frontiers | Research Topics

Advances and applications of artificial intelligence in geoscience and remote sensing

Edited by
Peng Zhenming, Sanyi Yuan, Xiaolan Qiu, Wenjuan Zhang and Anna Sowizdzal

Published in
Frontiers in Earth Science
Frontiers in Marine Science

ISSN 1664-8714
ISBN 978-2-8325-3157-0
DOI 10.3389/978-2-8325-3157-0

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Advances and applications of artificial intelligence in geoscience and remote sensing

Topic editors

Peng Zhenming — University of Electronic Science and Technology of China, China

Sanyi Yuan — China University of Petroleum, Beijing, China

Xiaolan Qiu — Aerospace Information Research Institute, Chinese Academy of Sciences (CAS), China

Wenjuan Zhang — Aerospace Information Research Institute, Chinese Academy of Sciences (CAS), China

Anna Sowizdzal — AGH University of Science and Technology, Poland

Citation

Zhenming, P., Yuan, S., Qiu, X., Zhang, W., Sowizdzal, A., eds. (2023). *Advances and applications of artificial intelligence in geoscience and remote sensing*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-3157-0

Table of contents

frontiers | Frontiers in Earth Science

TYPE Editorial
PUBLISHED 20 July 2023
DOI 10.3389/feart.2023.1234360

OPEN ACCESS

EDITED BY
Taskin Kavzoglu,
Gebze Technical University, Türkiye

REVIEWED BY
Bangyu Wu,
Xi'an Jiaotong University, China

*CORRESPONDENCE
Zhenming Peng,
zmpeng@uestc.edu.cn

RECEIVED 04 June 2023
ACCEPTED 10 July 2023
PUBLISHED 20 July 2023

CITATION
Peng Z, Yuan S, Qiu X, Zhang W and Sowizdzal A (2023), Editorial: Advances and applications of artificial intelligence in geoscience and remote sensing. *Front. Earth Sci.* 11:1234360. doi: 10.3389/feart.2023.1234360

Editorial: Advances and applications of artificial intelligence in geoscience and remote sensing

Zhenming Peng[1]*, Sanyi Yuan[2], Xiaolan Qiu[3], Wenjuan Zhang[3] and Anna Sowizdzal[4]

[1]School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, China, [2]College of Geophysics, China University of Petroleum, Beijing, China, [3]Aerospace Information Research Institute, Chinese Academy of Sciences (CAS), Beijing, China, [4]Department of Energy Resources, Faculty of Geology, Geophysics and Environmental Protection, AGH University of Krakow, Kraków, Poland

KEYWORDS

intelligent exploration geophysics, Earth observation, optical remote sensing, artificial intelligence, object detection

Editorial on the Research Topic
Advances and applications of artificial intelligence in geoscience and remote sensing

The Earth is the space that human beings depend on. Earth observation and remote sensing technology use modern information technology to carry out disaster and threat early warning, mineral and resource detection, climate change and other real-time monitoring, prediction and distribution law exploration of adjacent space, surface and internal structure of the Earth. It has been widely used in military, civil, energy and other fields, and has become an indispensable information technology means for the development of human society.

In recent years, with the continuous development of Earth Science and remote sensing technology, especially the continuous emergence of different detection sensors and new detection systems, and the continuous accumulation of historical data and samples, it is possible to use artificial intelligence (AI) for big data analysis, and it has become a research hotspot in this field.

In the field of oil and gas seismic exploration, technologies such as seismic data processing and reservoir prediction have shifted from classic signal processing methods to data-driven artificial intelligence methods, specifically including: 1) seismic data processing. In this Research Topic, newly developed artificial intelligence models are utilized to solve seismic denoising (He et al.), velocity analysis (Wang D. et al.), data reconstruction, etc., in order to minimize the negative impact of perceived factors as much as possible. 2) Reservoir parameter inversion and oil and gas prediction, automatic fault tracking shear wave velocity prediction (Wang H. et al.), logging modeling, seismic wave field forward modeling, seismic impedance inversion(), rock fracture detection, etc.

For remote sensing super-resolution image restoration and reconstruction, authors proposed a novel Auto-weighted low-rank Tensor Ring Factorization with

Hybrid Smoothness regularization (ATRFHS) for mixed noise removal in HIS (Wang Z. et al.).

For Remote sensing target detection and recognition, authors presented a system which uses multiple sensors and a convolutional neural network (CNN) architecture to test cross-sensor object detection resiliency (Mohan and Simske, 2023).

In the future fields of Earth science and remote sensing, artificial intelligence may play a more important role and have greater development space. Especially artificial intelligence models driven by sufficient knowledge, which do not rely on the neural network structure of large models and targeted interpretable networks, are worthy of attention.

A total of 20 submissions were received for the advances and applications of artificial intelligence in geoscience and remote sensing, and after peer review, 13 manuscripts were accepted, involving 55 authors.

Thanks all authors for sharing their latest achievements and contributions to promoting the application of artificial intelligence technology in the field of geoscience and remote sensing.

