

ARTICLE

A deep hybrid network for frequency-aware seismic impedance inversion

Yang Zhang^{1*}, Hong Cao^{2*}, Zhifang Yang¹, Hao Yang¹, Qiang Ge¹, Yuqi Qiu¹, Jiangbei Huang³, and Sen Zhao¹

¹Research Institute of Petroleum Exploration and Development, Beijing, China

²Bureau of Geophysical Prospecting, Zhuozhou, Hebei, China

³PetroChina Hangzhou Research Institute of Geology, Hangzhou, Zhejiang, China

Abstract

Seismic impedance inversion is a fundamental task in exploration geophysics. With the development of artificial intelligence, deep learning has been increasingly adopted for post-stack seismic impedance inversion. This paper proposes a frequency-aware deep hybrid framework that integrates continuous wavelet transform (CWT) representations, a transformer encoder, and atrous spatial pyramid pooling (ASPP) to improve both impedance fidelity and resolution. The CWT provides time–frequency components of seismic traces; the transformer aggregates non-local context to stabilize structural consistency; and ASPP captures local multi-scale features. In addition, we introduce a frequency-aware loss that rebalances optimization toward components that are more difficult to reconstruct, improving detail recovery in thin beds. Network training used a semi-supervised strategy. At well locations, well-log impedance values supervised the network prediction; at non-well locations, synthetic seismic data were generated by seismic forward modeling and compared with the observed seismic records. Validation on the Marmousi II model and a field dataset indicated that the proposed method achieves consistently higher accuracy than representative deep-learning baselines and a commercial post-stack inversion workflow.

Keywords: Seismic inversion; Frequency-aware loss; Deep learning

*Corresponding authors:

Yang Zhang
(zy_1211_zy@163.com);
Hong Cao
(caoho@petrochina.com.cn)

Citation: Zhang Y, Cao H, Yang Z, *et al.* A deep hybrid network for frequency-aware seismic impedance inversion. *J Seismic Explor.*
doi: 10.36922/JSE026270126

Received: July 3, 2026

Revised: August 10, 2026

Accepted: August 14, 2026

Published online: August 26, 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.

1. Introduction

Seismic inversion is a crucial task in the field of geophysical exploration. It provides an effective means to quantitatively estimate subsurface elastic and petrophysical properties from seismic observations. However, seismic inversion problems are inherently ill-posed and non-unique due to the band-limited nature of seismic data, limited acquisition geometry, and the inevitable presence of noise. These factors lead to instability in the inversion process and make it difficult to obtain reliable high-resolution subsurface parameter models. With the development of geophysical technology, various seismic inversion methods have been proposed and applied since the 1970s, such as least squares-based inversion methods,¹ model-based inversion methods,²⁻⁴ sparse-spike inversion methods,⁵⁻⁸ and full-waveform inversion (FWI).⁹⁻¹¹ Least squares-based inversion methods aim to minimize the difference between observed and

modeled seismic responses by formulating the problem as an optimization task. They typically use smoothness constraints or Tikhonov-type regularization to stabilize the solution and reduce the impact of noise and band-limited data. However, such methods often produce overly smooth results and may fail to recover sharp interfaces in complex geological settings. Model-based inversion methods further extend this framework by incorporating prior geological or petrophysical information to guide the inversion process. This incorporation of prior knowledge improves stability and robustness, especially in noisy environments, but also introduces dependence on the quality of the prior models. On the other hand, sparse-spike inversion methods assume that the subsurface reflectivity series is sparse, and therefore formulate inversion as a sparse optimization problem. These methods commonly employ L1-norm regularization or basis pursuit techniques to promote sparsity, which helps enhance temporal resolution and better preserve discontinuities. Among more advanced techniques, FWI has attracted significant attention due to its ability to use full-wavefield information and adjoint-state methods for gradient computation. Although FWI can theoretically achieve very high-resolution subsurface images, it is computationally expensive and highly nonlinear. These factors limit its robustness and practical applicability in complex real-world exploration scenarios.

For seismic impedance inversion, Das *et al.*¹² used a one-dimensional (1D) convolutional neural network (CNN) to estimate seismic impedance from seismic data. Wu *et al.*¹³ further improved upon this by replacing the 1D CNNs with a two-dimensional CNN and incorporating an initial impedance model as a constraint. The initial impedance model was used as an additional network input to capture low-frequency trends, which aided in generating stable impedance predictions. Biswas *et al.*¹⁴ discussed model-based and physics-based methods for post-stack inversion and pre-stack inversion problems. In the physics-based approach, the estimated elastic parameters are used to generate synthetic seismic data. These data are then compared with the field seismic data to calculate errors and update the weights. The results show that the physics-based approach significantly reduces the network's requirement for a realistic model. Zhang *et al.*¹⁵ introduced transfer learning to address the issue of limited labels. They used a priori low-frequency impedance model and seismic data as inputs to the network. The results indicate that transfer learning helps alleviate the network's dependence on large amounts of labeled field data, thereby enhancing the method's generalizability. Wu *et al.*¹⁶ proposed a supervised workflow for seismic impedance inversion based on generative adversarial networks, which consists of three networks: a generator, a discriminator,

and a forward model. The results show that the method makes better use of both labeled and unlabeled data, and predicts impedance more accurately than other inversion methods.

Most of these methods, along with some current inversion methods,^{17–29} rarely consider the time–frequency components, which is crucial for seismic inversion. Additionally, the performance of CNNs is limited by the size of their receptive fields, restricting their ability to capture global information. On the other hand, the continuous wavelet transform (CWT) can decompose seismic data into time–frequency components;^{30–32} different dilation rates in atrous spatial pyramid pooling (ASPP) expand the network's receptive field,³³ and transformers, equipped with multi-head attention mechanisms, can simultaneously learn information from different frequency bands while also demonstrating excellent capabilities in capturing global information.^{34,35}

This study accordingly proposes a frequency-aware seismic impedance inversion framework based on a deep hybrid network. The ASPP branch processes raw seismic data and low-frequency initial impedance through dilated convolutions, enhancing fine-grained detail extraction by capturing localized features across varying spatial resolutions. Concurrently, the transformer encoder branch analyzes CWT-derived time–frequency components, leveraging multi-head attention mechanisms to discern frequency-dependent seismic features while establishing long-range contextual dependencies. Feature representations from the two branches are adaptively integrated through a gating network, enabling the fusion of multiscale representations and enhancing overall representational capacity. To address limited labeled data availability, we implement a semi-supervised learning strategy. The supervised part incorporates mean squared error (MSE) between predicted and true impedance along with a spectral compensation mechanism that reweights frequency-domain discrepancies to prioritize underrepresented spectral components,^{36,37} thereby balancing local and global feature learning. The weakly supervised part employs a convolutional forward operator to generate synthetic seismic data from predicted impedance, optimized through correlation analysis with observed seismic data. Compared with traditional deep learning methods, the proposed method provides a reliable approach to seismic impedance estimation.

2. Methods

2.1. Seismic forward modeling

Seismic impedance inversion aims to estimate subsurface impedance models from seismic observations. The

relationship between impedance and seismic response can be described through the seismic forward modeling process. In this study, the acoustic impedance is defined as:

$$I = \rho V_p \quad (1)$$

where I denotes acoustic impedance, ρ represents density, V_p and is the P-wave velocity.

The seismic reflection coefficient is determined by the impedance contrast between adjacent layers:

$$r(n) = \frac{I(n+1) - I(n)}{I(n+1) + I(n)} \quad (2)$$

where $r(n)$ represents the reflection coefficient at the n -th sampling point. The seismic trace can be simulated by convolving the reflection coefficient series with the seismic wavelet:

$$s(n) = w(n) \times r(n) \quad (3)$$

where $s(n)$ denotes the synthetic seismic data, $w(n)$ represents the seismic wavelet, and $r(n)$ denotes the convolution operation.

According to the above forward model, seismic impedance inversion can be regarded as an inverse problem of recovering the impedance model from seismic responses. In other words, the seismic data contain indirect information about subsurface impedance through the reflection coefficients generated by impedance contrasts.

2.2. Network architecture

The proposed network architecture (as shown in Figure 1) employs parallel transformer encoder and ASPP branches to extract complementary information, adaptively fused through a gating block to generate impedance predictions.

