels among all voxels predicted as faults (i.e., the positive predictive value):

$$\text{Precision} = \frac{TP}{TP + FP + \varepsilon}, \tag{A2}$$

- **Recall:** Indicates the proportion of actual fault voxels that are correctly identified by the network (i.e., sensitivity):

$$\text{Recall} = \frac{TP}{TP + FN + \varepsilon}, \tag{A3}$$

- **IoU:** Quantifies the overlap between the predicted segmentation and the ground truth, also known as the Jaccard Index:

$$\text{IoU} = \frac{TP}{TP + FP + FN + \varepsilon}, \tag{A4}$$

- **Dice coefficient:** A harmonic mean of Precision and Recall, this metric is widely used to assess the similarity between predicted and ground truth segmentations:

$$\text{Dice} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN + \varepsilon}, \tag{A5}$$

  where ε is a smoothing coefficient, typically set to a small constant value (e.g., 1×10^{-7} in this research), to prevent division by zero errors in the denominator.

Appendix A.2. Fault Morphology and Structure Evaluation Indices

Appendix A.2.1. Fault Connectivity Score (FCS)

$$\text{FCS} = \exp\left(-\frac{\text{abs}(N_{pred} - N_{gt})}{N_{gt} + \varepsilon}\right), \tag{A6}$$

where ε is a small constant to prevent division by zero. The value of FCS ranges from (0, 1]. A value closer to 1 indicates that the fragmentation degree of the predicted fault is more consistent with the true label, and the structure is more complete.

Appendix A.2.2. Average Symmetric Surface Distance (ASSD)

$$\text{ASSD}(S_{pred}, S_{gt}) = \frac{1}{|S_{pred}|} \sum_{v_{pred} \in S_{pred}} \min_{v_{gt} \in S_{gt}} |v_{pred} - v_{gt}| + \frac{1}{|S_{gt}|} \sum_{v_{gt} \in S_{gt}} \min_{v_{pred} \in S_{pred}} |v_{gt} - v_{pred}|/2, \tag{A7}$$

where $|\cdot|$ represents the Euclidean distance. A smaller ASSD value indicates that the two surfaces are closer on average. The unit is voxels.

Appendix A.2.3. Hausdorff Distance 95% (HD95)

$$\mathrm{HD95}(S_{pred}, S_{gt}) = \max\left(K_{95\%}\left\{ \min_{v_{gt}\in S_{gt}} |v_{pred} - v_{gt}| \mid v_{pred} \in S_{pred} \right\}, K_{95\%}\left\{ \min_{v_{pred}\in S_{pred}} |v_{gt} - v_{pred}| \mid v_{gt} \in S_{gt} \right\} \right), \tag{A8}$$

where $\mathcal{K}_{95\%}\{\cdot\}$ denotes taking the 95th percentile of the set of distance values. A smaller HD95 indicates a smaller maximum deviation of the predicted surface. The unit is voxels.