References

Mohan, V., and Simske, S. J. (2023). Cross-sensor vision system for maritime object detection. *Front. Mar. Sci.* 10, 1112955. doi:10.3389/fmars.2023.1112955

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

frontiers | Frontiers in Earth Science

TYPE Original Research
PUBLISHED 05 January 2023
DOI 10.3389/feart.2022.1022874

OPEN ACCESS

EDITED BY
Wenjuan Zhang,
Aerospace Information Research Institute (CAS), China

REVIEWED BY
Jize Xue,
Northwestern Polytechnical University, China
Fengchao Xiong,
Nanjing University of Science and Technology, China
Yong Chen,
Jiangxi Normal University, China

*CORRESPONDENCE
Juan Wang,
wjuan0712@cwnu.edu.cn

SPECIALTY SECTION
This article was submitted to Environmental Informatics and Remote Sensing, a section of the journal Frontiers in Earth Science

RECEIVED 19 August 2022
ACCEPTED 11 November 2022
PUBLISHED 05 January 2023

CITATION
Luo X, Lv J, Wang B, Liu D and Wang J (2023), Hyperspectral image restoration *via* hybrid smoothness regularized auto-weighted low-rank tensor ring factorization.
Front. Earth Sci. 10:1022874.
doi: 10.3389/feart.2022.1022874

Hyperspectral image restoration *via* hybrid smoothness regularized auto-weighted low-rank tensor ring factorization

Xuegang Luo[1,2], Junrui Lv[2], Bo Wang[2], Dujin Liu[3] and Juan Wang[4]*

[1]College of Information and Communication Engineering, University of Electronic Science and Technology, Chengdu, Sichuan, China, [2]School of Mathematics and Computer Science, Panzhihua University, Panzhihua, Sichuan, China, [3]College of Intelligent Manufacturing, Sichuan University of Arts and Sciences, Dazhou, China, [4]College of Computer Science, China West Normal University, Nanchong, Sichuan, China

Restoration of hyperspectral images (HSI) is a crucial step in many potential applications as a preprocessing step. Recently, low-rank tensor ring factorization was applied for HSI reconstruction, which has high-order tensors' powerful and generalized representation ability. Although low-rank TR-based approaches with nuclear norm regularization achieved successful results for restoring hyperspectral images, there is still room for improved tensor low-rank approximation. In this article, we propose a novel Auto-weighted low-rank Tensor Ring Factorization with Hybrid Smoothness regularization (ATRFHS) for mixed noise removal in HSI. Nonlocal Cuboid Tensorization (NCT) is leveraged to transform HSI data into high-order tensors. TR factorization using latent factors rank minimization removes the mixed noise in HSI data. To highlight nuclear norms of factor tensors differently effective, an auto-weighted strategy is employed to reduce the more prominent factors while shrinking the smaller ones. A hybrid regularization combining total variation (TV) and phase congruency (PC) is incorporated into a low-rank tensor ring factorization model for the HSI noise removal problem. This efficient combination yields sharper edge preservation and resolves this weakness of existing pure TV regularization. Moreover, we develop an efficient algorithm for solving the resulting optimization problem using the framework of alternating minimization. Extensive experimental results demonstrate that our proposed method can significantly outperform existing approaches for mixed noise removal in HSI. The proposed algorithm is validated on synthetic and natural HSI data.

KEYWORDS

mixed noise removal, low-rank tensor ring, auto-weighted strategy, hybrid smoothness regularization, hyperspectral (H) image

1 Introduction

Hyperspectral imaging is acquired by employing specialized sensors to capture data at numerous narrow wavelengths, ranging from 400 nm to 2500 nm in the same region. It is generally represented as a three-dimension image in which each image represents one of the tens or hundreds of narrow wavelength ranges or spectral bands. However, HSIs are frequently contaminated by various noises during the capture and transmission process, including Gaussian noise, stripes, deadlines, impulse noise, and hybrids (Bioucas-Dias et al., 2013), making further analysis and use of HSIs challenging. Therefore, the noise removal from HSI is an essential task as a preprocessing step and attracted lots of attention (Dabov et al., 2007; Zhang et al., 2013; Chen et al., 2017; Wu et al., 2017; Aggarwal and Majumdar, 2016; Wang et al., 2017; Zhang et al., 2014; Huang et al., 2017; Fan et al., 2017; Chen et al., 2018; Liu et al., 2012)).

Because a high-dimensional HSI is composed of hundreds of separate images banded together, each band of HSIs is regarded as a two-dimensional image. Then, traditional image restoration methods are applied to remove noise band-by-band, such as BM3D (Dabov et al., 2007) and low-rank matrix approximation (Zhang et al., 2013; Zhang et al., 2014; Chen et al., 2017). The matrix-based denoising approach uses conventional two-dimensional image denoising methods and unfolds the three-dimensional tensor into a matrix or treats each band independently. Traditional HSIs denoising algorithms can only evaluate the structural properties of each pixel or band separately, neglecting the significant relationships between all spectral bands and global structure information. Various improved approaches have been developed to compensate for the shortcomings by considering the correlation between all spectral bands.

An HSI is a three-dimensional image stack having two spatial dimensions and one spectral dimension. Therefore, tensors are realistic representations of HSIs data. For the past few years, to fully capture the spatial-spectral correlation of the HSIs, many researchers have employed tensor decompositions to analyze HSI, such as the low-rank tensor method with total variation regularization (Wu et al., 2017), tensor completion with three-layer transform *via* sparsity prior (Xue et al., 2019a) and Laplacian scale mixture (Xue et al., 2021; Xue et al., 2022), missing data recovery (Liu et al., 2014; Yokota et al., 2016), hyperspectral image super-resolution (Dian et al., 2019; Dian and Li, 2019), hyperspectral image restoration with low-rank tensor factorization (Zeng et al., 2020; Xiong et al., 2019; Chen et al., 2019a; He et al., 2022), and hyperspectral image denoising (Chen et al., 2022a; Chen et al., 2022b). These tensor decomposition approaches have the advantage of simultaneously investigating the spatial-spectral correlation between the HSIs inside all bands and better preserving the image's spatial-spectral structure. Nevertheless, they fail to capture HSI's intrinsic high-order low-rank structure and cannot keep a sharper edge.