In the transformer encoder block, we first process the raw seismic data using CWT to obtain seismic data in the time–frequency domain (as shown in Figure 2). The transformer encoder receives a multichannel time–frequency representation of the seismic data, with channels corresponding to different frequency components. The equation for performing CWT on a time series $x(t)$ is shown below:

$$T(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \psi^* \left(\frac{t-b}{a} \right) dt \quad (4)$$

$$\psi(t) = \pi^{-1/4} e^{im_0 t} e^{-t^2/2} \quad (5)$$

where $\psi^*((t-b)/a)$ is the conjugate of the mother wavelet function $\psi(t)$ at scale a ($a > 0$) and time shift b , $T(a, b)$ represents the time–frequency components of the time series $x(t)$ after CWT, and ω_0 represents the central

frequency of the mother wavelet. Additionally, the real part of $T(a, b)$ represents amplitude information, while the imaginary part represents phase information. Therefore, we use the real part of the data to represent the amplitude at a specific time and frequency.

Spectral decomposition of the raw seismic data using various scales of mother wavelet functions facilitates the generation of high-resolution time–scale maps. Specifically, we used the Morlet mother wavelet for CWT of the time-domain seismic sequences. The Morlet wavelet was selected because it is a widely used complex wavelet for continuous wavelet transform analysis and is suitable for representing oscillatory and non-stationary signals. To express the scale axis in frequency units, the wavelet scales in the time–scale map were converted to their corresponding frequencies, yielding the time–frequency components. Figure 2A shows a time-domain seismic data sequence, while Figure 2B shows the time–frequency components obtained after CWT. The dominant frequency of the seismic data used in this study was approximately 30–40 Hz. The effective seismic energy is mainly concentrated below approximately 80 Hz, whereas the spectral amplitude above 80 Hz decreases significantly and contains little reliable reflection information. Therefore, the maximum frequency used in the CWT representation was set to 80 Hz.

The resulting time–frequency representation comprised 80 frequency channels derived from the original seismic data. This tensor serves as input to the transformer encoder (as shown in Figure 1B), where an initial convolutional layer (followed by batch normalization and rectified linear unit [ReLU] activation) projects frequency dimensions into the target feature space d_{model} . Positional encoding is then applied according to:

$$PE_{(pos, 2m)} = \sin(pos / 10000^{2m/d_{model}}) \quad (6)$$

$$PE_{(pos, 2m+1)} = \cos(pos / 10000^{2m/d_{model}}) \quad (7)$$

where pos denotes the position index, m denotes the feature dimension, and PE denotes the positional-encoding matrix. The positional encoding is fixed and is not updated during training. The output layer is designed to constrain the numerical range between 0 and 1, so the activation function is changed to sigmoid compared to the input layer.

In addition to the input layer and the output layer, the transformer encoder block mainly includes P sets of multi-head self-attention layers and feed-forward layers. A multi-head self-attention layer consists of multiple independent self-attention heads operating in parallel.

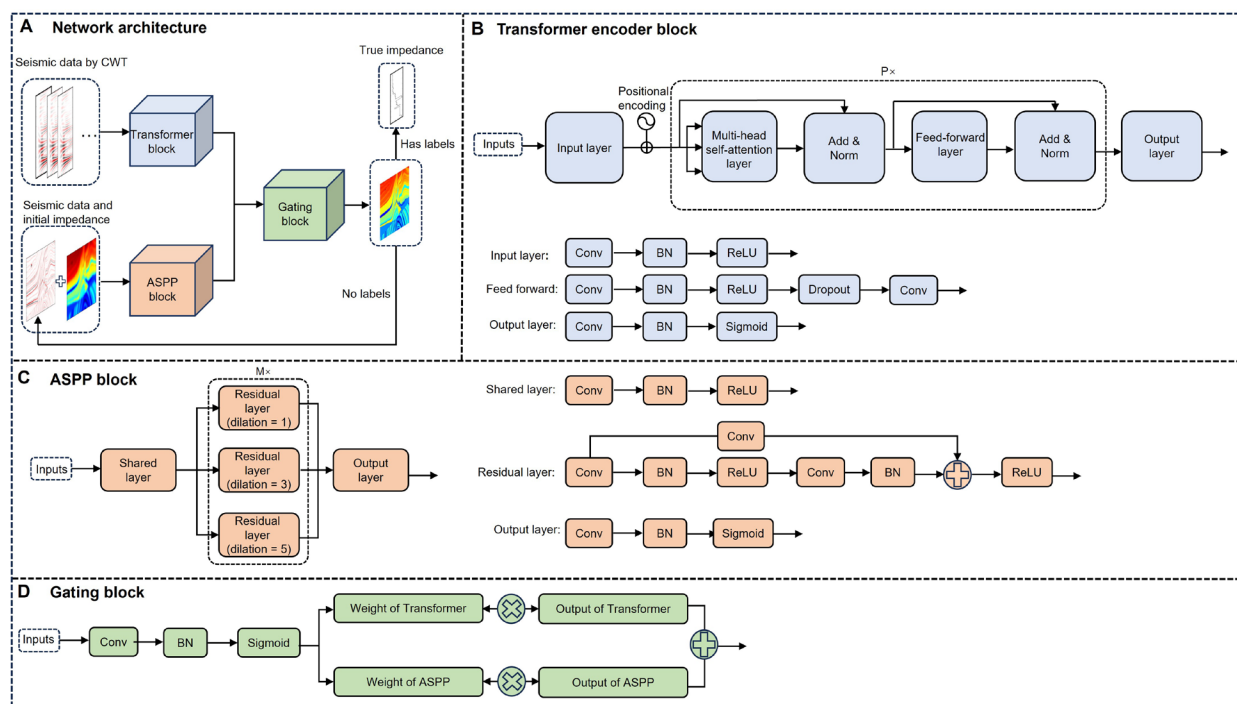


Figure 1. The framework of the proposed method for seismic impedance inversion. (A) Overall network architecture; (B) transformer encoder block; (C) ASPP block; and (D) gating block.

Abbreviations: ASPP: Atrous spatial pyramid pooling; BN: Batch normalization; Conv: Convolution; CWT: Continuous wavelet transform; ReLU: Rectified linear unit.

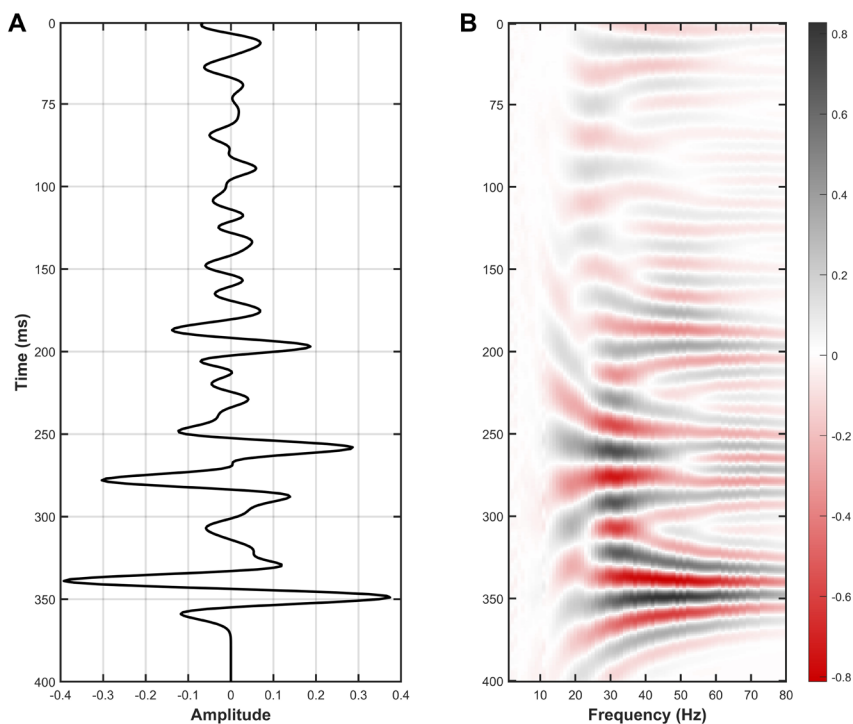


Figure 2. Continuous wavelet transform (CWT) of the seismic data. (A) Raw seismic data; (B) time–frequency representation obtained using CWT.

