

ARTICLE

DB-MS-CAN: A robust deep learning framework for microseismic velocity inversion via Dual-Branch Attention and Multi-Scale Feature Aggregation

Aoyu Feng¹, Binxin Hu^{2*} , Feng Zhu^{1*}, Yang Gao³, Feifei Jiang³, Danya Ma⁴, and Rong Liu²

¹Shandong Key Laboratory of Optoelectronic Sensing Technologies / National-Local Joint Engineering Laboratory for Energy and Environment Fiber Smart Sensing Technologies, Laser Institute, Qilu University of Technology (Shandong Academy of Sciences), Jinan, Shandong, China

²Qilu Institute of Technology, Jinan, Shandong, China

³Institute of Rock and Soil Mechanics, Chinese Academy of Sciences, Wuhan, Hubei, China

⁴Jinan Institute of Semiconductor Devices, Jinan, Shandong, China

Abstract

Accurate real-time assessment of underground stress fields is critical for mine safety, yet conventional inversion methods struggle to balance precision with computational efficiency. To address these challenges, we propose the Dual-Branch Multi-Scale Convolutional Attention Network (DB-MS-CAN), which employs a decoupling-and-aggregation strategy. The framework integrates a Dual-Branch Fusion Module (BFM) to isolate sparse transient signals from persistent mining noise and a Multi-Scale Dense Feature Aggregation (MDFA) module to reconstruct complex geological structures across varying spatial scales. To ensure rigorous validation, the model was evaluated on both geologically constrained real-world datasets and an independent synthetic benchmark governed by wave propagation physics. Results demonstrate that DB-MS-CAN achieves a structural similarity index (SSIM) of 0.925 and an root mean square error (RMSE) reduction of approximately 40% compared to traditional ray-based tomography. The proposed model achieved significant performance improvements compared to advanced baselines, including CNN-LSTM and Vision Transformers. Notably, the model maintains high fidelity (correlation coefficient > 0.90) under extreme 0 dB SNR conditions and achieves an inference speed of ~7 seconds per event on a single GPU, outperforming iterative solvers by orders of magnitude. This framework provides a robust and efficient solution for dynamic hazard early warning in deep mining environments.

*Corresponding authors:

Binxin Hu (bxhu@qlu.edu.cn);
Feng Zhu (feng.zhu@sdlaser.cn)

Citation: Feng A, Hu B, Zhu F, *et al.* DB-MS-CAN: A robust deep learning framework for microseismic velocity inversion via Dual-Branch Attention and Multi-Scale Feature Aggregation. *J Seismic Explor.* 2026;35(3):025480121. doi: 10.36922/JSE025480121

Received: November 26, 2025

Revised: February 15, 2026

Accepted: March 9, 2026

Published online: May 6, 2026

Copyright: © 2026 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Keywords: Microseismic inversion; Deep learning; Dual-Branch Attention; Multi-Scale Feature Aggregation

1. Introduction

As coal mining operations extend into deeper strata, dynamic disasters such as rockbursts have become increasingly frequent, posing severe threats to production

safety.¹ High stress concentration within coal-rock masses is the primary driver of such disasters; therefore, real-time and high-precision assessment of the subsurface stress state is critical for production safety.² Microseismic velocity inversion, which reconstructs the subsurface velocity field distribution, serves as a critical technique for safety assurance by intuitively reflecting stress concentration areas.³ However, the complex geological environment of deep mines, coupled with strong background noise and the stringent requirement for real-time warning, imposes significant challenges on existing inversion methods.⁴

Traditional physics-driven methods, such as full waveform inversion FWI and computerized tomography (CT), struggle to balance accuracy with efficiency. Although FWI theoretically achieves high precision, its non-convex objective function leads to a heavy reliance on the initial model, and its iterative computation is too costly for real-time monitoring.⁵⁻⁸ Conversely, while CT offers high efficiency, its reliance on ray theory often results in blurred boundaries and low resolution in complex heterogeneous structures.⁹⁻¹¹ Therefore, there is an urgent need for a data-driven approach that can deliver high-resolution imaging without the computational burden of iterative physics solvers.¹²

In recent years, deep learning (DL) has shown promise in geological inversion by learning non-linear mappings from data.^{13,14} However, directly applying generic computer vision architectures, such as U-Net or Convolutional Neural Networks (CNNs), to microseismic velocity inversion is challenging because these techniques overlook the unique physical characteristics of microseismic data.¹⁵⁻¹⁷ Unlike natural images with persistent objects, effective microseismic signals, such as P/S wave onsets, are sparse and transient, occupying only a fraction of the time series.^{18,19} Generic convolutional operations treat all time steps equally, often obscuring critical spikes in continuous background noise, leading to feature dilution and noise overfitting.²⁰

Furthermore, subsurface media exhibit extreme multi-scale heterogeneity, ranging from meter-scale micro-fractures to kilometer-scale fault zones.²¹ Standard encoders with fixed receptive fields either fail to capture minor anomalies, resulting in underfitting, or blur large geological boundaries, leading to over-smoothing.²²

Existing deep learning solvers usually handle spatiotemporal features ambiguously. Strictly spatial models, such as the standard U-Net, inherently struggle to capture the sequential dependencies in microseismic waveforms, resulting in temporal feature loss. While standard recurrent models, such as the Long Short-Term

Memory (LSTM), can capture temporal dependencies, they incur high computational costs and do not explicitly address the extreme spatial-scale differences inherent in subsurface structures.

Moreover, emerging advanced paradigms also face specific limitations in microseismic velocity inversion. While Physics-Informed Neural Networks (PINNs) enforce consistency, they often suffer from over-smoothing and boundary blurring due to soft regularization constraints, which attenuates the peak velocity values of stress concentration zones. Conversely, Vision Transformer-based models, despite their superior global context modeling, inherently introduce blocky artifacts and jagged edges owing to their patch-based processing mechanisms, thereby compromising the pixel-level fidelity required for precise hazard detection.

To address these challenges, this study presents the Dual-Branch Multi-Scale Convolutional Attention Network (DB-MS-CAN). Built upon a “Decoupling-and-Aggregation” principle, the framework establishes a unified mapping mechanism to resolve the inherent conflict between transient signal extraction and multi-scale structural reconstruction. Recognizing that mining environments require resolving sharp velocity gradients, we employ a U-Net-based CNN architecture; its local receptive fields and skip connections are inherently superior for preserving high-frequency spatial information and recovering velocity extremums compared to global optimization or patch-based methods. Specifically, the architecture integrates two core components: the Dual-Branch Fusion Module (BFM), which decouples high-frequency transients from background noise, and the Multi-Scale Dense Feature Aggregation (MDFA) module, designed to capture geological heterogeneities across varying spatial scales. To ensure industrial relevance, the model was validated against geologically constrained offline tomography derived from borehole logging and active surveys. Experimental results demonstrate that DB-MS-CAN achieves a structural similarity index (SSIM) of 0.925, significantly outperforming baseline models and proving its efficacy for high-precision monitoring in complex coal mine settings.

2. Methods

2.1. Overall model architecture

As illustrated in [Figure 1](#), the framework integrates two specialized components into a U-Net-like backbone.²³⁻²⁵ The overall data flow within the network advances through a progressive processing cascade, comprising three distinct integration phases:

The network extracts hierarchical features via successive convolutions.²⁶⁻²⁸ Unlike standard architectures, we embed the MDFA module directly into the deep encoding layers. By applying parallel dilated convolutions, MDFA encodes spatially enriched features that represent multi-scale structural information before significant downsampling occurs.

The data flow then converges at the network bottleneck, where the global context is compressed. At this critical junction, the BFM module is embedded to process the latent features. Through a dual-branch attention mechanism, BFM filters features along the temporal dimension, ensuring that only significant transient phases are preserved while background noise is suppressed.

Finally, the decoder reconstructs the velocity field via progressive upsampling.²⁹ The reconstruction process synergistically leverages the multi-scale spatial representations encoded by MDFA and the noise-suppressed temporal features refined by BFM, effectively mitigating the trade-off between structural continuity and boundary sharpness.

2.2. Multi-Scale Dense Feature Aggregation (MDFA)

To explicitly resolve the multi-scale heterogeneity of subsurface structures,³⁰ the MDFA module (Figure 2) employs parallel dilated convolutions with varying dilation rates as a scale-adaptive feature extractor.

We implemented the MDFA module using multi-scale

dilated convolution and a spatial recalibration mechanism.

Multi-Scale Dilated Convolution: Distinct from generic multi-scale modules, we designed the dilation rates (d) to correspond to specific geological targets.³¹ The small dilation branch with $d = 6$ is designed to capture high-frequency scattering effects induced by local micro-fractures and small voids. Conversely, the large dilation branches with $d = 12$ and 18 expand the receptive field to model macroscopic waveform variations caused by large-scale faults and lithological boundaries.

Mathematically, let the input feature tensor be $X \in \mathbb{R}^{C_{in} \times H \times W}$. The output Y_i of the i -th dilated convolutional branch is defined as:

$$Y_i = \text{ReLU}(W_i \otimes_{d_i} X + b_i), i \in \{1, 2, 3\} \quad (1)$$

where $\otimes_{d_i}$ denotes the 2D dilated convolution operation with a dilation rate of d_i ; W_i and b_i are the learnable weight matrix and bias vector for the i -th branch; $\text{ReLU}(\cdot)$ is the Rectified Linear Unit activation function.

By aggregating these multi-branch features, the MDFA module effectively constructs a feature pyramid to ensure that the inverted velocity model preserves both the sharpness of micro-anomalies and the continuity of macro-structures.

Feature Fusion and Spatial Recalibration: The final high-accuracy velocity model is a synthesized representation of all these multi-scale features. However, simple

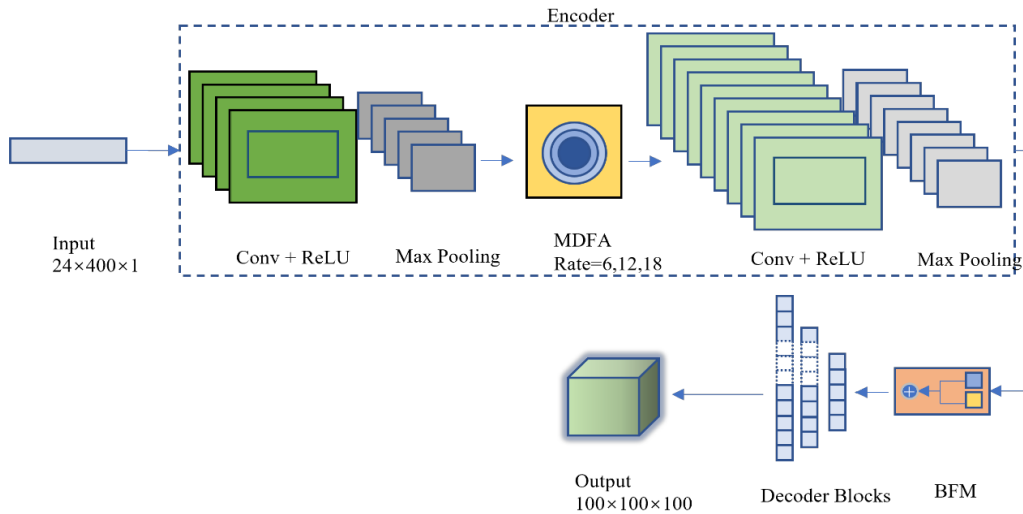


Figure 1. The overall network architecture

Abbreviations: BFM: Dual-Branch Fusion Module; MDFA: Multi-Scale Dense Feature Aggregation; ReLU: Rectified Linear Unit.