Many studies have demonstrated the advantages of low-rank tensor approximation techniques in dealing with high-order tensor data. Recently, tensor-ring (TR) (Zhao et al., 2016; Huang et al., 2020) was developed to describe a high-order tensor as a sequence of cyclically contracted third-order tensors, which is the extensional version of tensor train (TT) (Oseledets, 2011). Due to its ability to promise to represent complex interactions within high-dimensional data, TR has received increasing attention. It was utilized in many high-dimensional incomplete data recovery applications, such as HSI CS reconstruction (Chen et al., 2020; He et al., 2019), tensor ring networks (Wang et al., 2018), tensor completion (Yuan et al., 2020; Ding et al., 2022), missing data recovery in high-dimensional images (Wang et al., 2021), and HSI denoising (Chen et al., 2019b; Xue et al., 2019b; Xuegang et al., 2022).

Compared to traditional tensor decomposition, TR decomposition imposed on the tensor approximation has two superiorities. First, the TR factor can be rotated equivalently and circularly in the trace operation, but the traditional tensor decomposition technique cannot turn the core tensor. Second, Since TR provides a tensor-by-tensor representation architecture, the original data structure can be better maintained.

Two representative works on the TR low-rankness characterization are low-rank TR decomposition (LTRD) and TR rank minimization (TRRM) (He et al., 2019). introduced a TR decomposition and total-variation regularized method for the missing information reconstruction of remote sensing images (Chen et al., 2020). described a nonlocal TR Decomposition for HSI denoising. Although the TRD-based approaches have shown good denoising results, TR rank parameter estimation is an NP-hard problem.

The TRRM-based methods, based on the nuclear norm, are a biased approximation to the TR rank and do not need to choose the optimal TR rank. It is more efficient than the former (Wang et al., 2021). presented a weighted TR decomposition model with TR factors nuclear norms and total variation regularization for missing data recovery in high-dimensional optical RS images (Chen et al., 2018). introduced the sum of nuclear norms of all unfolding matrices by the mode-k matricization as the convex surrogate of tensor Tucker rank for the tensor completion problem. To explore the latent features of the whole HSI data, a TRRM model with TR nuclear norm minimization is proposed by (Yuan et al., 2020) and elaborated by a convex surrogate of TR rank of circularly unfolding matrices for high-order missing data completion (Yu et al., 2019). proposed a TRRM-based method with nuclear norm regularization on the latent TR factors by exploiting the rank relationship between the tensor and the TR latent space. An improved version (Ding et al., 2022) by penalizing the *logdet* function onto TR unfolding matrices is proposed as remedies. However, these approaches are predicated on the convex relaxation by weight nuclear norm of the

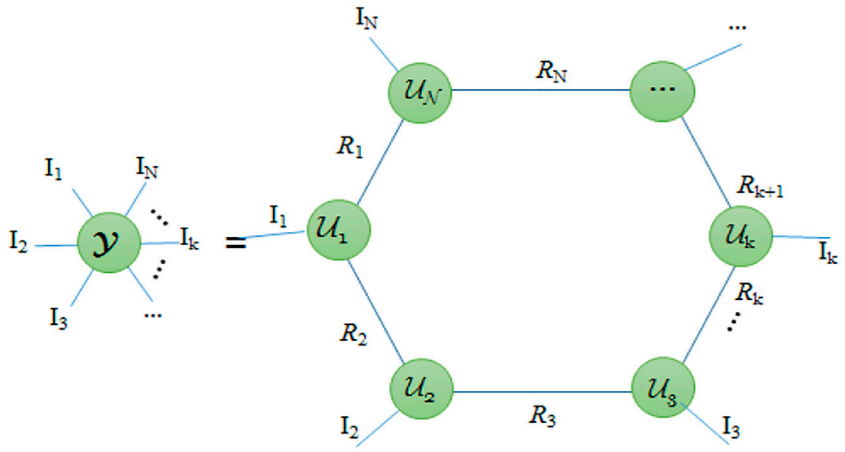


FIGURE 1
Illustration of TR representation of an Nth-order tensor $\mathcal{Y} \in \mathbb{R}^{I_1 \times I_2 \ldots \times I_N}$.

unbalanced TR unfolding matrices, which need manually choose the optimal weight values, resulting in poor solutions in execution. Furthermore, the conventional TR-based methods are inadequate to directly exploit the characteristics of low-rank by the original data and still have much room for improvement.

Due to the unfolding matrix with a much higher rank and larger size, the SVD operators of the rank minimization framework on the unfolding matrix in the TRRM-based methods are time-consuming (Wang et al., 2021). has employed three low-dimensional tensor factors of TR decomposition as a convex surrogate of TR rank for more convenient calculation. The SVD computation is considerably decreased due to the low dimension of TR factors. A better low-rank representation can be efficiently exploited by transforming lower-order tensors into higher-order tensors. As a result, TRRM-based approaches that leveraged low-rank and edge preservation on the original data were insufficient.