Each head computes attention values in different subspaces to capture distinct features. The specific computation process of multi-head self-attention is as follows:

$$\text{Multihead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h) W^O \quad (8)$$

$$\text{head}_i = \text{softmax}\left(\frac{Q_i K_i^T}{\sqrt{d_k}}\right) V_i \quad (9)$$

$$Q_i = S W_i^Q, K_i = S W_i^K, V_i = S W_i^V \quad (10)$$

where head_h denotes h heads, head_i denotes the i -th head, $W^O \in \mathbb{R}^{d_{\text{model}} \times d_{\text{model}}}$ denotes the weight matrix, Q_i denotes the i -th query matrix, K_i denotes the i -th key matrix, V_i denotes the i -th value matrix, K_i^T denotes the transpose of K_i , $\sqrt{d_k}$ denotes the scaling factor ($d_k = d_{\text{model}} / h$), $S \in \mathbb{R}^{L \times d_{\text{model}}}$ denotes the input sequence with length L , and $W_i \in \mathbb{R}^{d_{\text{model}} \times d_k}$ denotes the weight matrix. The feed-forward layer consists sequentially of a convolutional layer, a batch normalization layer, a ReLU layer, a dropout layer, and another convolutional layer, which is used to enhance the network's nonlinear representation capability. Additionally, each layer is preceded by layer normalization, and residual connections are applied to both the input and output of each layer. Through the above-mentioned layers, the model can better learn global information and information across different frequency bands within time-frequency components.

The ASPP block is primarily composed of a shared convolutional layer, M groups of residual layers with three different dilation rates, and an output layer (see Figure 1C). The dilation rate represents the spacing between points in the convolutional kernel. By employing dilated convolutions, the network's receptive field can be expanded, allowing it to capture a broader range of multi-scale information and enhance its ability to represent local multi-scale features. The input to the ASPP block is several time-domain seismic data.

Outputs from both branches are integrated through a gating block (as shown in Figure 1D). This block learns branch-specific attention weights, performs weighted fusion, and sums contributions to produce the final impedance prediction, dynamically balancing global and local feature representations.

2.3. Loss function and training implementation

After obtaining the predicted impedance, we calculate the loss and use backpropagation to compute gradients and update the network parameters. This paper employs a semi-supervised strategy to mitigate the issue of insufficient labeled data. In the supervised learning part

(i.e., at the well points), we obtain L_1 by calculating the MSE between the predicted impedance and the ground-truth impedance, as follows:

$$L_1 = \frac{1}{N} \sum_{n=1}^N (I_{\text{true}}(n) - I_{\text{pred}}(n))^2 \quad (11)$$

where N denotes the length of labeled data, I_{true} denotes the true impedance, and I_{pred} denotes the predicted impedance by the network.

In addition, this paper introduces a loss function for calculating errors in the frequency domain, as follows:

$$L_2 = \frac{1}{N} \sum_{k=1}^N w(k) (FI_{\text{true}}(k) - FI_{\text{pred}}(k))^2 \quad (12)$$

$$w(k) = \frac{|FI_{\text{true}}(k) - FI_{\text{pred}}(k)|}{\max_k (|FI_{\text{true}}(k) - FI_{\text{pred}}(k)|)} \quad (13)$$

$$FI(k) = \sum_{n=1}^N I(n) \left(\cos\left(\frac{2\pi kn}{N}\right) - i \cdot \sin\left(\frac{2\pi kn}{N}\right) \right) \quad (14)$$

where FI_{true} denotes the discrete Fourier transform of I_{true} , FI_{pred} denotes the discrete Fourier transform of I_{pred} , $w(k)$ denotes the weight for frequency loss ($w(k) \in [0, 1]$). When $w(k) = 1$, it means the most difficult frequency to recover and the maximum frequency loss. High-frequency components are often difficult to recover in seismic inversion. Here, by introducing $w(k)$ and calculating its value, we can focus the model on high-frequency information, thereby better recovering high-frequency features. It is worth mentioning that n represents the time sampling index in the original impedance sequence, and k denotes the frequency index after the discrete Fourier transform.

In the unsupervised learning part (i.e., at the non-well points), we first use the predicted impedance to calculate the synthetic seismic data based on the seismic forward modeling described above, and then calculate the correlation between the synthetic seismic data and the field seismic data, as follows:

$$\text{PCC} = \frac{\sum_{n=1}^N (s_{\text{true}}(n) - \bar{s}_{\text{true}})(s_{\text{pred}}(n) - \bar{s}_{\text{pred}})}{\sqrt{\sum_{n=1}^N (s_{\text{true}}(n) - \bar{s}_{\text{true}})^2} \sqrt{\sum_{n=1}^N (s_{\text{pred}}(n) - \bar{s}_{\text{pred}})^2}} \quad (15)$$

$$L_3 = \frac{1}{2} (1 - \text{PCC}) \quad (16)$$

where s_{pred} and $\bar{s}_{\text{pred}}$ respectively denote the predicted synthetic seismic data and its mean, s_{true} and $\bar{s}_{\text{true}}$ respectively denote the true seismic data and its mean, and PCC denotes the Pearson correlation coefficient.

After obtaining the above losses, we compute the final loss by performing a weighted summation, which is then used for backpropagation to update the gradients of the network, as follows:

$$L = \alpha_1 L_1 + \alpha_2 L_2 + \alpha_3 L_3 \quad (17)$$

$$\alpha_1 + \alpha_2 + \alpha_3 = 1 \quad (18)$$

where α_1 , α_2 , and α_3 respectively denote the weights of the corresponding loss terms. Based on hyperparameter tuning, the loss weights were set to $\alpha_1 = 0.5$, $\alpha_2 = 0.3$, $\alpha_3 = 0.2$. During the training process, the seismic data at well locations and non-well locations were concatenated and jointly used as the input of each training batch. Therefore, the well-location data required for calculating L_1 and L_2 were included in every training iteration. In addition, Kaiming initialization was used to initialize the network parameters.³⁸ In the transformer encoder block, d_{model} was set to 256, P was set to 3, and h was set to 8. In the ASPP block, M was set to 3. The initial learning rate of the network was set to 0.001, the optimizer was selected as Adam, and cosine annealing was used to gradually reduce the learning rate. The batch size was set to 8, the total number of training iterations to 1,000, and the weight decay to 1×10^{-6} . The proposed network was trained using an Nvidia GeForce RTX 4060 Ti graphics processor unit

with a memory capacity of 16 GB.

3. Results

3.1. Synthetic example test

We use the Marmousi II model, with a size of 460 time samples $\times$ 2,721 CDPs and a sampling interval of 1 ms, as a synthetic test case to evaluate the performance of our proposed method. Figure 3 shows the synthetic seismic data, true impedance, initial impedance, and wavelet used. Synthetic seismic data were generated by convolving a 40 Hz Ricker wavelet with a reflectivity model. The initial impedance is obtained by applying Gaussian low-pass filtering to the true impedance.

In this test, we select three well locations at common depth points (CDPs) 600, 1,350, and 2,100. The wells at CDPs 1,350 and 2,100 were used as labeled training wells, whereas the well at CDP 600 was withheld as the blind test well. To verify the effectiveness of the proposed method, we introduce five baseline methods for ablation experiments. Figure 4 shows the impedance predicted at the blind well by all methods. Among them, Figure 4A shows the predicted impedance of the proposed method, Figure 4B shows the predicted impedance without the frequency-aware components (i.e., without L_2 and CWT, the transformer encoder received the same input as the