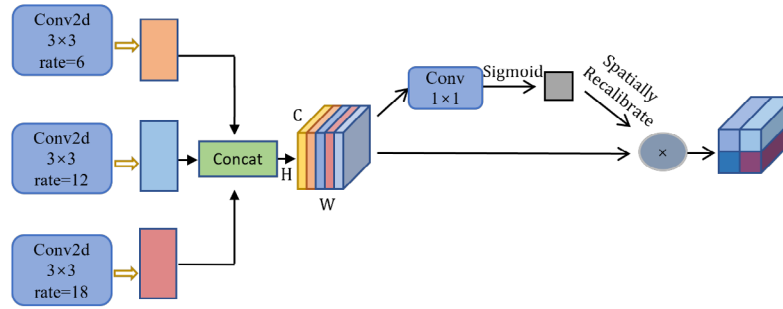


Figure 2. Multi-Scale Dense Feature Aggregation (MDFA) structure diagram

feature concatenation often struggles with semantic inconsistencies among features extracted at different scales. To ensure the coherent integration of information from micro-anomalies to macroscopic structures, we introduced a Spatial Recalibration Mechanism into our proposed MDFA module. This mechanism adaptively learns spatial attention weights to screen and strengthen the spatial locations most critical for the velocity inversion task.

First, the multi-scale features from all branches are concatenated along the channel dimension to obtain the aggregated feature U :

$$U = \text{Concat}(Y_1, Y_2, Y_3) \in \mathbb{R}^{3C_{out} \times H \times W} \quad (2)$$

where Y_1, Y_2 , and Y_3 correspond to the feature maps extracted by the branches with dilation rates $d = 6, 12$, and 18 , respectively. C_{out} denotes the number of output channels for each individual branch (set to 64 in our experiments), and H and W represent the spatial height and width of the feature maps, which remain consistent due to padding operations.

Subsequently, a 1×1 convolutional layer is applied to the aggregated feature U to perform dimensionality reduction and cross-channel interaction, followed by a Sigmoid function to generate the spatial attention map A :

$$A = \sigma(W_{cal} \otimes_1 U + b_{cal}) \in \mathbb{R}^{1 \times H \times W} \quad (3)$$

where W_{cal} and b_{cal} denote the learnable weight matrix and bias vector of the spatial calibration convolution, respectively. The symbol $\otimes_1$ represents the 1×1 convolution operation, and $\sigma(\cdot)$ denotes the Sigmoid activation function used to normalize the attention weights to the range $(0, 1)$.

Finally, the output Y_{out} —the spatially recalibrated feature—is obtained by applying the attention map to the aggregated feature via element-wise multiplication (Hadamard product):

$$Y_{out} = A \odot U \quad (4)$$

where $\odot$ is the Hadamard product.

This mechanism ensures that the model dynamically adjusts its focus based on the complexity of the geological structure. It produces a high-precision velocity model that retains local fracture details while accurately representing overall structural trends.

Figure 3 shows the intermediate feature maps from the encoder's bottleneck layer, illustrating how our proposed MDFA module addresses spatial-scale conflicts for a typical microseismic event. The red dashed lines are the ground truth reference for the velocity anomaly boundary. As observed in Figure 3A, the branch driven by small dilation rates ($d = 6$) prioritizes high-frequency scattering information. The resulting map exhibits rich textural details indicative of micro-fracture zones but suffers from a diffuse spatial distribution. In contrast, the large dilation branch ($d = 18$), shown in Figure 3C, captures the low-frequency envelope of the geological body, demonstrating strong morphological consistency with the ground-truth boundary. However, the internal activation intensity remains relatively subdued. Ultimately, the final spatially recalibrated output (Figure 3D) effectively synthesizes these complementary features. It retains the sharp, coherent structural boundaries derived from the macro-branch while simultaneously preserving the high-intensity internal details from the micro-branch. This visual evidence confirms that MDFA successfully mitigates the trade-off between boundary precision and internal detail resolution.

2.3. Dual-Branch Fusion Module (BFM)

The Dual-Branch Fusion Module (BFM), illustrated in Figure 4, employs a dual-path architecture. It is designed to mitigate the blurring of sparse high-frequency transients by decoupling the input into two streams and fusing them via a channel-spatial attention mechanism.^{32,33}

Dual-Branch Structure: The local branch functions as a transient feature extractor. To adaptively capture waveform spikes with varying pulse widths, we employed a multi-kernel structure consisting of parallel 3×3 , 5×5 , and 7×7 convolutions to adaptively emphasize instantaneous P/S onsets. Mathematically, given the input feature X , the local feature map Y_{local} is defined as:

$$Y_{local} = \text{ReLU}\left(\sum_{k \in \{3, 5, 7\}} W_{k \times k} \otimes X + b_{k \times k}\right) \quad (5)$$

where $W_{k \times k}$ and $b_{k \times k}$ denote the learnable weights and biases of the convolutional kernels with sizes 3×3 , 5×5 , 7×7 , respectively; and $\otimes$ denotes the convolution operation; and $\text{ReLU}(\cdot)$ is the activation function.

We created a Global Branch to serve as a context encoder, using Global Average Pooling (GAP) to model

signal decay and background noise trends.³⁴ This technique provides the necessary context to distinguish real events from coherent noise.^{35,36} The global feature vector V_{global} is computed as:

$$V_{global} = W_{1 \times 1} \otimes (\text{GAP}(X)) + b_{1 \times 1} \quad (6)$$

where $\text{GAP}(\cdot)$ denotes the Global Average Pooling operation, and $W_{1 \times 1}$ and $b_{1 \times 1}$ are the 1×1 Convolutional kernel weights and bias, respectively. Finally, the global feature is upsampled back to the dimension of Y_{local} to form the global feature map Y_{global} .

Channel-Spatial Attention Mechanism and Feature Fusion: Following the feature extraction, the dual streams are integrated via element-wise summation ($F = Y_{local} \oplus Y_{global}$) and refined through an attention mechanism specifically tailored for multi-channel microseismic data. The Channel Attention mechanism serves as a sensor weighting filter. The mechanism autonomously learns to prioritize channels with coherent signal patterns. This is formulated as:

$$M_c(F) = \sigma(\text{FC}(\text{GAP}(F)) + \text{FC}(\text{GMP}(F))) \quad (7)$$

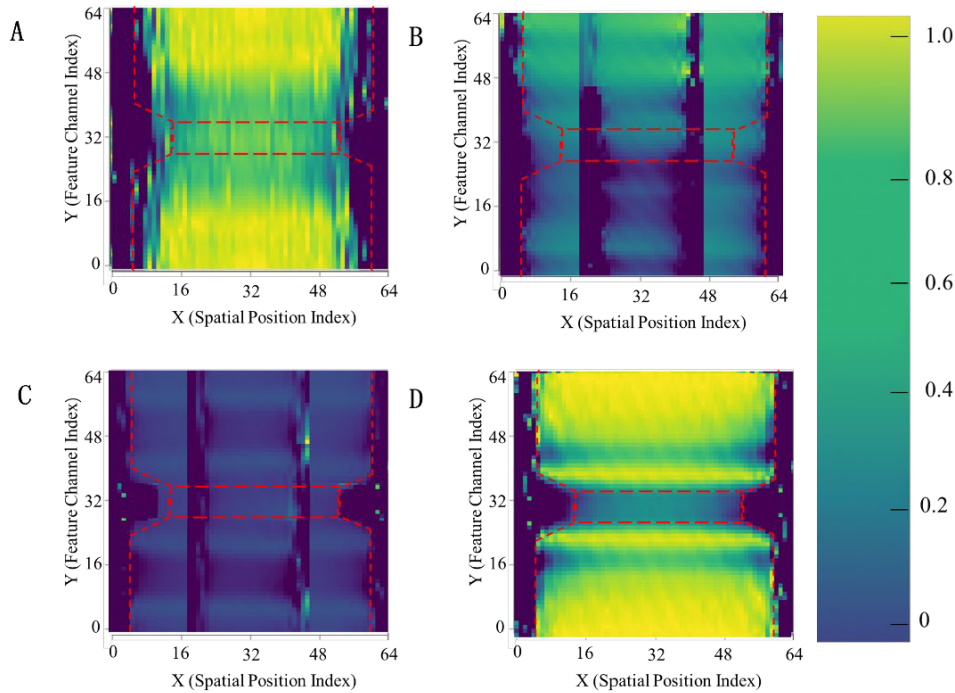


Figure 3. Visualization of intermediate feature maps from Multi-Scale Dense Feature Aggregation (MDFA) branches. The color scale indicates normalized activation intensity (0.0–1.0). The X and Y axes represent spatial position indices (64×64). Red dashed lines denote the reference boundary of the velocity anomaly. (A)–(C) Feature maps corresponding to dilation rates $d = 6, 12$, and 18 , respectively. (D) The final spatially recalibrated feature map after fusing (A), (B), and (C).

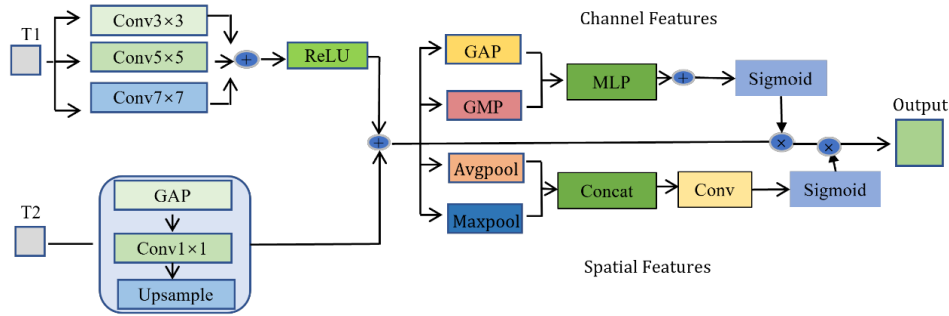


Figure 4. Branch Fusion Module (BFM) structure diagram

Abbreviations: GAP: Global Average Pooling; GMP: Global Max Pooling; MLP: Multilayer Perceptron; ReLU: Rectified Linear Unit.

where $GAP(\cdot)$ and $GMP(\cdot)$ denote the Global Average Pooling and Global Max Pooling operations, respectively. FC denotes the shared Fully Connected layers applied to the feature vectors, and σ denotes the Sigmoid activation function.