Inspired by the high effectiveness of rank minimization on TR latent factor for tensor completion, in this paper, to effectively promote the low-rankness of the solution, we introduce an auto-weight TR factors nuclear norm minimization with hybrid smoothness regularization by total variation (TV) and phase congruency (PC) to restore HSI image, which can more accurately approximate the TR rank and sharper promote edge preservation.

Contributions to this article are as follows.

1) To fully exploit the high-dimensional structure information and the low-rankness of HSI, an auto-weight TR nuclear norm, based on the convex relaxation by penalizing the weighted sum of nuclear norm of TR factors unfolding matrices, is proposed to recover the clean HSI part.
2) To highlight TR unfolding matrices differently effectively, an auto-weighted strategy is utilized to shrink the larger matrices while shrinking the smaller ones. By jointly regularizing TV and PC to promote local smoothness, this efficient combination yields sharper edge preservation and resolves this weakness of existing pure TV regularization.
3) An optimization algorithm with an alternating minimization framework is developed to solve the proposed approach efficiently. Experiments demonstrate that the proposed approach can effectively deal with gauss, strip, and mixed noise and outperform the state-of-the-art competitors in evaluation index and visual assessment.

This paper is organized as follows. To facilitate our presentation, we first introduce some notations, TR decomposition, tensor augmentation, and phase congruency regularization in Section 2. In Section 3, our proposed model is presented. We then develop an efficient framework of alternating minimization for solving the proposed model. In Section 4, extensive experiments on both simulated and real datasets were carried out to illustrate the merits of our model. We finally conclude this paper with some discussions on future research in Section 5.

2 Preliminaries

2.1 Background and notations

We deploy lowercase letters to denote scalars, e.g., $m \in \mathbb{R}$. And vectors are denoted by boldface lowercase letters, e.g., $\boldsymbol{y}$. The upper case letters are represented for matrices, e.g., $\boldsymbol{Y}$. An Nth-order tensor is given by lowercase letters calligraphic letters

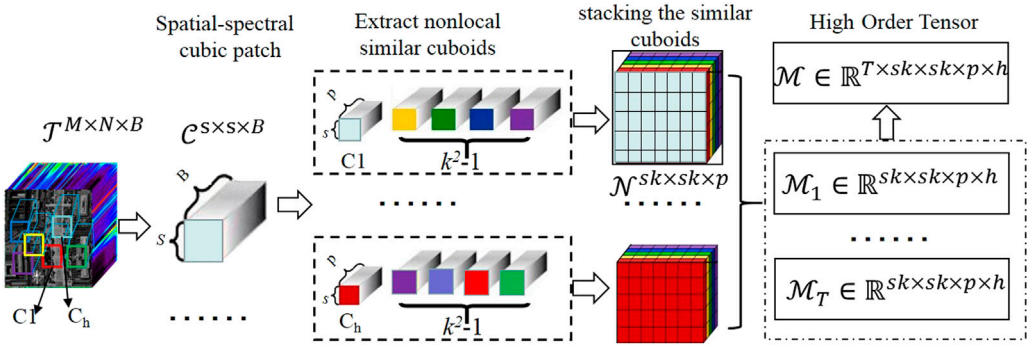


FIGURE 2
Illustration of the procedure to construct a high-order tensor by spatial and spectral similarities of HSI.

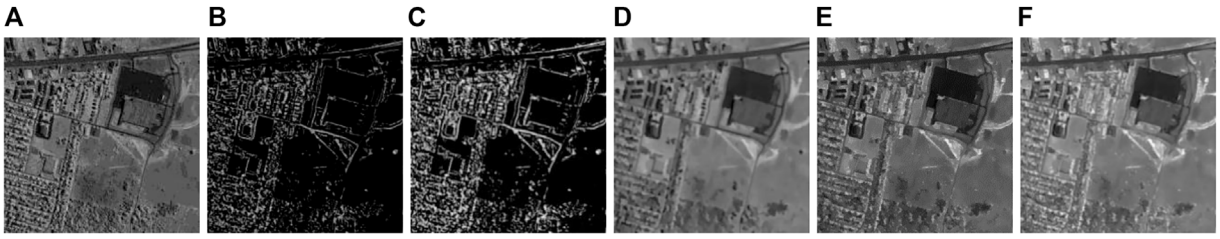


FIGURE 3
Comparison of the denoising results with TV regularization and PC regularization. **(A)** Noisy image from HYDICE urban HSI data, **(B)** Gradient magnitude maps, **(C)** PC feature maps,**(D)** Restored by the model with TV regularization [PSNR:33.58dB; SSIM:0.9345],**(E)** Restored by the model with PC regularization [PSNR:33.41dB; SSIM:0.9541] **(F)** Restored by the model with hybrid smoothness regularization combining TV and PC [PSNR: 34.27dB; SSIM:0.9741].

throughout this paper, e.g., $\mathcal{Y} \in \mathbb{R}^{I_1\times I_2\ldots\times I_N}$ where I_j is the dimension of mode j and j=1,2 . . . , N. The (i_1, i_2, . . . , i_n) entry of tensor $\mathcal{Y}$ is given by $\mathcal{Y}(i_1, i_2, \ldots, i_N)$. A tensor sequence is defined by the set $\{\mathcal{Y}^{(k)}\}_{k=1}^{N}$: = $\{\mathcal{Y}^{(1)}, \mathcal{Y}^{(2)}, ..., \mathcal{Y}^{(N)}\}$, where $\mathcal{Y}^{(k)}$ is the kth tensor of the sequence. $diag(\boldsymbol{Y})$ denoted a column vector consisting of the diagonal elements of $\boldsymbol{Y}$. We use $\mathbf{E}$ to represent an identity matrix. The Frobenius norm of $\boldsymbol{Y}$ is defined as $\|\boldsymbol{Y}\|_2^F = \sqrt{\langle \boldsymbol{Y}, \boldsymbol{Y}\rangle}$. The nuclear norm $\|\boldsymbol{Y}\|_*$ is the sum of singular values of a matrix $\boldsymbol{Y}$.