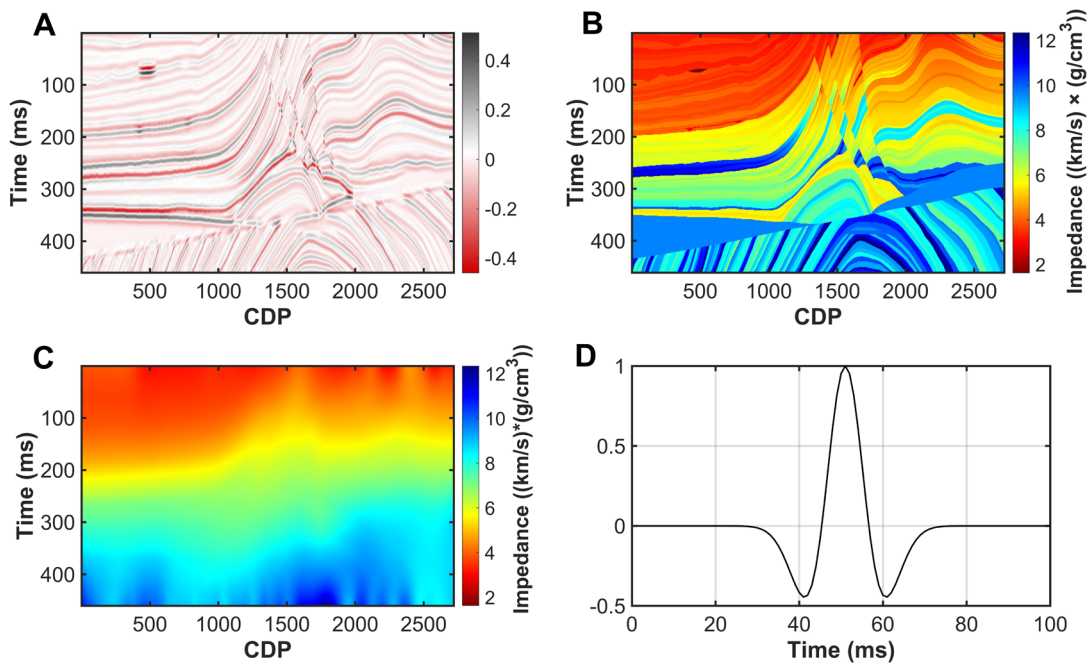


Figure 3. The synthetic seismic data, true impedance, initial impedance, and the wavelet used. (A) Synthetic seismic data; (B) true impedance; (C) initial impedance; and (D) wavelet.

Abbreviation: CDP: Common depth point.

ASPP branch), Figure 4C shows the predicted impedance using only the transformer encoder block, Figure 4D shows the predicted impedance using only the ASPP block, Figure 4E shows the predicted impedance using CNN, and Figure 4F shows the predicted impedance using the model-based inversion implemented in Hampson–Russell. In addition, we calculate the MSE and Pearson correlation coefficient (PCC) between the predicted impedance and the true impedance, and the results are shown in Table 1, with the best-performing values indicated. The equations for calculating MSE and PCC are presented in Equations 11 and 15, respectively. Figure 5 shows the impedance profile predicted by all six methods. Both the highlighted region (white circle) and the overall profile indicate that the proposed method provides higher apparent resolution and more clearly characterizes small-scale structures.

Table 1. MSE and PCC for different methods on the synthetic dataset

Method	MSE	PCC
Proposed	0.1394 ^a	0.9856 ^a
Without frequency-aware components	0.2726	0.9713
Transformer	0.3832	0.9626
ASPP	0.3270	0.9614
CNN	0.4614	0.9585
Model-based	0.5214	0.9348

Note: ^aBest-performing value.

Abbreviations: ASPP: Atrous spatial pyramid pooling; CNN: Convolutional neural network; MSE: Mean squared error; PCC: Pearson correlation coefficient.

3.2. Field data test

We applied the proposed method to a survey area in southwestern China, as shown in Figure 6A. The dataset comprised 900 CDPs over a 140-ms time window, with a sampling interval of 1 ms. There are three wells located at CDPs 106 (W1), 456 (W2), and 864 (W3). In addition, W3 was withheld as the blind test well. The initial impedance model was constructed by low-pass filtering the training-well impedance logs and interpolating the resulting low-frequency trends along interpreted horizons, as shown in Figure 6B. The seismic wavelet was statistically estimated by Hampson–Russell software based on seismic data, as shown in Figure 6C. To evaluate the prediction accuracy at the blind well, Figure 7 shows the impedance predicted at the blind well by all methods. In addition, we calculate the MSE and PCC between the predicted impedance and the true impedance, and the results are shown in Table 2,

with the best-performing values indicated. The equations for calculating MSE and PCC are shown in Equations 11 and 15. Figure 8 shows the impedance predicted profiles by all methods. The regions highlighted by the two black circles, near CDPs 300 and 800, show that the proposed method better preserves local impedance details and spatial variations.

Table 2. MSE and PCC of different methods on the field dataset

Method	MSE	PCC
Proposed	0.0858 ^a	0.8164 ^a
Without frequency-aware components	0.1213	0.7655
Transformer	0.1425	0.6392
ASPP	0.1581	0.6370
CNN	0.1554	0.6125
Model-based	0.1802	0.6031

Note: ^aBest-performing value.

Abbreviations: ASPP: Atrous spatial pyramid pooling; CNN: Convolutional neural network; MSE: Mean squared error; PCC: Pearson correlation coefficient.

4. Discussion

Seismic inversion is important for imaging subsurface structures and characterizing lithology and reservoirs. However, band-limited seismic data and wavelet uncertainty make seismic inversion inherently uncertain. Deep-learning methods can learn nonlinear relationships between seismic data and impedance, thereby improving the accuracy of inversion.

However, most current impedance inversion methods pay little attention to the frequency and global characteristics of seismic data, which makes it difficult to achieve ideal results. To address these issues, this study proposes a frequency-aware joint learning framework for seismic impedance inversion. In addition, we quantified the contribution of the different components by evaluating changes in MSE and PCC.

For the synthetic dataset, the complete model achieves an MSE of 0.1394 and a PCC of 0.9856. When the frequency-aware mechanism is removed, the MSE increases to 0.2726, corresponding to an increase of 95.55%, while the PCC decreases from 0.9856 to 0.9713 (1.45% reduction). This demonstrates that the frequency-aware mechanism plays an important role in recovering high-frequency impedance variations and thin-bed details. When only