At the same time, the Spatial Attention mechanism provides temporal focus by highlighting critical arrival windows, reducing pre-event noise and coda waves. Its calculation is defined as:

$$M_s(F) = \sigma(W_{1 \times 1} \otimes [AvgPool(F); MaxPool(F)]) \quad (8)$$

where $AvgPool(\cdot)$ and $MaxPool(\cdot)$ represent average-pooling and max-pooling operations along the channel axis. The symbol $[\cdot]$ denotes the concatenation operation. $W_{1 \times 1}$ represents the learnable weights of the 1×1 convolution (denoted as $\otimes$) used to compress the channel information into a single spatial map.

The final output Y_{fused} is obtained by modulating the fused features via element-wise multiplication, ensuring the network prioritizes significant spatiotemporal regions:

$$Y_{fused} = (M_c(F) \odot F) \odot M_s(F) \quad (9)$$

where $\odot$ denotes the Hadamard product (element-wise multiplication).

Figure 5 visualizes the intermediate activations of the BFM module to verify its decoupling capability. As shown in Figure 5A, the local branch exhibits a sharp activation peak at $t=1.8s$ (normalized intensity approaching 0.95) effectively isolating the high-frequency P/S wave onsets from the background; whereas the global branch (Figure 5B) suppresses these transient pulses to retain the wavefield background trend. Regarding channel reliability, Figure 5C demonstrates distinct selectivity: the network assigns dominant weights to sensors with strong

coherence (e.g., Channel 7 = 0.78 and Channel 3 = 0.70), while heavily suppressing the noise-dominated Channel 2 (weight = 0.13). This significant quantitative disparity (> 6 -fold difference) confirms that the mechanism has implicitly learned to prioritize high-quality signals through optimization. Finally, the spatial attention mask in Figure 5D reaches its maximum intensity (> 0.9) precisely at the 1.8 s window, aligning accurately with the signal arrival time, verifying that the model autonomously achieves temporal focusing without explicit physical programming.

3. Results

3.1. Data acquisition and preprocessing

The model proposed in this study was trained and validated using real-world microseismic monitoring data collected from an operating mine. The data acquisition zone is located near the deep mining face of a large-scale coal mine in Shanxi Province. Within this region, a high-density microseismic monitoring network comprising 24 triaxial geophones was deployed to capture microseismic events and subsurface velocity variations with high precision. The detailed sensor layout is illustrated in Figure 6.

3.1.1. Construction of the ground truth velocity field

To ensure the physical fidelity of the training labels, the velocity models used as ground truth, denoted as V_{gt} , were constructed through a geologically constrained offline inversion workflow, integrating *in situ* exploration data. This distinguishes them from the unconstrained real-time CT results used for baseline comparison.

The label generation process involves three key stages:

- (i) **Reference Model Construction:** We utilized detailed geological logging data and exploration reports to build an initial velocity model. Precise travel-time pickings of P-wave and S-wave arrivals from a substantial historical dataset (exceeding 1,200 events)

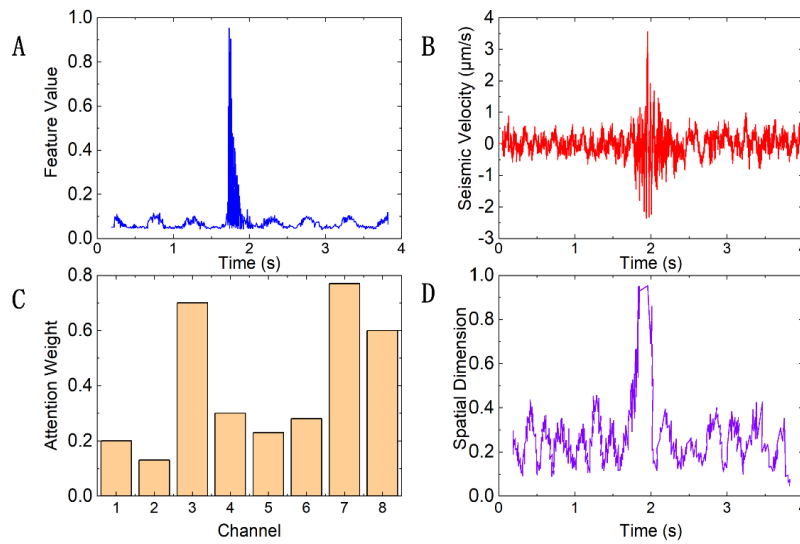


Figure 5. Visualization of intermediate variables within the Branch Fusion Module (BFM) module. (A) Local branch output captures high-frequency transient features, specifically responding to the sharp P/S wave onsets. (B) Global branch output represents the low-frequency background trends and global wavefield context. (C) Channel attention weights indicates the learned reliability of each sensor channel (treating sensors as feature channels). (D) Spatial attention weights defines the attention mask along the feature map's spatial dimension (which corresponds to the temporal axis in microseismic waveforms).

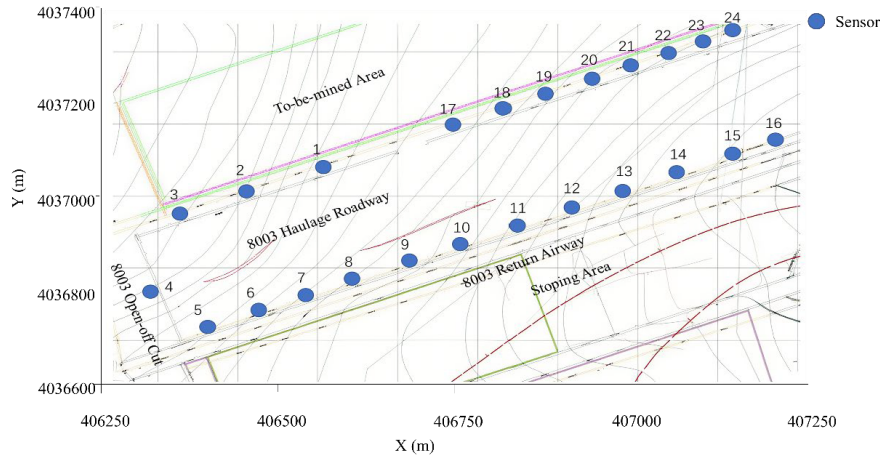


Figure 6. Sensor distribution diagram

were employed as constraints.³⁷

- (ii) **Constrained Offline Tomography:** Based on the reference model, we executed a Geologically Constrained offline inversion utilizing the Simultaneous Iterative Reconstruction Technique (SIRT). To ensure the high fidelity of the ground truth (V_{gt}), the input travel-time data underwent rigorous manual inspection by experienced geophysicists, achieving a picking precision of approximately 10 ms, followed by 30 iterations of recursive refinement to

minimize the RMS error.

- (iii) **Grid Discretization:** The final inverted velocity field was discretized into a $100 \times 100 \times 100$ three-dimensional grid with a uniform spatial resolution of $10\text{m} \times 10\text{m} \times 10\text{m}$. The resulting velocity values range between 3500 m/s and 5500 m/s.

The physically constrained tomographic result (V_{gt}) integrates actual geological measurements with large-scale microseismic data,³⁸ providing a high-quality target for the

deep learning model.

3.1.2. Data input and standardization

The model input consists of multi-channel waveforms recorded by 24 sensors for each microseismic event. All waveforms were uniformly truncated to 400 time sampling points, with a sampling rate of 2 kHz. The input tensor dimension is therefore $X \in \mathbb{R}^{24 \times 400 \times 1}$.

The dataset comprises 1,600 microseismic events. To ensure a rigorous evaluation and prevent data leakage, this dataset is strictly temporally mutually exclusive from the historical dataset (exceeding 1,200 events) used for ground truth construction. We partitioned the data into a training set (1,120 events), a validation set (240 events), and a testing set (240 events).

To eliminate dimensional differences among sensor channels, we standardized the input waveforms using the Z-score transformation. This standardization transformed the data to have a zero mean and unit standard deviation before passing it to the model for training. The output velocity model labels were also Min-Max normalized within a range of [0,1].

To comprehensively evaluate the proposed model's performance, a statistical analysis of the dataset's characteristics was conducted. The training, validation, and testing sets maintained consistent distributions across key attributes, including event magnitude, average signal-to-noise ratio (SNR), and spatial coverage, as detailed in Table 1. This process ensured fairness and the model's generalization assessment. Furthermore, the high-dimensional input tensor provides a rich feature space, offering sufficient information density for the network to extract robust physical representations.

3.2. Model training strategy and configuration

The proposed DB-MS-CAN was implemented using the PyTorch framework on a workstation equipped with an

Intel Core i9-14900HX CPU and an NVIDIA GeForce RTX 4060 GPU. The training process utilized the Adam optimizer³⁹ with an initial learning rate of 1×10^{-4} , dynamically modulated by a Cosine Annealing schedule. We configured a batch size of 16 with a maximum duration of 500 epochs, employing an Early Stopping mechanism (patience = 20) based on validation loss to mitigate overfitting. The mean squared error (MSE) served as the objective function to minimize voxel-wise discrepancies between the predicted velocity field $\hat{V}$ and the ground truth V_{gt} . Regarding model complexity, the network comprises approximately 31.2 million trainable parameters. Notably, the proposed core innovations are designed to be extremely lightweight: the MDFA and BFM modules contain only 35,712 and 170,596 parameters, respectively. Together, they constitute less than 1% of the total model size while contributing to significant performance gains. The training phase required approximately 1.5 minutes per epoch, resulting in a total computation time of roughly 12 hours.

Figure 7 illustrates the training and validation loss curves of the DB-MS-CAN model, which show distinct trends reflecting the model's learning dynamics. As observed, the training loss exhibits a rapid descent within the first 50 epochs, indicating that the gradient descent algorithm effectively optimized the network weights. The validation loss closely follows the training loss trend without diverging, stabilizing at approximately 0.05 after 150 epochs. The minimal gap between the training and validation curves confirms that the Early Stopping mechanism (patience = 20) and Dropout layers successfully prevented overfitting. The smooth convergence behavior in the later stages demonstrates the stability of the proposed DB-MS-CAN architecture under the Adam optimizer.