2.2 Tensor ring low-rank factors

Tensor ring (TR) decomposition is briefly introduced in this subsection. TR representation is to decompose a tensor of higher order into a sequence of latent tensors. As shown in Figure 1, a linear tensor network can graphically represent the TR representation by circular multilinear products over a series of third-order tensors. The number of edges denotes the order of a tensor (which includes matrix and vector). The size of each mode is indicated by the number beside the edges (or dimension). A multilinear product operator between two tensors in a specific manner, also known as tensor contraction, corresponds to the summation over the indices of that mode when two nodes are connected.

For i=1, . . . , N, the TR factors are denoted by a third-order tensor $\mathcal{U}_{(i)} \in \mathbb{R}^{r_{i-1}\times R_i\times r_i}$. The syntax {$R_1$, R_2, . . . , R_N, R_{N+1}}is indicated by the TR rank, which controls the model complexity of TR decomposition and satisfies the R_1= R_{N+1}. Then, $\mathcal{Y}$ can be estimated by tensor with TR format, as $\mathfrak{M}([\mathcal{U}]) = <\mathcal{U}^{(1)}, \mathcal{U}^{(2)}, \ldots, \mathcal{U}^{(N)}>$. Therefore, the element-wise form can be expressed by $\mathcal{Y}(i_1, i_2, \ldots, i_N)$ = Trace($\mathcal{U}^{(1)}(r_1, i_1, r_2), \mathcal{U}^{(2)}(r_2, i_2, r_3), \ldots, \mathcal{U}^{(N)}(r_N, i_N, r_1)$). Trace($Y$) is the matrix trace operation. $Y_{(n)}$ denotes the standard mode-n unfolding of tensor $\mathcal{Y}$.

The relationship between the tensor rank and the corresponding core tensor rank is elaborated, which can be explained by the following theorem. For the nth core tensor $\mathcal{U}^{(n)}$, according to the work of (Yuan et al., 2020; Chen et al., 2020), we define the $Y_{<n>}$ is another mode-n unfolding of tensor $\mathcal{Y}$ used in TR operations denoted by $Y_{<n>} \in \mathbb{R}^{I_n\times I_{n+1}\ldots I_N I_1 I_2\ldots I_{n-1}}$. Thus, we have $Y_{<n>} = U_{(2)}^{(n)}(U_{<2>}^{(\neq n)})^T$, where $U_{<2>}^{(\neq n)}$ is a matrix

by the second unfolding along mode-2 by sequentially merging all core tensors except the nth one and $U_{(i)}^{(n)} \in \mathbb{R}^{R_i \times I_{n-1} I_n}$ is the mode-i unfolding of the nth core tensor. The relation of tensor ring rank and the corresponding factors rank have the following inequality for all $n = 1, \ldots, N$.

$$Rank\left(Y_{(n)}\right) \leq Rank\left(U_{(2)}^{(n)}\right) \tag{1}$$

The detailed proof is available in (Yuan et al., 2020; Chen et al., 2020). The rank of the mode-n unfolding of the tensor $\mathcal{Y}$ is upper bounded by the rank of the dimension-mode unfolding of the corresponding core tensor $\mathcal{U}^{(n)}$, allowing us to impose a low-rank constraint on $[\mathcal{U}]$ to investigate the underlying tensor's more low-rank structure.

2.3 Nonlocal cuboid tensorization for HSI augmentation

Tensor augmentation is an essential preprocessing step for exploiting the local structures and low-rank characteristics since higher-order tensors provide more significant image structure *via* TR decomposition. There are three main ways to transform a tensor into a higher-order one, namely the Reshape Operation (RO) (Yuan et al., 2019), high-order Hankelization (Yokota et al., 2018), and Ket Augmentation (KA) (Yuan et al., 2019). However, the recovered tensors applied RO and KA often have apparent blocking artifacts, while the data were permuted and rearranged without exploiting any neighborhood information. Patch Multiway Delay Embedding Transform (Yokota et al., 2018) is a high-order Hankelization approach, which provides a patch-wise procedure to extract more local information. But this technology increases the amount of HSI data, which makes high computational complexity. An augmented scheme called Nonlocal Cuboid Tensorization (NCT) (Xuegang et al., 2022) can represent HSI data into a high-order one for exploiting low-rank structure representation preferably, simultaneously exploring the nonlocal self-similarity and the spatial-spectral correlation. Therefore, our proposed ATRFHS approach leverages NCT to build HSI augmentation by grouping nonlocal similar cuboids in HSI.