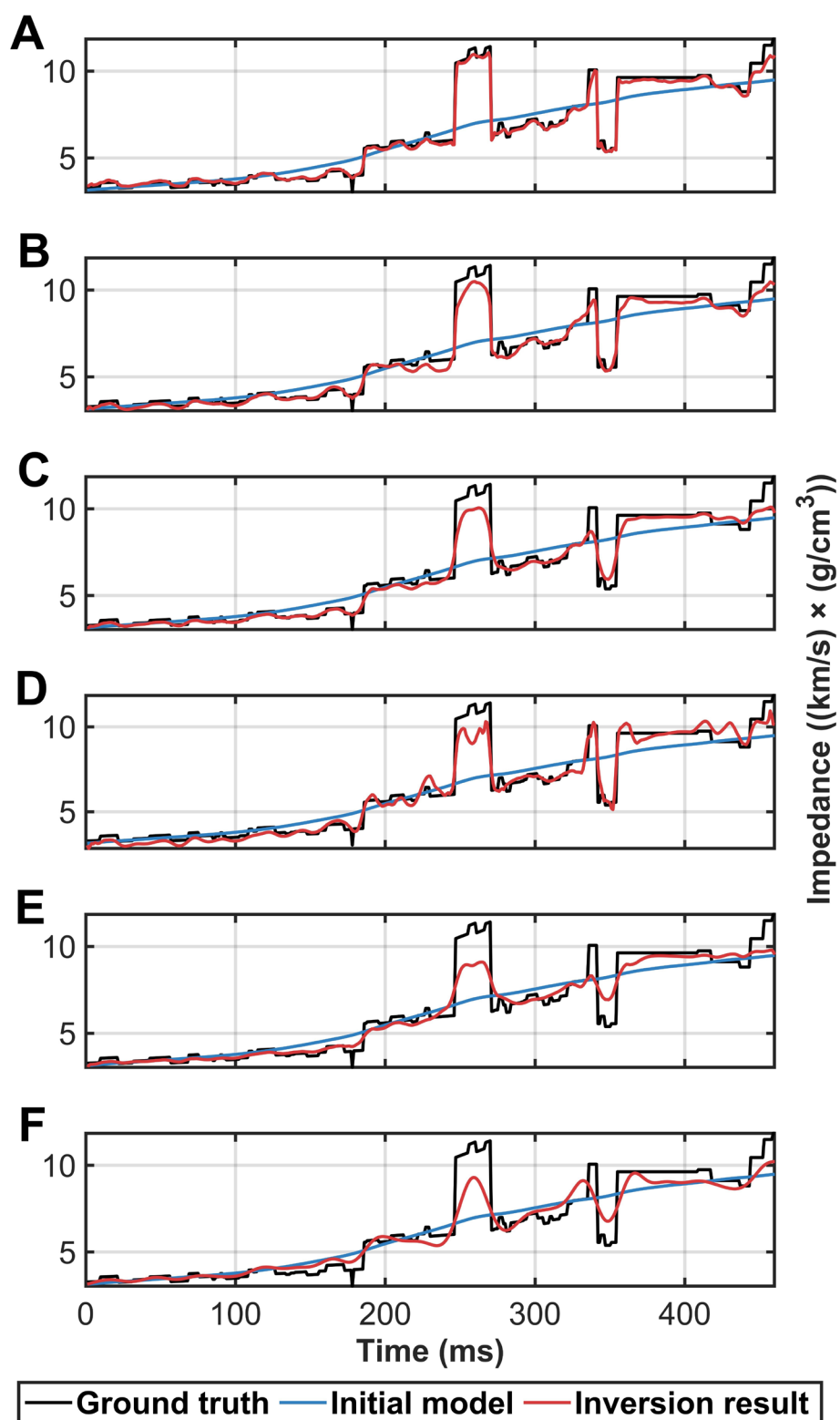


Figure 4. Blind-well impedance predictions obtained using six methods for the Marmousi II model (common depth point = 600). (A) Proposed methods; (B) ablation without the frequency-aware components; (C) transformer-only model; (D) atrous spatial pyramid pooling-only model; (E) convolutional neural network baseline; and (F) model-based inversion.

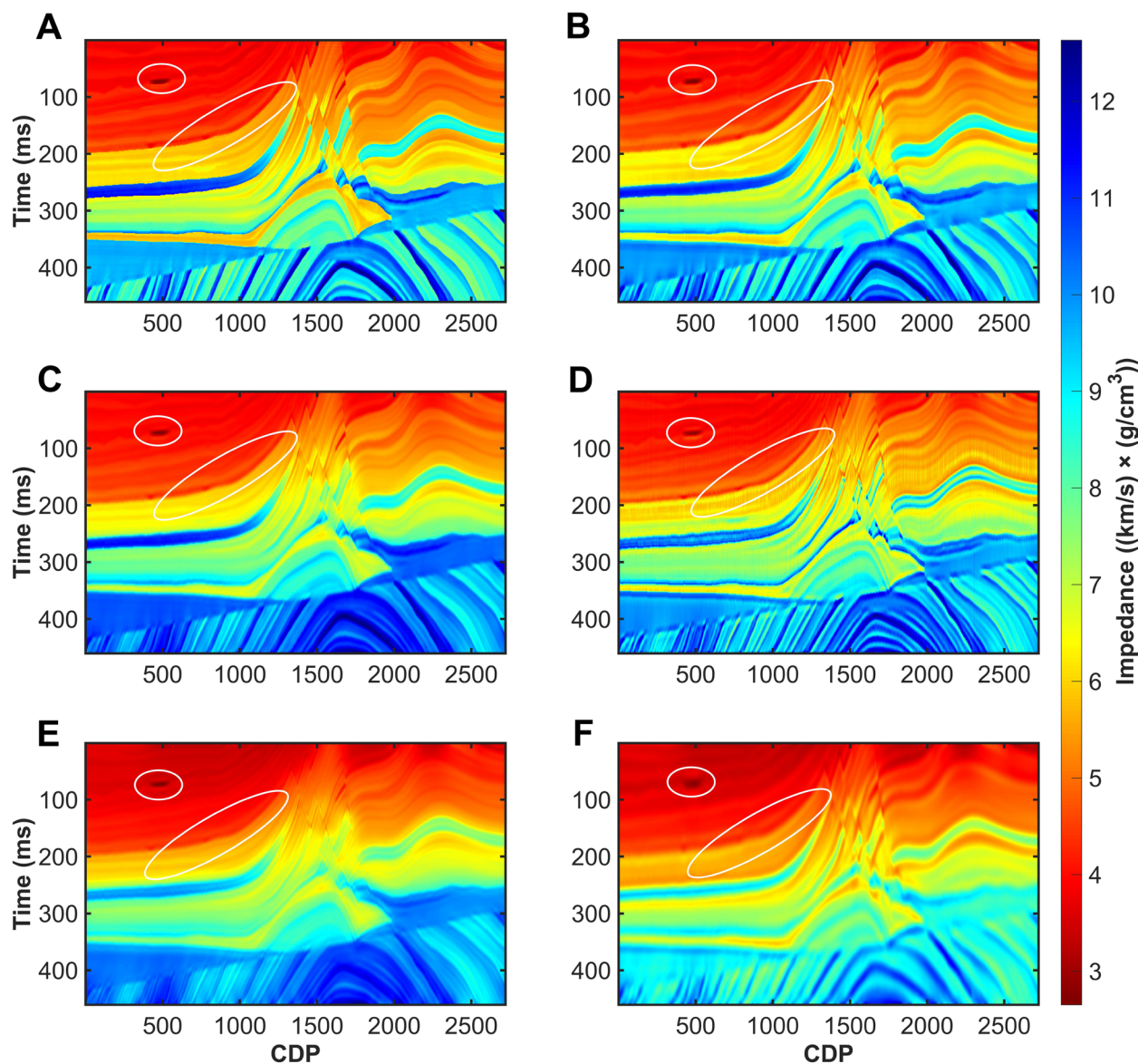


Figure 5. The impedance predicted by the six methods. (A) Proposed method; (B) proposed method without frequency-aware components; (C) the transformer-only model; (D) atrous spatial pyramid pooling-only model; (E) convolutional neural network baseline; and (F) model-based inversion. Abbreviation: CDP: Common depth point.

the Transformer branch is retained, the MSE increases to 0.3832 (174.96% increase) and the PCC decreases to 0.9626 (2.33% reduction) compared with the complete model. Similarly, using only the ASPP branch increases the MSE to 0.3270 (134.58% increase) and reduces the PCC to 0.9614 (2.45% reduction). These results indicate that both global dependency modeling provided by Transformer and local multi-scale feature extraction provided by ASPP are essential, and their combination produces complementary improvements.

For the field dataset, the proposed method achieves an MSE of 0.0858 and a PCC of 0.8164. Removing the frequency-aware mechanism increases the MSE to 0.1213 (41.38% increase) and decreases the PCC to 0.7655 (6.24% reduction). The Transformer-only model produces an MSE of 0.1425 (66.08% increase) and a PCC of 0.6392 (21.70% reduction). The ASPP-only model results in an MSE of 0.1581 (84.27% increase) and a PCC of 0.6370 (21.96% reduction). These results confirm that the proposed method achieves more reliable and accurate inversion