3.3. Experimental results and analysis

3.3.1. Comparison with advanced deep learning models

Table 1. Microseismic dataset partition and feature statistics for the deep learning model

Dataset	Number of events	Magnitude range (ML)	Average SNR (dB)	Spatial distribution area
Training set	1120	[−1.5, 2.0]	15.2	Covers the entire working face
Validation set	240	[−1.3, 1.8]	14.8	Covers the entire working face
Test set	240	[−1.6, 1.9]	15.1	Covers the entire working face

To validate the effectiveness and superiority of the proposed DB-MS-CAN model, the U-Net, Attention U-Net, CNN,^{40,41} Attention CNN,⁴² CNN-LSTM,⁴³ and DB-MS-CAN architectures were trained and evaluated separately on the same dataset. The quantitative performance comparison is summarized in Table 2. All metrics in this and subsequent tables are reported as mean $\pm$ standard deviation, calculated across the 240 independent events in the test set. The visual comparison is presented in Figure 8.

Figure 8 visually compares the velocity inversion performance across different models, spanning from background low-velocity zones (~ 3500 – 4000 m/s) to high-velocity stress concentration areas (4800 – 5500 m/s). Visual inspection reveals clear limitations in the baseline architectures. The CNN and U-Net models (Figure 8A and 8C) suffer from significant over-smoothing, characterized by blurred boundaries in the left low-velocity anomaly and an inability to resolve fine textures in the bottom-right zone. They also tend to spatially exaggerate the central high-velocity core (> 5200 m/s), leading to an overestimated stress concentration area. These visual deficiencies align with their stagnant performance metrics, with root mean square error (RMSE) values persisting at 2.63 – 2.65 %. While adding spatial attention mechanisms (Figure 8B and 8D) yields slight morphological improvements in transitional layers, these models fail to correct global velocity deviations. Similarly, although the CNN-LSTM model (Figure 8E) improves structural continuity, it introduces local distortions by excessively enlarging the upper high-velocity region.

In contrast, the proposed DB-MS-CAN (Figure 8F) delivers an inversion result most consistent with the geological ground truth. It precisely reconstructs the low-velocity core (3600 – 3800 m/s) and resolves complex velocity gradients, effectively correcting the spatial artifacts observed in the baselines. This visual superiority is supported by rigorous statistical evidence. The model achieves a state-of-the-art SSIM of 0.925 ± 0.008 and reduces the RMSE to 2.17 ± 0.11 %. A paired t -test against the best-performing baseline (CNN-LSTM) yielded a p -value < 0.001 , confirming that this performance gain is statistically significant rather than a result of random chance. Moreover, our method exhibits the lowest standard deviation among all models (*e.g.*, compared to CNN's ± 0.021), indicating that the architecture effectively mitigates stochastic instability and ensures consistent inversion results across diverse geological conditions.

3.3.2. Module ablation experiment

To further assess the specific contributions of the BFM and MDFA modules to the overall model performance, a module ablation study was designed in this research. Four distinct model variants were constructed for comparative analysis:

- (i) Baseline Model (U-Net): Employs only the traditional U-Net structure, without incorporating any additional mechanisms.
- (ii) BFM-Only (U-Net + BFM): Introduces the BFM solely at the bottleneck layer.
- (iii) MDFA-Only (U-Net + MDFA): Integrates the

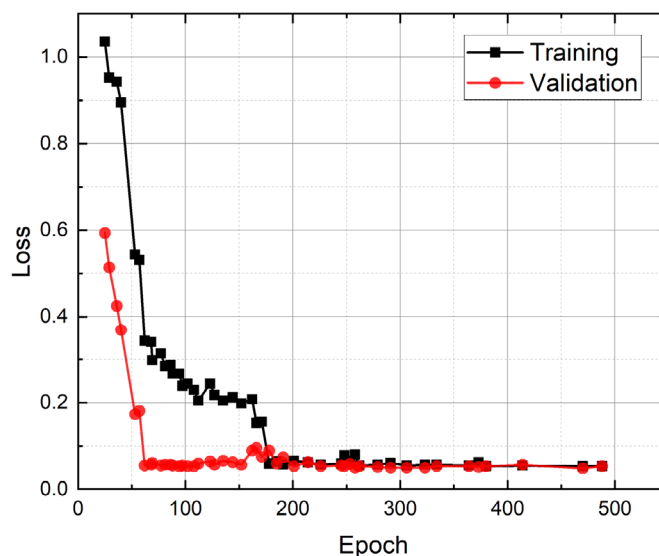


Figure 7. Training loss and validation loss

Table 2. Comparison of inversion results between DB-MS-CAN and existing deep learning models

Model	SSIM (mean $\pm$ SD)	RMSE (%) (mean $\pm$ SD)	R (mean $\pm$ SD)
CNN	0.715 $\pm$ 0.021	2.65 $\pm$ 0.28	0.940 $\pm$ 0.015
U-Net	0.735 $\pm$ 0.019	2.63 $\pm$ 0.25	0.939 $\pm$ 0.013
Attention CNN	0.785 $\pm$ 0.016	2.57 $\pm$ 0.19	0.945 $\pm$ 0.010
CNN-LSTM	0.865 $\pm$ 0.012	2.43 $\pm$ 0.15	0.954 $\pm$ 0.007
Attention U-Net	0.795 $\pm$ 0.014	2.59 $\pm$ 0.17	0.941 $\pm$ 0.009
DB-MS-CAN	0.925 $\pm$ 0.008	2.17 $\pm$ 0.11	0.968 $\pm$ 0.004

Abbreviations: CNN: Convolutional Neural Network; DB-MS-CAN: Dual-Branch Multi-Scale Convolutional Attention Network; LSTM: Long Short-Term Memory; RMSE: Root mean square error; SSIM: Structural similarity index.

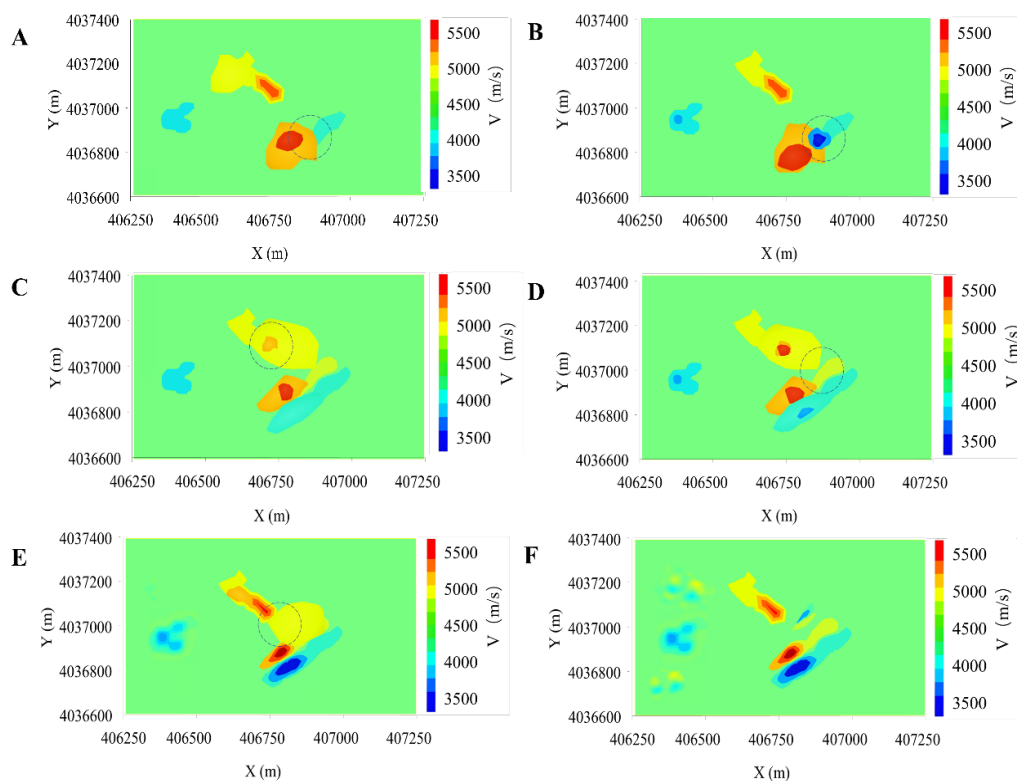


Figure 8. Visual comparison of microseismic velocity inversion results generated by various deep learning models. (A) U-Net; (B) Attention U-Net; (C) CNN; (D) Attention CNN; (E) CNN-LSTM; (F) DB-MS-CAN. The dashed-line circles point out the failure of baseline models to accurately reconstruct the local velocity anomalies, which are well preserved in (F).

Abbreviations: CNN: Convolutional Neural Network; DB-MS-CAN: Dual-Branch Multi-Scale Convolutional Attention Network; LSTM: Long Short-Term Memory.

MDFA mechanism solely within the intermediate layers of the encoder.

- (iv) DB-MS-CAN: Represents the final, comprehensive network architecture proposed in this paper.

Figure 9 visualizes the ablation study results against this benchmark. The baseline U-Net (Figure 9A) exhibits characteristic over-smoothing, failing to resolve the high-velocity core or define the low-velocity zone (< 4000 m/s). This structural deficiency is reflected in the poorest performance metrics: an SSIM of 0.733 ± 0.019 and an RMSE of 2.65 ± 0.26 %.

Integrating specific modules yields distinct improvements. The MDFA module (Figure 9C) enhances macroscopic spatial continuity, evolving the patchy low-velocity structure into a coherent geological body. This improves the SSIM to 0.812 ± 0.015 ; however, the model struggles to recover peak high-velocity intensities, limiting RMSE reduction to 2.54 ± 0.18 %. In contrast, the U-Net-BFM variant (Figure 9B) excels at sharpening local boundaries and filtering noise. It successfully recovers the high-velocity core (> 5000 m/s), reducing the RMSE to 2.48 ± 0.14 %, though the background strata morphology remains discontinuous.

The proposed DB-MS-CAN (Figure 9D) integrates these complementary advantages, retaining MDFA-derived stratification and BFM-captured high-intensity cores. This completeness translates to state-of-the-art performance, achieving a peak SSIM of 0.925 ± 0.008 and a minimum RMSE of 2.17 ± 0.10 %.

Table 3 reveals a progressive reduction in metric variance as modules are integrated (SSIM standard deviation decreases from ± 0.019 to ± 0.008). This trend confirms that introducing domain knowledge—multi-scale constraints and transient signal decoupling—acts

as a strong regularization factor. By constraining the solution space to geologically plausible manifolds, the full model significantly enhances inversion stability and reproducibility.

As additional modules are integrated, the overall running time of the model increases, as illustrated in Figure 10. However, this computational cost is acceptable considering the significant performance improvements.