Subsequently, we present the principle of the NCT method. For the recovery processing of an HSI image, $\mathcal{T} \in \mathbb{R}^{x \times y \times b}$ and $\mathcal{X} \in \mathbb{R}^{x \times y \times b}$ with $x \times y$ spatial size and b spectral bands denote the observed and recovered images, respectively. Firstly, for exhibiting rich redundancy in spectra, all cuboid patches $\mathcal{C}_i$ with the size of $s \times s \times p$ across full bands $\mathcal{C} \in \mathbb{R}^{s \times s \times b}$ of HSI in the same spatial locations along the spectral direction with the interval $\frac{p}{2}$ are extracted, we search for its k^2-1 nearest neighbors patches in a local window by Euclidean distance in the same spectra band for each cuboid patch. The k^2-1 similar cuboid patches are stacked into a third-order tensor $\mathcal{N} \in \mathbb{R}^{sk \times sk \times p}$. There are $(\frac{2b}{p} - 1)$ cuboid patches in the same spatial locations with different spectra bands. Thus, as shown in Figure 2, they are grouped into a four-order tensor $\mathcal{M}_i \in \mathbb{R}^{sk \times sk \times p \times h}$ where $h = (\frac{2b}{p} - 1)$.The part HSI with the size of $x \times y \times p$ is divided into $T = \frac{xy}{s^2}$ cuboid patches with the size of $s \times s \times p$.

2.4 Phase congruency regularization

The regularization term can be regarded as the prior knowledge from underlying properties on recovered HS images. Total variation (TV) (Wang et al., 2017) is one of the prevalent regularization approaches applied for image restoration. TV regularization has long been acknowledged as a practical approach for improving image processing smoothness. For third-order hyperspectral data $\mathcal{T}$, the total variation of HSI is denoted by

$$\begin{aligned}\|\mathcal{T}\|_{TV} = \sum_{x,y,b} \Big(&\left|\mathcal{T}_{x,y,b} - \mathcal{T}_{x-1,y,b}\right| + \left|\mathcal{T}_{x,y,b} - \mathcal{T}_{x,y-1,b}\right| \\ &+ \left|\mathcal{T}_{x,y,b} - \mathcal{T}_{x,y,b-1}\right| \Big)\end{aligned} \tag{2}$$

The TV model can effectively remove noise while simultaneously preserving the fine details of the image's edge. However, it is prone to misdiagnose the noises as the edge when the image edge is substantially contaminated by noise and cannot disentangle the noises from the edge. Furthermore, an edge-preserving regularization with gradient magnitudes diffusing along the edges rather than across them results in a staircase (blocky) effect.

To alleviate this shortcoming, phase congruency features are employed in this research to accurately preserve edge information and improve region structure smoothness from a noisy image. Since phase congruency (Morrone and Owens, 1987) is compatible with the properties of signals from corresponding points, it can adequately detect image features. Figure 3 compared the denoising results with TV regular and PC regular. We can see the discrepancy from Figure 3 that high-order information with PC feature maps from Figure 3C is more affluent than the first-order information with horizontal and vertical gradients from Compared to Figure 3D and Figure 3E, restored results in Figure 3F using hybrid smoothness regular combing TV and PC regularization can preserve more details of original images.

Monogenic Phase Congruency (MPC) (Luo et al., 2015; Yuan et al., 2019) has recently increased the precision of feature localization and demonstrated superior computational efficiency and accuracy compared to standard phase congruency. At any specific point x in an image, MPC can be mathematically formulated as

$$C(x) = E(x)\left\lfloor 1 - \xi \times \operatorname{acos}\left(\frac{W'(x)}{B'(x)}\right)\right\rfloor \frac{\lfloor W'(x) - M \rfloor}{B'(x) + \eta} \tag{3}$$

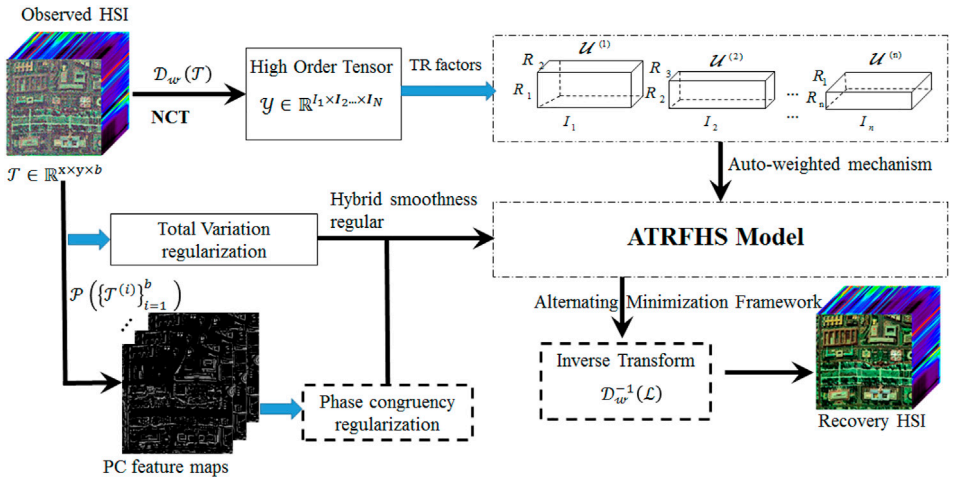


FIGURE 4
Illustration of the proposed ATRFHS for HSI Denoising.