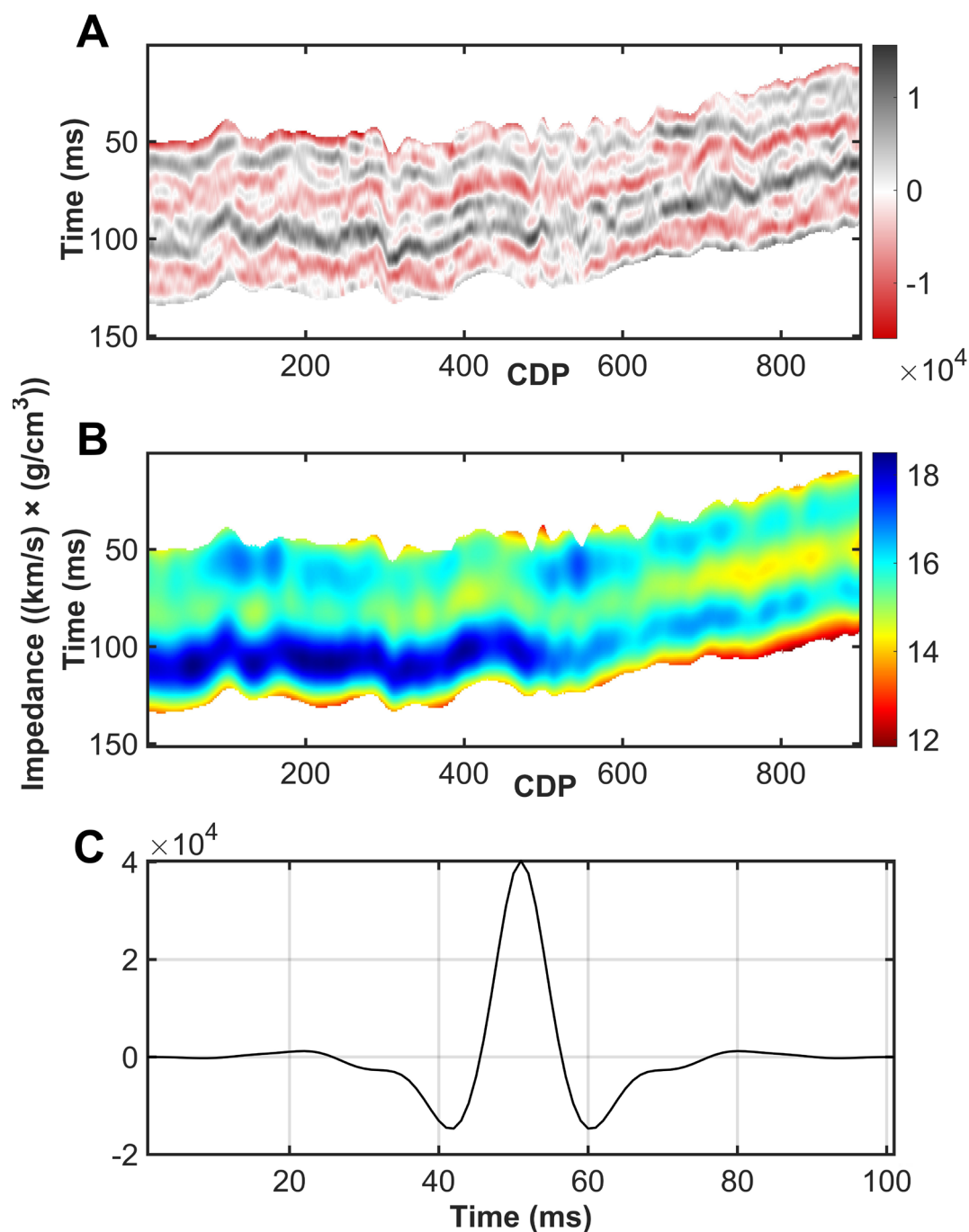


Figure 6. The field seismic data, initial impedance, and the wavelet used. (A) Field seismic data; (B) initial impedance; and (C) wavelet. Abbreviation: CDP: Common depth point.

outcomes.

Despite the effectiveness of the proposed method, several limitations still exist that may affect its practical applicability in real-world seismic inversion tasks. First, the performance of the model is highly sensitive to

hyperparameter settings, including the learning rate and loss weights. We conduct a sensitivity analysis for loss weight and learning rate based on actual data from this study. Table 3 shows the MSE and PCC with a fixed learning rate of 0.001 and varying loss weights. Table

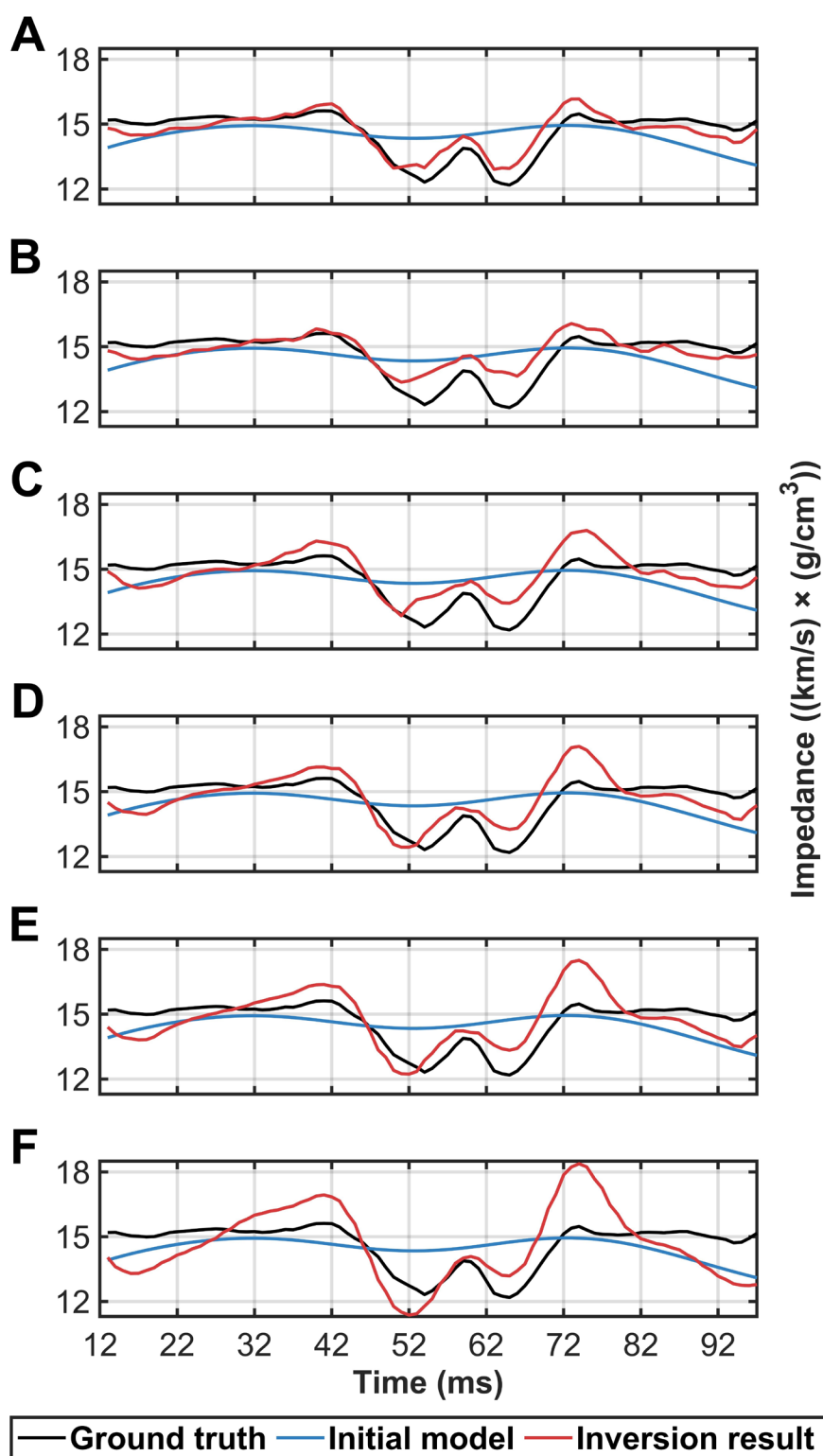


Figure 7. Blind-well (W3) impedance predictions obtained from the field data using six methods. (A) Proposed methods; (B) the proposed method without frequency-aware components; (C) transformer-only model; (D) atrous spatial pyramid pooling-only model; (E) convolutional neural network; and (F) model-based inversion.

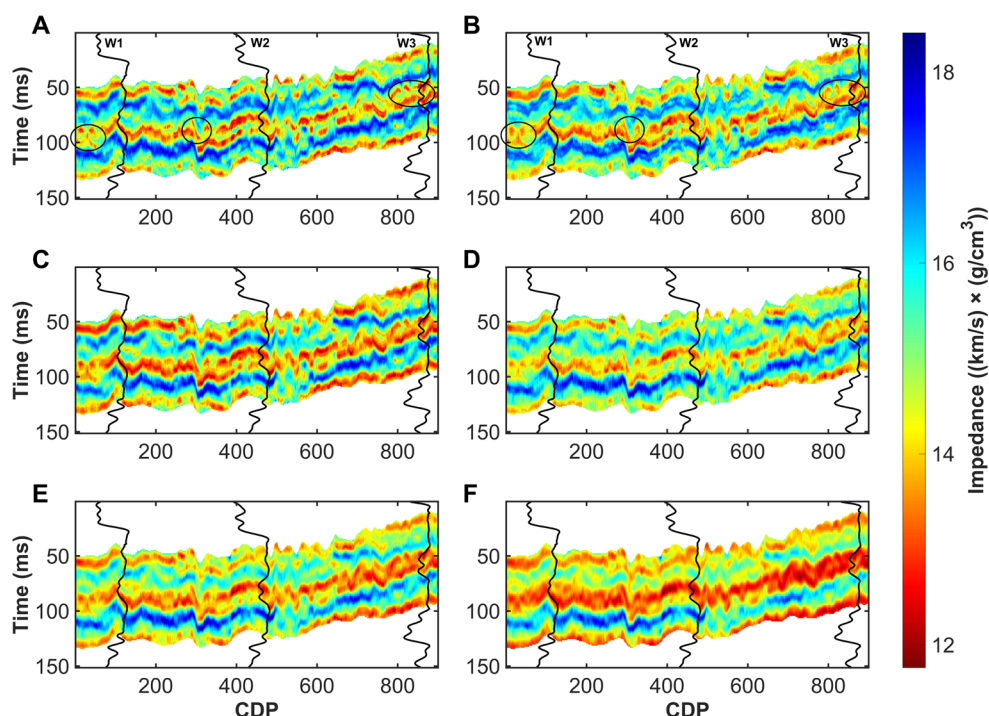


Figure 8. The impedance predicted by the six methods. (A) Proposed method; (B) the proposed method without frequency-aware components; (C) transformer-only model; (D) atrous spatial pyramid pooling-only method; (E) convolutional neural network baseline; and (F) model-based inversion. Abbreviation: CDP: Common depth point.