References

1. Bahorich, M.; Farmer, S. 3-D seismic discontinuity for faults and stratigraphic features: The coherence cube. *Lead. Edge* **1995**, *14*, 1053–1058. [CrossRef]
2. Marfurt, K.J.; Kirlin, R.L.; Farmer, S.L.; Bahorich, M.S. 3-D seismic attributes using a semblance-based coherency algorithm. *Geophysics* **1998**, *63*, 1150–1165. [CrossRef]
3. Gersztenkorn, A.; Marfurt, K.J. Eigenstructure-based coherence computations as an aid to 3-D structural and stratigraphic mapping. *Geophysics* **1999**, *64*, 1468–1479. [CrossRef]
4. Van Bemmel, P.P.; Pepper, R.E. Seismic Signal Processing Method and Apparatus for Generating a Cube of Variance Values. U.S. Patent 6,151,555, 21 November 2000.
5. Randen, T.; Pedersen, S.I.; Sønneland, L. Automatic extraction of fault surfaces from three-dimensional seismic data. In Proceedings of the SEG International Exposition and Annual Meeting, San Antonio, TX, USA, 9–14 September 2001; p. SEG-2001.
6. Aqrawi, A.A.; Boe, T.H. Improved fault segmentation using a dip guided and modified 3D Sobel filter. In *SEG Technical Program Expanded Abstracts 2011*; Society of Exploration Geophysicists: Houston, TX, USA, 2011; pp. 999–1003.
7. Bakker, P. *Image Structure Analysis for Seismic Interpretation*; Citeseer: Princeton, NJ, USA, 2002.
8. Hale, D. Structure-oriented smoothing and semblance. *CWP Rep.* **2009**, *635*, 261–270.
9. Wu, X. Directional structure-tensor-based coherence to detect seismic faults and channels. *Geophysics* **2017**, *82*, A13–A17. [CrossRef]
10. Hale, D. Methods to compute fault images, extract fault surfaces, and estimate fault throws from 3D seismic images. *Geophysics* **2013**, *78*, O33–O43. [CrossRef]
11. Wu, X.; Hale, D. 3D seismic image processing for faults. *Geophysics* **2016**, *81*, IM1–IM11. [CrossRef]
12. Neff, D.B.; Grismore, J.R.; Lucas, W.A. Automated Seismic Fault Detection and Picking. U.S. Patent 6,018,498, 16 March 2000
13. Cohen, I.; Coult, N.; Vassiliou, A.A. Detection and extraction of fault surfaces in 3D seismic data. *Geophysics* **2006**, *71*, P21–P27. [CrossRef]
14. Wu, X.; Zhu, Z. Methods to enhance seismic faults and construct fault surfaces. *Comput. Geosci.* **2017**, *107*, 37–48. [CrossRef]
15. Pedersen, S.I.; Randen, T.; Sønneland, L.; Steen, Ø. Automatic fault extraction using artificial ants. In Proceedings of the Seg International Exposition and Annual Meeting, Salt Lake City, UT, USA, 6–11 October 2002; p. SEG-2002.
16. Pedersen, S.; Skov, T.; Hetlelid, A.; Fayemendy, P.; Randen, T.; Sønneland, L. New paradigm of fault interpretation. In Proceedings of the 73rd Annual International Meeting, SEG, Expanded Abstracts, Dallas, TX, USA, 26–31 October 2003; Society of Exploration Geophysicists: Houston, TX, USA, 2003; pp. 350–353.
17. Wu, X.; Fomel, S. Automatic fault interpretation with optimal surface voting. *Geophysics* **2018**, *83*, O67–O82. [CrossRef]
18. Dorn, G.; James, H. Automatic fault extraction of faults and a salt body in a 3-D survey from the Eugene Island area, Gulf of Mexico. In Proceedings of the Aapg International Conference and Exhibition, Expanded Abstracts, Paris, France, 11–14 September 2005; Volume 19.
19. Admasu, F.; Back, S.; Toennies, K. Autotracking of faults on 3D seismic data. *Geophysics* **2006**, *71*, A49–A53. [CrossRef]
20. Di, H.; Wang, Z.; AlRegib, G. Why using CNN for seismic interpretation? An investigation. In Proceedings of the SEG International Exposition and Annual Meeting, Anaheim, CA, USA, 14–19 October 2018; p. SEG-2018.
21. Wu, X.; Shi, Y.; Fomel, S.; Liang, L. Convolutional neural networks for fault interpretation in seismic images. In Proceedings of the SEG International Exposition and Annual Meeting, Anaheim, CA, USA, 14–19 October 2018; p. SEG-2018.
22. Wu, X.; Liang, L.; Shi, Y.; Fomel, S. FaultSeg3D: Using synthetic data sets to train an end-to-end convolutional neural network for 3D seismic fault segmentation. *Geophysics* **2019**, *84*, IM35–IM45. [CrossRef]
23. Yang, D.; Cai, Y.; Hu, G.; Yao, X.; Zou, W. Seismic fault detection based on 3D Unet++ model. In Proceedings of the SEG International Exposition and Annual Meeting, Online, 11–16 October 2020; p. D031S039R002.
24. Yan, B.; Qian, L.; Zhao, J.; Li, M.; Pan, R. Fault Identification Based on W-Net in 3-D Seismic Images. *IEEE Geosci. Remote Sens. Lett.* **2024**, *21*, 7504805. [CrossRef]

25. Cui, L.; Huang, Y.; Niu, Y.; Cui, H.; Tao, Y.; Qian, L.; Zhao, J. MS-Unet: A Multi-Scale Feature Fusion U-Net for 3D Seismic Fault Detection. *Processes* **2025**, *13*, 1976. [CrossRef]
26. Hu, G.; Hu, Z.; Liu, J.; Cheng, F.; Peng, D. Seismic fault interpretation using deep learning-based semantic segmentation method. *IEEE Geosci. Remote Sens. Lett.* **2020**, *19*, 7500905. [CrossRef]
27. Roy, S.; Koehler, G.; Ulrich, C.; Baumgartner, M.; Petersen, J.; Isensee, F.; Jaeger, P.F.; Maier-Hein, K.H. Mednext: Transformer-driven scaling of convnets for medical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer—Assisted Intervention, Vancouver, BC, Canada, 8–12 October 2023; pp. 405–415.
28. Peng, Y.; Chen, D.Z.; Sonka, M. U-net v2: Rethinking the skip connections of u-net for medical image segmentation. In Proceedings of the 2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI), Houston, TX, USA, 14–17 April 2025; pp. 1–5.
29. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
30. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
31. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
32. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.