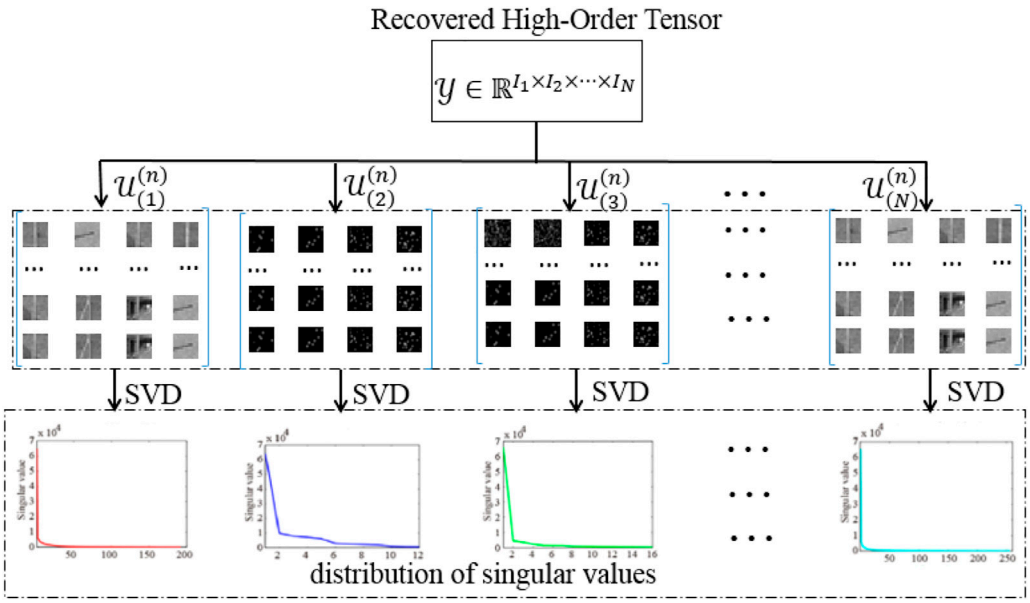


FIGURE 5
The distribution of the singular values of unfolding matrixes.

where $E(x)$ is a weighting function constructed by applying a sigmoid function to the filter response spread value, which is given by (Luo et al., 2015) in detail. Both ξ and η are gain factors approximately from 1 to 2, which sharpen the edge response. M compensates for the influence of noise. $W'(x)$ is local energy information. Similarly, $B'(x)$ is the local amplitude at point x. MPC is capable of retaining both the irregular structure and being impervious to impulse noise. The l_1-norm with phase congruency regularization is generally employed in the fidelity term for impulse noise, similar to total variation regularization.

MPC feature maps are calculated by Eq. 3. Then, monogenic phase congruency regular is obtained by

$$\|\mathcal{P}(\mathcal{T})\|_{\mathrm{PC}} = \sum_{i=1}^{b} C(\mathcal{T}(:,:,i)) \tag{4}$$

TABLE 1 Quantitative results of all the methods under different noise cases FOR WDC and urban DATASETs.

Noise case	Datasets	Index	QRNN3D	LRTF-L_0	LRTDGS	SBNTRD	$L_{1\text{-}2}$ SSTV	ANTRRM	OURS
Case 1	WDC mall	MPSNR(dB)	37.58	37.92	37.68	38.03	37.12	38.12	**38.24**
		MSSIM	0.9234	0.9108	0.9242	0.9524	0.9085	0.9417	**0.9534**
		MFSIM	0.9785	0.9692	0.9741	0.9793	0.9764	0.9742	**0.9798**
		ERGAS	152.79	98.48	164.64	108.48	292.47	102.47	**91.47**
	Indian pines	MPSNR(dB)	34.57	33.98	33.806	34.27	33.48	33.12	**33.78**
		MSSIM	0.9019	0.8918	0.8906	0.9130	0.9067	0.9078	**0.9192**
		MFSIM	0.9729	0.9701	0.9736	0.9430	0.9741	0.9712	**0.9781**
		ERGAS	134.17	76.58	89.47	78.98	80.55	80.24	**74.12**
Case 2	WDC mall	MPSNR(dB)	32.78	33.78	34.89	34.87	34.89	35.02	**35.19**
		MSSIM	0.9262	0.9235	0.9115	0.9231	0.9320	0.9312	**0.9387**
		MFSIM	0.9818	0.9788	0.9784	0.9797	**0.9872**	0.9789	0.9814
		ERGAS	282.16	78.98	88.83	74.43	70.27	71.48	**68.69**
	Indian pines	MPSNR(dB)	29.47	30.24	31.98	32.12	31.34	32.74	**33.12**
		MSSIM	0.7914	0.8645	0.8947	0.9014	0.9124	0.9014	**0.9145**
		MFSIM	0.9691	0.9147	0.9657	0.9678	**0.9665**	0.9602	0.9624
		ERGAS	424.75	142.19	325.18	213.79	**87.93**	132.47	121.97
Case 3	WDC mall	MPSNR(dB)	29.67	30.47	30.65	31.21	30.84	31.02	**31.47**
		MSSIM	0.8701	0.8963	0.8837	0.8947	0.8667	0.8941	**0.9024**
		MFSIM	0.9258	0.9247	0.9347	0.9419	**0.9513**	0.9284	0.9258
		ERGAS	146.9	169.4	164.5	174.9	161.8	145.9	**124.3**
	Indian pines	MPSNR(dB)	33.47	33.12	34.85	34.67	35.31	35.04	**35.42**
		MSSIM	0.9047	0.8987	0.9102	0.9147	**0.9204**	0.9147	0.9194
		MFSIM	0.9541	0.9412	0.9567	0.9641	**0.9678**	0.9412	0.9524
		ERGAS	247.6	183.4	194.7	357.6	368.6	143.7	**134.7**
Case 4	WDC mall	MPSNR(dB)	31.68	31.83	31.20	31.87	30.89	32.11	**32.24**
		MSSIM	0.8762	0.8635	0.8615	0.8831	0.8320	0.9014	**0.9087**
		MFSIM	0.9714	0.9678	**0.9745**	0.9743	0.9578	0.9618	0.9724
		ERGAS	124.47	104.75	135.71	89.65	90.67	88.65	**83.37**
	Indian pines	MPSNR(dB)	28.98	29.47	30.67	30.25	28.04	30.27	**30.96**
		MSSIM	0.8114	**0.8997**	0.8974	0.8914	0.8378	0.8914	0.8987
		MFSIM	0.9214	**0.9404**	0.9374	0.9378	0.9375	0.9289	0.9345
		ERGAS	187.63	104.19	125.18	113.79	286.78	114.64	**102.32**
Case 5	WDC mall	MPSNR(dB)	29.78	31.45	31.86	30.89	31.32	32.11	**32.53**
		MSSIM	0.8867	0.9017	0.8947	0.8897	0.9220	0.9314	**0.9337**
		MFSIM	0.9145	0.9457	0.9378	0.9401	**0.9591**	0.9498	0.9507
		ERGAS	286.3	186.7	157.8	148.6	178.3	133.8	**101.5**
	Indian pines	MPSNR(dB)	31.58	33.38	33.57	33.75	34.08	35.01	**35.47**
		MSSIM	0.8979	0.9378	0.9378	0.9265	0.9304	0.9527	**0.9574**
		MFSIM	0.9970	0.9577	0.9687	0.9555	**0.9654**	0.9504	0.9542
		ERGAS	245.8	189.6	201.5	347.7	147.6	134.5	**114.5**