4 shows the MSE and PCC when the learning rate is changed with fixed loss weights of 0.5, 0.3, and 0.2. As shown in Table 3, different combinations of loss weights result in different inversion performance. The optimal configuration (0.5, 0.3, 0.2) achieves the lowest MSE (0.0858) and the highest PCC (0.8164). Table 4 shows the influence of learning rate on model performance. The best performance is obtained with a learning rate of 0.001 (MSE = 0.0858, PCC = 0.8164). The experiments demonstrate that the proposed model is affected by hyperparameter selection, and the selected parameters provide a suitable balance between reconstruction accuracy and training stability. However, at present, there is limited theoretical guidance for selecting optimal configurations under different geological conditions or data distributions. As a result, model development often relies heavily on empirical trial and error. In some cases, automated hyperparameter optimization strategies such as grid search, random search, or Bayesian optimization can be employed to improve efficiency and performance. Nevertheless, these approaches significantly increase computational cost and training time, especially when dealing with large-scale seismic datasets or complex hybrid network architectures, thereby limiting their practicality in time-sensitive exploration scenarios or industrial applications.

Second, the generalization ability of the proposed model across different seismic surveys remains limited. This issue is primarily caused by domain shifts between datasets, which may arise from differences in acquisition geometry, source and receiver characteristics, sampling intervals, processing flows, noise levels, wavelet signatures, and even underlying geological settings. As a result, a model trained on one dataset may not directly transfer to another dataset with different statistical or physical properties. In practical applications, this often necessitates retraining or fine-tuning the model using newly labeled pairs of seismic traces and impedance logs for each survey area. However, only one field seismic dataset with reliable well-log impedance labels was available in this study. Therefore, a comprehensive cross-survey evaluation was beyond the scope of this work. This limitation reflects a common challenge in data-driven geophysical inversion frameworks. Cross-survey generalization therefore warrants further investigation. Future work will focus on improving model transferability through domain adaptation, transfer learning, and self-supervised learning strategies.

Lastly, similar to most CNN-based and hybrid deep learning inversion approaches, the proposed method may

be prone to overfitting when training data are limited. Although the incorporation of frequency constraints, multi-branch feature extraction, and global context modeling can partially mitigate this issue, the network may still memorize dataset-specific patterns rather than learning generalizable, physically consistent representations. This problem becomes more severe in complex geological environments where seismic responses are highly non-stationary and exhibit strong lateral heterogeneity. In such cases, the network may produce overly smooth impedance models or introduce artificial discontinuities when applied to unseen data, which can negatively impact subsequent geological interpretation. Regularization strategies and data augmentation techniques can alleviate these issues to some extent, but they do not fully resolve the underlying problem of dependence on training data.

Table 3. MSE and PCC for different loss-weight combinations at a fixed learning rate

Hyperparameters			MSE	PCC
α_1	α_2	α_3		
0.5 ^a	0.3 ^a	0.2 ^a	0.0858 ^a	0.8164 ^a
0.5	0.1	0.4	0.1103	0.7705
0.5	0.4	0.1	0.1087	0.7812
0.6	0.3	0.1	0.1025	0.7835
0.4	0.3	0.3	0.1128	0.7713
0.7	0.1	0.2	0.0997	0.7983
0.3	0.5	0.2	0.1014	0.7896

Note: ^aBest-performing value.

Abbreviations: MSE: Mean squared error; PCC: Pearson correlation coefficient.

Table 4. MSE and PCC for different learning rates with fixed loss weights

Learning rate	MSE	PCC
0.001 ^a	0.0858 ^a	0.8164 ^a
0.0001	0.1336	0.7203
0.0002	0.1194	0.7458
0.0005	0.0995	0.7784
0.002	0.1032	0.7653
0.005	0.1183	0.7502
0.01	0.1558	0.6984

Note: ^aBest-performing value.

Abbreviations: MSE: Mean squared error; PCC: Pearson correlation coefficient.

In future work, we will focus on addressing these

limitations by improving model robustness, enhancing cross-domain generalization, and reducing reliance on labeled training data through the integration of physics-informed learning and advanced self-supervised frameworks. In addition, we will explore extending the proposed method to pre-stack seismic inversion, enabling the joint estimation of elastic and reservoir parameters from angle-dependent seismic data.

5. Conclusion

We demonstrate that the proposed hybrid network combined with a frequency-aware loss can improve the accuracy of post-stack seismic impedance inversion. This design effectively enhances the sensitivity of the network to subtle spectral variations, which are often associated with thin-bed responses and small-scale geological heterogeneities. The ablation results indicate that the transformer can capture global contextual information, the ASPP can capture local multi-scale features, and the frequency-aware loss encourages the network to focus more on frequency components that are harder to fit, thereby enhancing detail recovery in thin-bed responses. Experimental results on both a synthetic dataset and a field dataset further validate the robustness and effectiveness of the proposed method. On the synthetic dataset, where ground truth is available, the proposed model achieves lower reconstruction error and better structural consistency compared with baseline methods. On the field seismic dataset, the predicted impedance exhibits higher resolution and clearer stratigraphic boundaries. These improvements are particularly beneficial for subsequent interpretation tasks such as horizon tracking, fault identification, and reservoir characterization. Overall, the results indicate that combining hybrid deep learning architectures with frequency-aware learning strategies offers a promising strategy for improving the accuracy, resolution, and geological consistency of seismic impedance inversion.

Acknowledgments

None.

Funding

This research was funded by the CNPC Science and Technology Major Project (grant nos. 2023ZZ05, 2023ZZ05-05, and 2023ZZ18-03).

Conflict of interest

The authors declare that they have no competing interests.

Author contributions

Conceptualization: Hong Cao, Zhifang Yang, Hao Yang,

Qiang Ge

Formal analysis: Yuqi Qiu, Jiangbei Huang, Sen Zhao

Investigation: Hao Yang

Methodology: Yang Zhang

Writing—original draft: Yang Zhang

Writing—review & editing: Yang Zhang

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

The non-confidential data used in this article can be obtained by contacting the corresponding author.