3.3.3. Comparative verification with known ground truth

To mitigate field label uncertainty and evaluate the theoretical limits of DB-MS-CAN, we developed a synthetic benchmark following the microseismic perturbation framework. Unlike simplified layered models, our benchmark employs a physically-constrained stochastic strategy to embed high-contrast polygonal anomalies ($V > 5200$ m/s) and low-velocity voids ($V < 3800$ m/s). To bridge the ‘simulation-to-reality’ gap, high-order finite-difference modeling was employed, with site-specific industrial noise superimposed to ensure model robustness under complex field SNR conditions.

Utilizing this benchmark, we compare our method against two advanced baselines: PINN⁴⁴ (physics-driven) and a Vision Transformer-based network⁴⁵ (data-driven). To ensure a fair comparison, all baseline models were implemented using standard training protocols widely adopted in recent literature and were trained to convergence under identical hardware environments.

Figure 11A presents the ground truth velocity model, constructed based on rock mechanics principles to simulate a heterogeneous mining environment. Featuring a high-velocity stress concentration zone (> 5000 m/s) and a low-velocity void (< 3800 m/s), the significant contrast

Table 3. Ablation study results of components in the DB-MS-CAN model

Model	SSIM (mean $\pm$ SD)	RMSE (%) (mean $\pm$ SD)	R (mean $\pm$ SD)
U-Net	0.733 ± 0.019	2.65 ± 0.26	0.940 ± 0.014
U-Net-MDFA	0.812 ± 0.015	2.54 ± 0.18	0.947 ± 0.009
U-Net-BFM	0.855 ± 0.011	2.48 ± 0.14	0.951 ± 0.007
DB-MS-CAN	0.925 ± 0.008	2.17 ± 0.10	0.968 ± 0.004

Abbreviations: BFM: Dual-Branch Fusion Module; DB-MS-CAN: Dual-Branch Multi-Scale Convolutional Attention Network; MDFA: Multi-Scale Dense Feature Aggregation; RMSE: Root mean square error; SSIM: Structural similarity index.

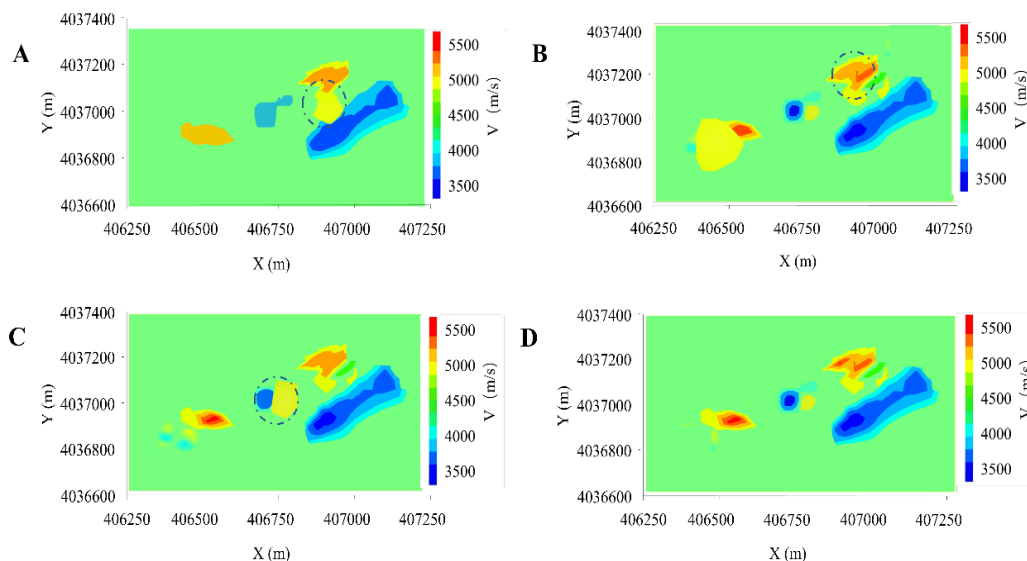


Figure 9. Velocity inversion images of (A) U-Net, (B) U-Net-BFM, (C) U-Net-MDFA, and (D) DB-MS-CAN. The dashed-line circles in (A) through (C) point out the reconstruction defects of the baseline and ablated models compared to the proposed architecture.

Abbreviations: BFM: Dual-Branch Fusion Module; DB-MS-CAN: Dual-Branch Multi-Scale Convolutional Attention Network; MDFA: Multi-Scale Dense Feature Aggregation.

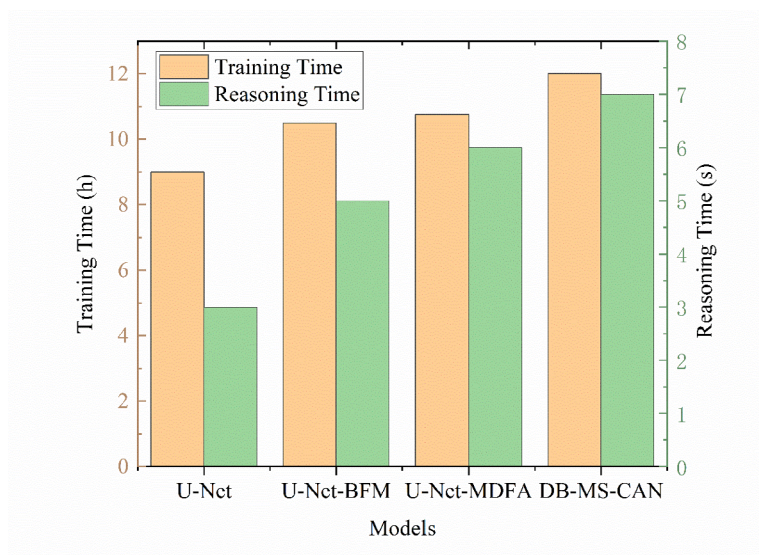


Figure 10. Training time and inference time

Abbreviations: BFM: Dual-Branch Fusion Module; DB-MS-CAN: Dual-Branch Multi-Scale Convolutional Attention Network; MDFA: Multi-Scale Dense Feature Aggregation.

(> 1200 m/s) serves as a precise benchmark for evaluating boundary delineation capabilities.

Against this benchmark, the PINN method (Figure 11B) successfully identifies the relative positions of the low-velocity void (blue region) and the high-velocity stress zone, demonstrating that physical constraints effectively guide the model toward a geologically plausible

solution. However, visual inspection reveals significant over-smoothing, characterized by blurred high-stress zone boundaries and a failure to recover peak intensities. This amplitude attenuation, induced by soft physical regularization, results in suboptimal quantitative metrics: an SSIM of 0.875 ± 0.024 and an RMSE of $2.90 \pm 0.32\%$. The notably large standard deviation (± 0.32) suggests

that convergence remains highly sensitive to initial weight distributions and the balancing of loss terms.

Similarly, the Vision Transformer-based network (Figure 11C) exhibits exceptional global context modeling capabilities. Benefiting from its self-attention mechanism, it achieves accurate macro-scale positioning of anomalies, which is reflected in a high Pearson correlation coefficient ($R = 0.952$). Nevertheless, the result is compromised by visible blocky artifacts and jagged discontinuities along anomaly boundaries, a consequence of its patch-based processing mechanism. This local fragmentation limits pixel-level fidelity, yielding an RMSE of $2.87 \pm 0.25\%$, which offers no significant improvement over the PINN baseline.

In contrast, the proposed DB-MS-CAN (Figure 11D) delivers superior reconstruction performance. It is worth noting that its reconstruction of the homogeneous background is slightly less smooth than that of the PINN, with minor textural fluctuations observable. This is a characteristic behavior of CNN-based architectures inferring from local receptive fields rather than global governing equations. However, this minor background noise does not affect the identification of the high-stress anomaly, which is the primary objective of safety monitoring. Crucially, the model effectively eliminates baseline artifacts, precisely recovering sharp geometric boundaries and physical extremum values (> 5200 m/s). This fidelity translates into state-of-the-art metrics: a peak SSIM of 0.918 ± 0.009 and a minimum RMSE of $1.96 \pm 0.12\%$. The significantly lower variance (± 0.009) demonstrates the architecture's exceptional robustness in resolving complex geological heterogeneities.

3.4. Field validation

3.4.1. Comparison with traditional methods

To comprehensively evaluate the performance of the proposed method, this section compares the DB-MS-

CAN model with two representative traditional velocity inversion techniques: passive seismic tomography (CT) and FWI. For a rigorous comparison, strictly identical test data from the mining area described in Section 3.1 were used for all inversion tasks. The implementation details and configuration strategies for the traditional methods are as follows:

- (i) Passive CT: Implemented using the fast sweeping method (FSM) for forward modeling and the Least Squares with QR-factorization (LSQR) algorithm for inversion, with a grid resolution of 10×10 m.
- (ii) FWI configuration: We employed a time-domain acoustic FWI with a multi-scale inversion strategy to effectively mitigate cycle-skipping. Specifically, a smoothed one-dimensional velocity model derived from pilot surveys was utilized as the initial model. The inversion process was executed sequentially across three frequency bands: 0–30 Hz, 0–60 Hz, and 0–120 Hz. We adopted the L-BFGS optimizer (memory depth of 10) with strict stopping criteria, terminating iterations when the relative change in the objective function fell below 1×10^{-4} or reached a maximum of 50 iterations per band. Finally, the inversion grid was refined to a resolution of 5×5 m to align with the output of the deep learning model.

Traditional methods relied on a workstation with an Intel Xeon Gold 6248R CPU (standard for parallelized iterative solvers), while the DB-MS-CAN inference was conducted on a single NVIDIA RTX 4060 GPU.

The quantitative performance comparison is summarized in Table 5 and Figure 12, while a comparison of the running time is provided in Figure 13.

Constrained by ray-theory approximations, the CT inversion (Figure 12A) exhibits severe spectral leakage and over-smoothing. It fails to resolve fine-grained geological textures or recover velocity extrema, missing both the deep

Table 4. Quantitative performance comparison of different inversion methods on the synthetic benchmark

Method	SSIM (mean $\pm$ SD)	RMSE (%) (mean $\pm$ SD)	R (Pearson) (mean $\pm$ SD)
PINN	0.875 ± 0.024	2.90 ± 0.32	0.924 ± 0.018
Transformer	0.885 ± 0.018	2.87 ± 0.25	0.952 ± 0.012
DB-MS-CAN	0.918 ± 0.009	1.96 ± 0.12	0.976 ± 0.005

Abbreviations: DB-MS-CAN: Dual-Branch Multi-Scale Convolutional Attention Network; PINN: Physics-Informed Neural Network; RMSE: Root mean square error; SSIM: Structural similarity index.