References

- Wright SJ, Holt JN. A new non-linear least squares algorithm for the seismic inversion problem. *Geophys J Int*. 1986;87(3):1041-1056.
doi: 10.1111/j.1365-246X.1986.tb01982.x
- Smith GC, Gidlow PM. Weighted stacking for rock property estimation and detection of gas. *Geophys Prospect*. 1987;35(9):993-1014.
doi: 10.1111/j.1365-2478.1987.tb00856.x
- Fu L. Joint inversion of seismic data for acoustic impedance. *Geophysics*. 2004;69(4):994-1004.
doi: 10.1190/1.1778242
- Wu X. Structure-, stratigraphy- and fault-guided regularization in geophysical inversion. *Geophys J Int*. 2017;210(1):184-195.
doi: 10.1093/gji/ggx150
- Zhang R, Castagna J. Seismic sparse-layer reflectivity inversion using basis pursuit decomposition. *Geophysics*. 2011;76(6):R147-R158.
doi: 10.1190/geo2011-0103.1
- Chai X, Wang S, Yuan S, Zhao J, Sun L, Wei X. Sparse reflectivity inversion for nonstationary seismic data. *Geophysics*. 2014;79(3):V93-V105.
doi: 10.1190/geo2013-0313.1
- Yuan S, Wang S, Luo C, He Y. Simultaneous multitrace impedance inversion with transform-domain sparsity promotion. *Geophysics*. 2015;80(2):R71-R80.
doi: 10.1190/geo2014-0065.1
- Wang L, Zhao Q, Gao J, Xu Z, Fehler M, Jiang X. Seismic sparse-spike deconvolution via Toeplitz-sparse matrix factorization. *Geophysics*. 2016;81(3):V169-V182.
doi: 10.1190/geo2015-0151.1
- Chen G, Yang W, Liu Y, Wang H, Huang X. Salt structure elastic full waveform inversion based on the multiscale signed envelope. *IEEE Trans Geosci Remote Sens*. 2022;60:1-12.
doi: 10.1109/TGRS.2022.3166028
- Chen GX, Wu RS, Chen SC. Multiscale direct envelope inversion: Algorithm and methodology for application to the salt structure inversion. *Earth Space Sci*. 2019;6(1):174-190.
doi: 10.1029/2018EA000453
- Chen G, Yang W, Wang H, Zhou H, Huang X. Elastic full waveform inversion based on full-band seismic data reconstructed by dual deconvolution. *IEEE Geosci Remote Sens Lett*. 2022;19:1-5.
doi: 10.1109/LGRS.2022.3178915
- Das V, Pollack A, Wollner U, Mukerji T. Convolutional neural network for seismic impedance inversion. In: *SEG Technical Program Expanded Abstracts 2018*. Tulsa, OK: Society of Exploration Geophysicists; 2018:2071-2075.
doi: 10.1190/segam2018-2994378.1
- Wu X, Yan S, Bi Z, Zhang S, Si H. Deep learning for multidimensional seismic impedance inversion. *Geophysics*. 2021;86(5):R735-R745.
doi: 10.1190/geo2020-0564.1
- Biswas R, Sen MK, Das V, Mukerji T. Prestack and poststack inversion using a physics-guided convolutional neural network. *Interpretation*. 2019;7(3):SE161-SE174.
doi: 10.1190/INT-2018-0236.1
- Zhang J, Li J, Chen X, Li Y, Huang G, Chen Y. Robust deep learning seismic inversion with a priori initial model constraint. *Geophys J Int*. 2021;225(3):2001-2019.
doi: 10.1093/gji/ggab074
- Wu B, Meng D, Zhao H. Semi-Supervised Learning for Seismic Impedance Inversion Using Generative Adversarial Networks. *Remote Sens*. 2021;13(5):909.
doi: 10.3390/rs13050909
- Shi S, Li M, Wang J, Chang W, Li L, Xie D. Seismic impedance inversion based on semi-supervised learning. *Comput Geosci*. 2024;182:105497.
doi: 10.1016/j.cageo.2023.105497
- Wei Z, Li S, Ning J, Chen X, Yang X. Low-frequency constrained seismic impedance inversion combining large kernel attention and long short-term memory. *Acta Geophys*. 2024.
doi: 10.1007/s11600-024-01298-3
- Shu J, Zou C, Peng C, et al. Evaluating gas saturation in unconventional gas reservoirs using acoustic logs: A case

- study of the Baiyun Depression in the northern South China Sea. *J Mar Sci Eng*. 2025;13(11):2078.
doi: 10.3390/jmse13112078
20. Tao L, Gu Z, Ren H. Improving the seismic impedance inversion by fully convolutional neural network. *J Mar Sci Eng*. 2025;13(2):262.
doi: 10.3390/jmse13020262
 21. Li H, Huang X, Yang S, Cui X, Li Y, Yang R. Integrated rock physics-based interpretation of time-lapse seismic data for residual oil detection in offshore waterflooded reservoirs. *J Mar Sci Eng*. 2026;14(1):91.
doi: 10.3390/jmse14010091
 22. Li Z, Li M, Liu G, Dong Y, Wang Y, Wang Y. Complex reservoir lithology prediction using sedimentary facies-controlled seismic inversion constrained by high-frequency stratigraphy. *J Mar Sci Eng*. 2025;13(8):1390.
doi: 10.3390/jmse13081390
 23. Long S, Wang X. Seismic impedance inversion based on multi-scale dual-branch semi-supervised learning. *J Appl Geophys*. 2026;251:106326.
doi: 10.1016/j.jappgeo.2026.106326
 24. Liu Z, Liu D, Sacchi MD, Li J, Chen X, Zhang W. Diffusion posterior sampling with 3D lateral constraints for post-stack impedance inversion. *J Appl Geophys*. 2026;251:106248.
doi: 10.1016/j.jappgeo.2026.106248
 25. Zhang Y, Wang Y, Geng W, She Y, Zhou H. Total variation-regularized poststack seismic impedance inversion based on the L1–2 norm. *Geophysics*. 2026;91(4):R191–R203.
doi: 10.1190/GEO-2025-0455
 26. Peng J, Liu Y, Wang M, Li Y, Feng W. Encoder–Inverter framework for seismic acoustic impedance inversion. *Eng Appl Artif Intell*. 2026;176:114673.
doi: 10.1016/j.engappai.2026.114673
 27. Peng J, Wang X, Liu Y, Li Y, Wang M. Physics-informed cross-learning for seismic acoustic impedance inversion and wavelet extraction. *Comput Geotech*. 2026;194:108010.
doi: 10.1016/j.compgeo.2026.108010
 28. Xiao X, Jiang S, Dong X, et al. MSCNN-transformer-based logging curve reconstruction for acoustic impedance inversion in shale reservoirs: A case study of the Kong2 member in the Cangdong sag. *J Appl Geophys*. 2026;248:106169.
doi: 10.1016/j.jappgeo.2026.106169
 29. Xiang Q, Dai R. Seismic Impedance Inversion With Adaptive Group Sparse Regularization. *IEEE Geosci Remote Sens Lett*. 2026;23:1–5.
doi: 10.1109/lgrs.2026.3680028
 30. Sinha S, Routh PS, Anno PD, Castagna JP. Spectral decomposition of seismic data with continuous-wavelet transform. *Geophysics*. 2005;70(6):P19–P25.
doi: 10.1190/1.2127113
 31. De Matos MC, Osorio PL, Johann PR. Unsupervised seismic facies analysis using wavelet transform and self-organizing maps. *Geophysics*. 2007;72(1):P9–P21.
doi: 10.1190/1.2392789
 32. Kazemeini SH, Juhlin C, Zinck-Jørgensen K, Norden B. Application of the continuous wavelet transform on seismic data for mapping of channel deposits and gas detection at the CO2 SINK site, Ketzin, Germany. *Geophys Prospect*. 2009;57(1):111–123.
doi: 10.1111/j.1365-2478.2008.00723.x
 33. Chen LC, Zhu Y, Papandreou G, Schroff F, Adam H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In: Ferrari V, Hebert M, Sminchisescu C, Weiss Y, eds. *Computer Vision – ECCV 2018*. Lecture Notes in Computer Science. Vol 11211. Cham, Switzerland: Springer International Publishing; 2018:833–851.
doi: 10.1007/978-3-030-01234-2_49
 34. Lin Y. Multi-scale seismic impedance inversion based on Transformer model and deep learning. *Eng Res Express*. 2025;7(1):015209.
doi: 10.1088/2631-8695/ada48d
 35. Vaswani A, Shazeer N, Parmar N, et al. Attention Is All You Need. In: *Advances in Neural Information Processing Systems* 30. 2017.
 36. Jiang L, Dai B, Wu W, Loy CC. Focal Frequency Loss for Image Reconstruction and Synthesis. In: *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. New York, NY: IEEE; 2021:13899–13909.
doi: 10.1109/ICCV48922.2021.01366
 37. Lin TY, Goyal P, Girshick R, He K, Dollár P. Focal Loss for Dense Object Detection. In: *2017 IEEE International Conference on Computer Vision (ICCV)*. New York, NY: IEEE; 2017:2999–3007.
doi: 10.1109/ICCV.2017.324
 38. He K, Zhang X, Ren S, Sun J. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. In: *2015 IEEE International Conference on Computer Vision (ICCV)*. New York, NY: IEEE; 2015:1026–1034.
doi: 10.1109/ICCV.2015.123