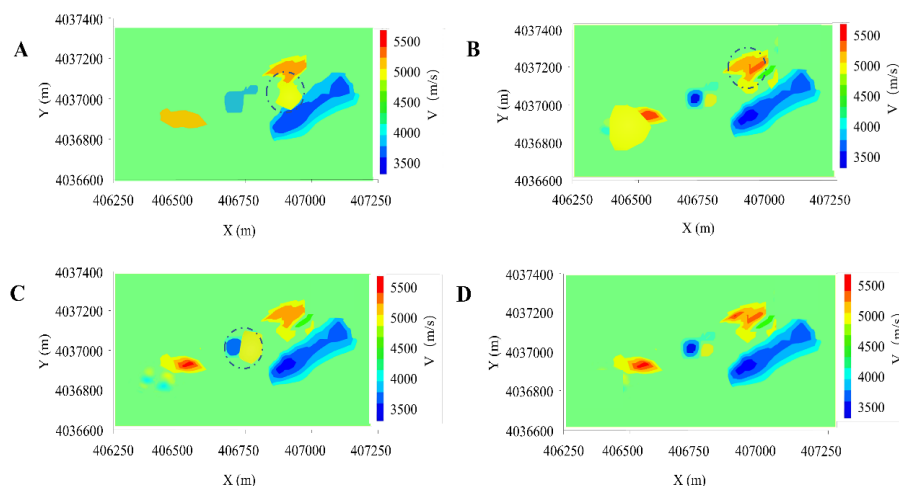


Figure 11. Comparison of velocity inversion results on the synthetic benchmark: (A) Ground truth model; (B) PINN; (C) Vision Transformer; (D) DB-MS-CAN. The dashed-line circles point out the key high-velocity and low-velocity anomalies, serving as a visual reference to compare the exact structures in (A) with the inaccurately reconstructed regions in (B) and (C).

Abbreviations: DB-MS-CAN: Dual-Branch Multi-Scale Convolutional Attention Network; PINN: Physics-Informed Neural Network.

blue intervals (< 3600 m/s) and deep red stress zones (> 5200 m/s). This limitation yields the poorest quantitative performance, characterized by an RMSE of $3.65 \pm 0.52\%$ and an SSIM of 0.720 ± 0.045 . The notably large standard deviation (± 0.52) highlights the inherent instability of ray-based tomography, which remains highly sensitive to ray coverage density and initial model selection.

The FWI result (Figure 12B) demonstrates marginal visual improvement by incorporating waveform information, reducing the RMSE to $2.87 \pm 0.38\%$. However, structural definition remains indistinct, and boundary integrity is compromised by non-physical artifacts arising from cycle-skipping. The suboptimal SSIM (0.815 ± 0.032) and substantial metric variance indicate that optimization-based methods struggle to converge consistently under complex non-linear conditions, often trapping in local minima.

In contrast, the proposed DB-MS-CAN (Figure 12C) delivers the most geologically consistent reconstruction. It precisely delineates complex stratification within low-velocity zones and recovers the full dynamic range of the velocity field. This fidelity translates into state-of-the-art metrics: a peak SSIM of 0.925 ± 0.008 and a minimum RMSE of $2.17 \pm 0.11\%$ (a $\sim 40\%$ error reduction compared to CT). The minimal variance significantly outperforms traditional methods (CT's ± 0.52 and FWI's ± 0.38), demonstrating superior reproducibility for safety-critical monitoring. Additionally, the inference time of ~ 7 seconds achieves near-real-time performance, providing an optimal balance between accuracy and efficiency.

3.4.2. Robustness and noise tolerance experiment

The test dataset utilized in this study is derived from raw *in situ* microseismic monitoring. Consequently, the input waveforms inherently contain complex, real-world mining interference, including mechanical vibrations, electrical hum, and background blasting noise (as visualized in Figure 14). The model's high performance on this test set implicitly validates its robustness against realistic non-Gaussian noise.

To rigorously evaluate the model's theoretical stability limits under strictly controlled SNR conditions, we introduced Gaussian white noise as an additional stress test variable. We constructed a stress-test dataset by superimposing Gaussian white noise onto the test waveforms at SNRs of 10 dB, 5 dB, and 0 dB. The performance degradation of the three methods is quantified in Figure 15.

To rigorously evaluate the theoretical stability limits under strictly controlled SNR conditions, we introduced Gaussian white noise as an additional stress test variable. We constructed a stress-test dataset by superimposing noise onto the test waveforms at varying levels: Noise-free, 10 dB, 5 dB, and 0 dB. The quantitative performance degradation of the three methods is illustrated in Figure 15.

As the noise level increases (*i.e.*, SNR decreases), a general decline in performance is observed across all inversion methods. However, the proposed DB-MS-CAN demonstrates exceptional robustness compared to

Table 5. Performance comparison of DB-MS-CAN with conventional inversion methods

Method	SSIM (mean ± SD)	RMSE (%) (mean ± SD)	R (mean ± SD)
CT wave velocity	0.720 ± 0.045	3.65 ± 0.52	0.85 ± 0.035
FWI	0.815 ± 0.032	2.87 ± 0.38	0.931 ± 0.022
DB-MS-CAN	0.925 ± 0.008	2.17 ± 0.11	0.968 ± 0.004

Abbreviations: CT: Computed tomography; DB-MS-CAN: Dual-Branch Multi-Scale Convolutional Attention Network; FWI: Full waveform inversion; RMSE: Root mean square error; SSIM: Structural similarity index.

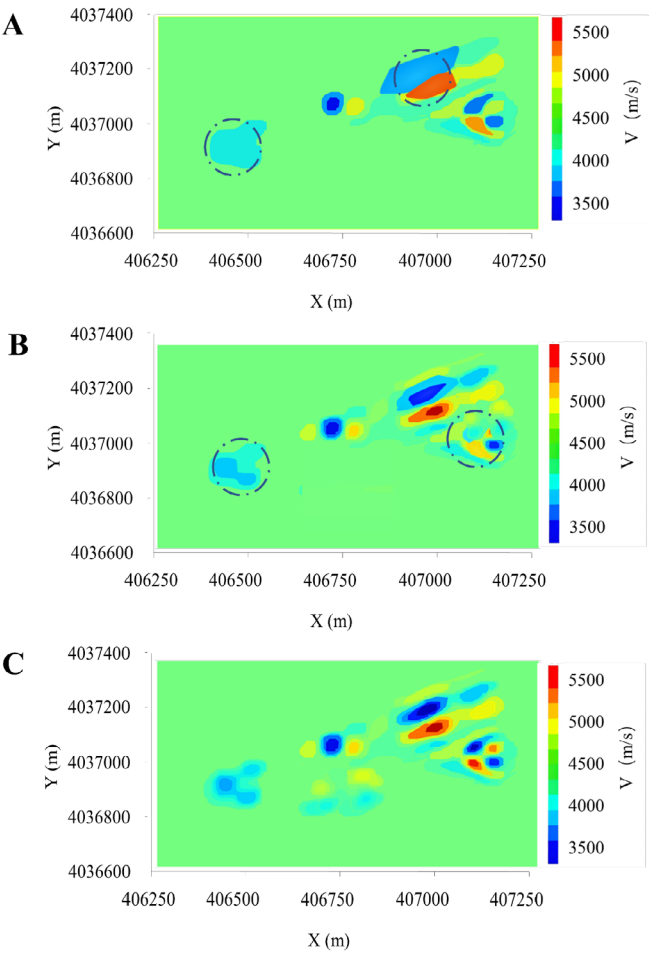


Figure 12. Visual comparison of velocity inversion results between traditional methods and the proposed DB-MS-CAN model. (A) CT velocity inversion image. (B) FWI inversion image. (C) DB-MS-CAN. The dashed-line circles point out local inversion inaccuracies and structural distortions in the traditional methods, which are effectively overcome by the DB-MS-CAN model in (C).
Abbreviations: CT: Computed tomography; DB-MS-CAN: Dual-Branch Multi-Scale Convolutional Attention Network; FWI: Full waveform inversion.

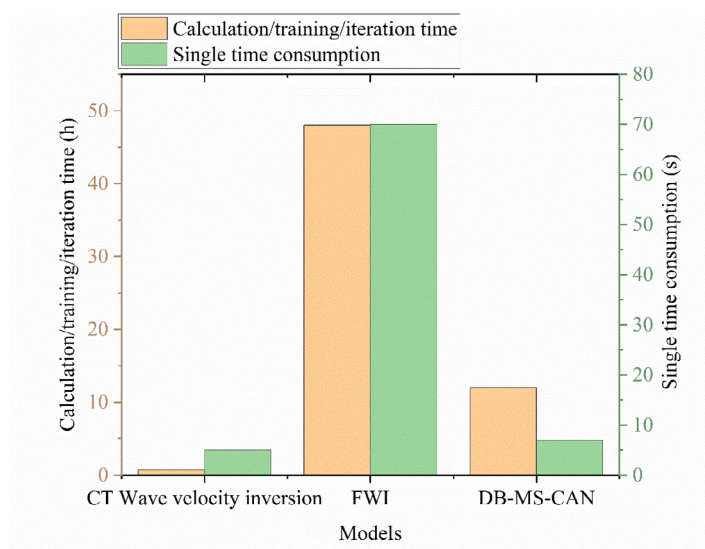


Figure 13. Running time of the three methods

Abbreviations: CT: Computed tomography; DB-MS-CAN: Dual-Branch Multi-Scale Convolutional Attention Network; FWI: Full waveform inversion.

traditional approaches.

First, regarding structural fidelity (Figure 15A), the SSIM of the traditional CT method (black line) drops precipitously from ~ 0.79 to ~ 0.63 as noise intensifies to 0 dB, indicating severe structural distortion. In contrast, DB-MS-CAN (blue line) maintains a high SSIM (> 0.92) even under the harshest 0 dB conditions, significantly outperforming FWI and CT. This suggests that the multi-scale aggregation mechanism effectively preserves geological structural details despite heavy interference.

Second, in terms of numerical accuracy (Figure 15B), the RMSE of the CT method exhibits a sharp linear increase, exceeding 5.4% at 0 dB. Conversely, DB-MS-CAN maintains the lowest error rate, with the RMSE rising only marginally to $\sim 2.6\%$. The gentle slope of the blue curve confirms the model's stability in minimizing velocity deviations.

Finally, the correlation coefficient (R value) in Figure 15 further validates the model's reliability. While the R value for CT and FWI falls to approximately 0.70 and 0.85 respectively at 0 dB, DB-MS-CAN sustains a correlation above 0.92. This superior performance is attributed to the BFM, which adaptively assigns low attention weights to noise-dominated time steps, effectively "locking on" to effective signal phases and filtering out random interference.

4. Discussion

The primary bottleneck in microseismic monitoring for mine safety has consistently been the trade-off between inversion accuracy and computational efficiency.

Traditional methods represent two extremes: travel-time tomography (CT) is fast but lacks detail, while FWI is precise but computationally intensive. The DB-MS-CAN model proposed in this study effectively bridges this gap. While there are formal differences in comparing an offline-trained deep learning model with online iteratively optimized traditional methods, this comparison possesses substantial fairness and significant industrial value within the specific context of dynamic mine safety monitoring. Traditional methods, particularly FWI, treat each microseismic event as an independent problem and must solve the wave equation from scratch. While this strategy is reasonable in exploration seismology, it appears redundant in real-time early warning scenarios. By shifting the heavy computational burden to the offline training phase, DB-MS-CAN achieves an online inference speed of approximately 7 seconds per event. This near-real-time processing capability allows the stress field to be updated dynamically at a speed synchronous with event occurrence, enabling engineers to promptly identify hazardous precursors such as high-stress concentration zones. Furthermore, both CT and FWI used for comparison in this paper employ industry-leading optimization algorithms to ensure performance, which further highlights the generational efficiency advantage of the proposed model.

Moreover, compared with current state-of-the-art models, the model in this paper maintains an advantage. Ablation experiments confirm the strengths and roles of each module within the model: the MDFA module focuses on the spatial continuity of low-frequency geological structures, while the BFM module focuses on the transient scattering of high-frequency P/S waves.

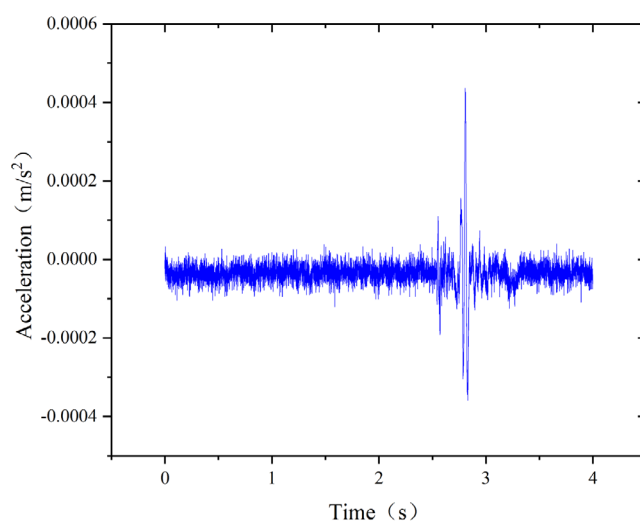


Figure 14. Raw microseismic signals featuring complex real-world noise

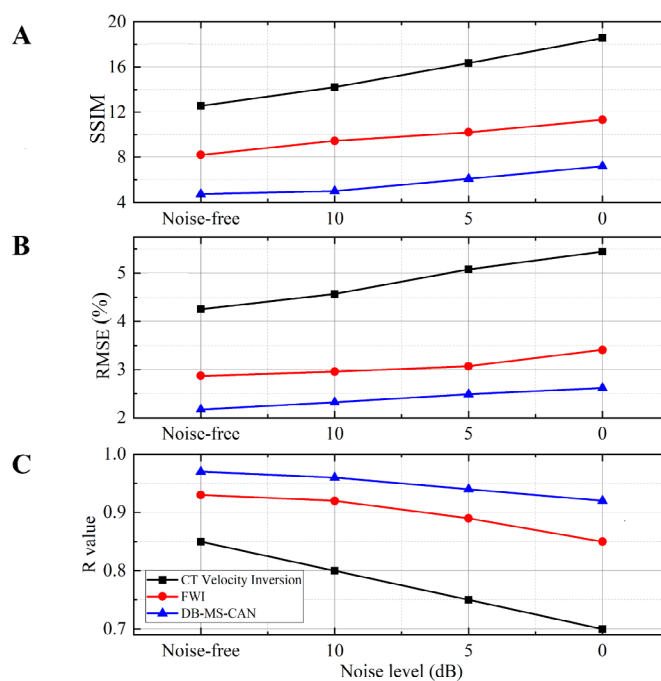


Figure 15. Evaluation metrics of the three methods after adding noise

Abbreviations: CT: Computed tomography; DB-MS-CAN: Dual-Branch Multi-Scale Convolutional Attention Network; FWI: Full waveform inversion; RMSE: Root mean square error; SSIM: Structural similarity index.

Although this study has achieved encouraging results in specific mining areas, to meet rigorous scientific standards, we must frankly discuss the limitations of the current method: synthetic vs. field data. Laboratory synthetic data are simpler than field data and struggle to restore the complex geological structures of mining

environments. This paper adopts a strategy of synthesizing ground truth based on the actual physical environment of the mining area to ensure that the validation of the model under complex working conditions has practical significance. Additionally, to ensure data accuracy, our field labels are not purely data-driven results but are

subjected to strict geological constraints from currently publicly available borehole logging and active source detection, followed by manual screening by experienced experts. This minimizes the non-uniqueness errors inherent in offline CT itself. Nonetheless, we acknowledge that systematic errors in offline inversion may still set an upper bound on model accuracy, potentially introducing risks of bias transfer and circular validation. To verify the reliability of the proposed model, we introduced a synthetic benchmark dataset generated via physical equation forward modeling. In this experiment, the ground truth is absolutely precise and independent of any inversion algorithm. The performance of DB-MS-CAN over PINN and Transformer on synthetic data proves that the network has indeed learned spatiotemporal patterns consistent with wavefield propagation physics, rather than merely fitting specific artifacts of offline tomography. In comparative experiments, we found that PINN possesses a natural physical advantage in background smoothness, while Transformer excels in global macro-positioning. Future work will explore unsupervised learning or reference the advantages of PINN and Transformer by using wave equation residuals directly as a loss function to fundamentally escape dependence on labeled data.

Current validation is based on data from a single mine. While the dataset provides comprehensive spatial coverage, its “universal applicability” should currently be limited to engineering environments with similar geological backgrounds. The main challenge facing cross-mine application is domain drift. The different dilation rate features captured by MDFA correspond respectively to fractures and structures of different scales in the geological body. This “scale-independent” feature extraction capability gives it potential structural robustness when processing heterogeneities of different scales across mines. Future work for this model will focus on physics-driven transfer learning, utilizing the architecture’s ability to capture physical laws to achieve small-sample fine-tuning for new scenarios.

The current model assumes a relatively fixed sensor layout. However, in actual mining, rock movement may lead to sensor displacement or failure, thereby causing changes in input tensor patterns. It is crucial to highlight that the Channel Attention mechanism within the BFM module provides an inherent degree of topological robustness against sensor malfunctions. As demonstrated in our quantitative analysis (Figure 5C), the proposed mechanism autonomously learns to evaluate sensor reliability. For noise-dominated or faulty channels, such as Channel 2, the network assigns a significantly suppressed

weight (0.13), effectively filtering out misleading data. Conversely, the model prioritizes high-fidelity signals by assigning dominant weights to coherent channels like Channel 7 (0.78). This adaptive weighting capability ensures that the inversion process remains stable even under partial sensor failure, significantly enhancing the practical utility of DB-MS-CAN for continuous safety-critical monitoring in harsh industrial settings.

5. Conclusion

This study proposes the DB-MS-CAN deep learning framework to address the difficulty of balancing inversion accuracy and computational efficiency in microseismic passive tomography. By integrating multi-scale feature aggregation and a dual-branch attention mechanism, this method demonstrates potential superior to traditional methods in the studied mining environment. The core findings and contributions of this research are summarized as follows: The proposed framework, through an offline learning strategy, reduces the inference time for a single event to approximately 7 seconds while maintaining structural fidelity (SSIM 0.925) comparable to high-precision physical inversion, providing a feasible technical path for real-time dynamic disaster warning. The temporal decoupling mechanism of the BFM module endows the model with stability in strong noise environments. Even under stress tests with a 0 dB signal-to-noise ratio, the model maintains a high correlation coefficient, indicating its ability to adapt to complex industrial site noise environments. Compared to general computer vision models, the network architecture integrating the MDFA and BFM modules shows significant advantages in resolving micro-scale fractures and macro-scale fault zones.

Acknowledgments

None.

Funding

This work was supported by the Open Fund of State Key Laboratory of Geomechanics and Geotechnical Engineering (Grant No. SKLGGES-024034), the Key Research and Development Program of Jining City (Grant No. 2024AQGX009), and the Research Planning Project of Qilu Institute of Technology (Grant Nos. QIT24BZ002 and QIT25TP006).

Conflict of interest

The authors declare they have no competing interests.

Author contributions

Conceptualization: Aoyu Feng, Binxin Hu, Feng Zhu

Data curation: Aoyu Feng, Feifei Jiang

Formal analysis: Aoyu Feng, Danya Ma

Funding acquisition: Feng Zhu

Investigation: Aoyu Feng, Feifei Jiang

Methodology: Aoyu Feng, Binxin Hu, Yang Gao

Project administration: Binxin Hu

Resources: Yang Gao, Rong Liu

Software: Aoyu Feng

Supervision: Binxin Hu, Feng Zhu

Validation: Binxin Hu, Feng Zhu, Rong Liu

Visualization: Aoyu Feng, Danya Ma

Writing—original draft: Aoyu Feng

Writing—review & editing: Binxin Hu, Feng Zhu, Yang Gao

Availability of data

The authenticity of this study is grounded in a high-density microseismic dataset collected from an operational coal mine in Shanxi Province. As this is proprietary, sensitive monitoring data critical to mine safety and bound by non-disclosure agreements, the data cannot be shared publicly.

AI use statement

The authors affirm that AI tools were utilized solely for minor language refinement, grammar checking, and translation assistance during the preparation of this manuscript. All conceptualization, methodological design, experimental execution, result analysis, and interpretation were performed solely by the authors. The authors have reviewed and edited the content to ensure accuracy and original thought.

References

1. Feng F, Chen S, Zhao X, Li D, Wang X, Cui J. Effects of external dynamic disturbances and structural plane on rock fracturing around deep underground cavern. *Int J Coal Sci Technol*. 2022;9(1):15.
doi: 10.1007/s40789-022-00487-z
2. Tang Z, Liu X, Xu Q, Li C, Qin P. Stability evaluation of deep-buried TBM construction tunnel based on microseismic monitoring technology. *Tunnelling and Underground Space Technology*. 2018;81:512-524.
doi: 10.1016/j.tust.2018.08.02
3. Liu JP, Si YT, Wei DC, Shi HX, Wang R. Developments and prospects of microseismic monitoring technology in underground metal mines in China. *J Cent South Univ*. 2021;28(10):3074-3098.
doi: 10.1007/s11771-021-4839-y
4. Dai R, Wang Y, Zhang D, Ji H. Automatic P-phase-onset-time-picking method of microseismic monitoring signal of underground mine based on noise reduction and multiple detection indexes. *Entropy*. 2023;25(10):1451.
doi: 10.3390/e25101451
5. Aghamiry HS, Gholami A, Operto S, Malcolm A. ADMM-based full-waveform inversion for microseismic imaging. *Geophys J Int*. 2022;228(1):259-274.
doi: 10.1093/gji/ggab332
6. Zhang B, Xia D, Tan J, et al. Acoustic wave equation full waveform inversion based on optimal transport and L2 norm hybrid weighting. *J Ocean Univ China*. Published online March 18, 2025.
doi: 10.1007/s11802-026-6167-1
7. Zhang S, Li M, Cui G, Mao G. Jiyu xishu zhengzehua FWI he mOWL-QN de weidizhen zhenyuan weizhi yu zhenyuan zibo tongshi fanyan fangfa [A simultaneous inversion method for microseismic source location and source waveform based on sparse regularization FWI and mOWL-QN]. *Oil Geophysical Prospecting*. 2025;60(5):1146-1159. [In Chinese].
doi: 10.13810/j.cnki.issn.1000-7210.20240309
8. Sun P, Zheng J, Zhao J, Yang Y, Wang Y. Fully automatic differentiation with coupling deep neural networks for full-waveform inversion. *J Seismic Explor*. 2025;34(6):16-28.
doi: 10.36922/JSE025410085
9. Zhang J, Song G, Zhu F, Hu B, Zhang H. Elastic wave CT with multistage inversion process based on improved SIRT. *Phys Lett A*. 2024;522:129739.
doi: 10.1016/j.physleta.2024.129739
10. Cai S. Zhu beidong shuangyuan weizhen CT fanyan suanfa yanjiu ji yingyong [Research and application of the active-passive dual-source microseismic CT inversion algorithm] [master's thesis]. Xuzhou, China: China University of Mining and Technology; 2024. [In Chinese].
doi: 10.27623/d.cnki.gzkyu.2024.002106
11. Tian Z, Pei S, Yu Q. Inversion imaging research of cross hole seismic CT in complex Karst areas: Guiyang Metro Line 3 under construction presented as an example. *CT Theory and Applications*. 2024;33(5):561-567. [In Chinese].
doi: 10.15953/j.ctta.2023.132
12. Yuan C, Zhang J. Vertical slowness-constrained joint anisotropic parameters and event-location inversion for downhole microseismic monitoring. *Remote Sensing*. 2025;17(3):529.
doi: 10.3390/rs17030529
13. Olya BAM, Mohebian R, Moradzadeh A. A new approach for seismic inversion with GAN algorithm. *J Seismic Explor*.

- 2024;33(3):1-36.
14. Zhou Y, Han L, Zhang P, Zeng J, Shang X, Huang W. Microseismic data-direct velocity modeling method based on a modified attention U-Net architecture. *Applied Sciences*. 2023;13(20):11166.
doi: 10.3390/app132011166
 15. Li L, Zeng X, Pan X, Peng L, Tan Y, Liu J. Microseismic velocity inversion based on deep learning and data augmentation. *Applied Sciences*. 2024;14(5):2194.
doi: 10.3390/app14052194
 16. Chang A, Gross L, Hörning S. Full waveform inversion constrained to well log velocity data and seismic data using random mixing. *J Appl Geophys*. 2025;243:105968.
doi: 10.1016/j.jappgeo.2025.105968
 17. Gou R, Song D, He X, *et al.* Microseismic response characteristics and stress anomaly zoning in deep outburst-prone coal seams with multi-concealed structures: A case study. *J Appl Geophys*. 2025;242:105932.
doi: 10.1016/j.jappgeo.2025.105932
 18. Ma XP. Dadi dianci he yuanzhen zoushi shuju sanwei lianhe fanyan yanjiu [Research on three-dimensional joint inversion of magnetotelluric and teleseismic arrival time data] [PhD dissertation]. Changchun, China: Jilin University; 2025. [In Chinese].
doi: 10.27162/d.cnki.gjlin.2025.000803
 19. Wamriew D, Charara M, Pissarenko D. Joint event location and velocity model update in real-time for downhole microseismic monitoring: A deep learning approach. *Comput Geosci*. 2022;158:104965.
doi: 10.1016/j.cageo.2021.104965
 20. Cao D, Su Y, Cui R. Multi-parameter pre-stack seismic inversion based on deep learning with sparse reflection coefficient constraints. *J Pet Sci Eng*. 2022;209:109836.
doi: 10.1016/j.petrol.2021.109836
 21. Song C, Lu M, Li Y, *et al.* Spatial pattern learning: dip structure constraint multi-view convolutional neural network for pre-stacked seismic inversion. *IEEE Trans Geosci Remote Sens*. 2023;61:1-16.
doi: 10.1109/TGRS.2023.3307897
 22. Zou L, Jia J, Ye Y, Wu B. STA-Fault3D: a lightweight 3D seismic fault detection network based on spatial-temporal asymmetric convolution set. *Applied Sciences*. 2025;15(22):12153.
doi: 10.3390/app152212153
 23. Ou L, Huang L, Zhao Y, Wang Z, Shen H, Li X. Intelligent phase picking of microseismic signals based on ResUNet in underground engineering. *J Cent South Univ*. 2025;32(9):3314-3335. [In Chinese].
doi: 10.1007/s11771-025-6077-1
 24. Ma B, Han L, Liu W, Wu Z, Li C. A deep learning velocity modeling method based on a novel attention mechanism networks. *J Seismic Explor*. 2024;33(5):1-14.
 25. Yang R. Convolutional neural network for image classification research - based on VGG16. In: *Proceedings of the 2024 International Conference on Image Processing, Computer Vision and Machine Learning (ICICML)*; November 22-24, 2024; Shenzhen, China. IEEE; 2024:213-217.
doi: 10.1109/ICICML63543.2024.10958042
 26. Li S, Xie B. Boosting UNet performance via VGG-based encoder for medical image segmentation. In: *Proceedings of the 2025 6th International Conference on Computer Engineering and Application (ICCEA)*. April 25-27, 2025; Hangzhou, China. IEEE; 2025:1964-1968.
doi: 10.1109/ICCEA65460.2025.11102997
 27. Tang W, Sun J, Wang S, Zhang Y. Review of AlexNet for Medical Image Classification. *EAI Endorsed Trans e-Learn*. 2023;9.
doi: 10.4108/eetel.4389
 28. Wei K, Liu H, Cao H, Chen M. Deep learning inversion research on frequency-domain borehole-to-surface electromagnetic method based on CNN-LSTM network. *Comput Tech Geophys Geochem Explor*. 2025. [In Chinese].
doi: 10.12474/wthtjs.20250226-0001
 29. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: *Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017)*. July 21-26, 2017; Honolulu, Hawaii, USA. IEEE; 2017:2261-2269.
doi: 10.1109/CVPR.2017.243
 30. Liu L, Song W, Zeng C, Yang X. Microseismic event detection and classification based on convolutional neural network. *J Appl Geophys*. 2021;192:104380.
doi: 10.1016/j.jappgeo.2021.104380
 31. Libby SA. Modelling faults as fracture swarms to introduce multi-scale heterogeneity. In: *Proceedings of the 58th U.S. Rock Mechanics/Geomechanics Symposium*. June 23-26, 2024; Golden, Colorado, USA. American Rock Mechanics Association; 2024.
doi: 10.56952/ARMA-2024-0934
 32. Chen F, Liang Z, Cao A. ConvLSTM for predicting short-term spatiotemporal distribution of seismic risk induced by large-scale coal mining. *Nat Resour Res*. 2023;32(3):1459-1479.
doi: 10.1007/s11053-023-10193-5
 33. Luo H, Xu X, Pan Y, Yu J, Zhang Y, Zhang L. The CGAS

- deep learning algorithm for P-wave arrival time picking of mining microseismic events. *IEEE Access*. 2023;11:102961-102970.
doi: 10.1109/ACCESS.2023.3317084
34. He Z, Jia M, Wang L. UACNet: a universal automatic classification network for microseismic signals regardless of waveform size and sampling rate. *Eng Appl Artif Intell*. 2023;126:107088.
doi: 10.1016/j.engappai.2023.107088
 35. Xu H, Li K, Shan P, *et al.* Automatic picking method for the first arrival time of microseismic signals based on fractal theory and feature fusion. *Fractal and Fractional*. 2025;9(11):679.
doi: 10.3390/fractalfract9110679
 36. Yang Y, Wang H, Li Y, Birnie CE, Alkhalifah T. Automatic microseismic event location using deep neural networks in anisotropic media. In: *Proceedings of the 83rd EAGE Annual Conference & Exhibition*. June 6-9, 2022; Madrid, Spain. European Association of Geoscientists & Engineers; 2022:1-5.
doi: 10.3997/2214-4609.202210621
 37. Aguiar AC, Myers SC. Microseismic event relocation based on PageRank linkage at the Newberry Volcano geothermal site. *Bull Seismol Soc Am*. 2018;108(6):3656-3667.
doi: 10.1785/0120180115
 38. Antolik M, Ekstrom G, Dziewonski AM, Boschi L, Kustowski B, Pan J. Multi-resolution global models for teleseismic and regional event location. In: *Proceedings of the 23rd Seismic Research Review: Worldwide Monitoring of Nuclear Explosions*. October 2-5, 2001; Jackson Hole, Wyoming, USA. National Nuclear Security Administration; 2001:136-146.
 39. Yang L. Theoretical analysis of Adam optimizer in the presence of gradient skewness. *Int J Appl Sci*. 2024;7(2):27-42.
doi: 10.30560/ijas.v7n2p27
 40. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM*. 2017;60(6):84-90.
doi: 10.1145/3065386
 41. Bazulin M, Sabitov D, Charara M. Determination of the elastic parameters of a VTI medium from sonic logging data using deep learning. *Comput Geosci*. 2021;152:104759.
doi: 10.1016/j.cageo.2021.104759
 42. Li W, Liu H, Wu T, Huo S. A high-resolution velocity inversion method based on attention convolutional neural network. *IEEE Trans Geosci Remote Sens*. 2023;61:1-14.
doi: 10.1109/TGRS.2023.3311788
 43. Cao W, Guo XB, Tian F, *et al.* Seismic velocity inversion based on CNN-LSTM fusion deep neural network. *Appl Geophys*. 2021;18(4):499-514.
doi: 10.1007/s11770-021-0913-3
 44. Sun QH, Zong ZY, Li X. Probabilistic seismic inversion based on physics-guided deep mixture density network. *Pet Sci*. 2024;21(3):1611-1631.
doi: 10.1016/j.petsci.2023.12.015
 45. Lu J, Wu C, Huang J, Li G, Yuan S. Multi-receptive field distillation network for seismic velocity model building. *Eng Appl Artif Intell*. 2024;133:108547.
doi: 10.1016/j.engappai.2024.108547