The best results for each quality index are shown in bold.

3 Proposed model and optimization solution

In this section, we propose a new model for HSI denoising based on weighted low-rank TR factorization using latent factors rank minimization with TV and PC regularization. Then, we introduce an auto-weighted mechanism to establish a tensor completion model and develop the corresponding algorithm based on an alternating minimization framework to solve the model. Figure 4 illustrates the proposed ATRFHS for HSI denoising.

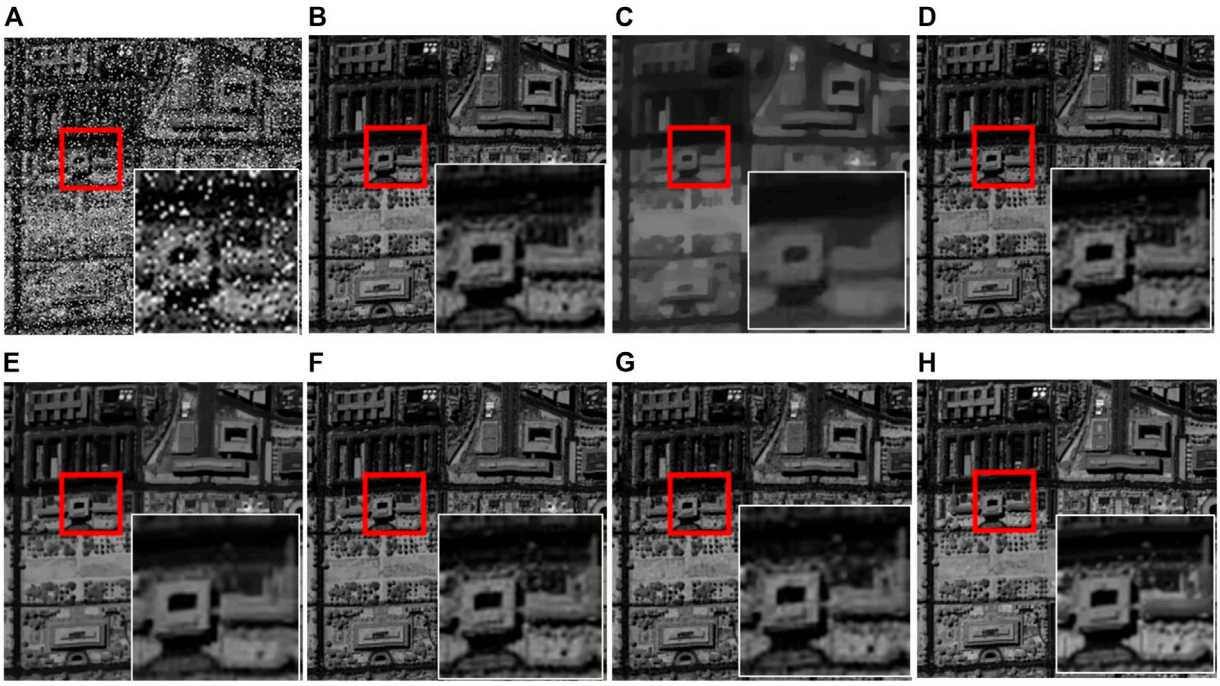


FIGURE 6
Restored results of all comparison methods for band 68 of WDC HSI data under Case 5: **(A)** Noisy, **(B)** $L_{1\text{-}2}$ SSTV, **(C)** QRNN3D, **(D)** LRTDGS, **(E)** SBNTRD, **(F)** LRTF-L_0, **(G)** ANTRRM, **(H)** OURS.

3.1 The proposed algorithm

To facilitate the presentation, for recovering a clean HSI from an observed HSI, by imposing the nuclear norm regularizations on the TR factors, we first review that a high-order tensor is decomposed into a sequence of 3-order tensors using the TR model to find the TR cores of an uncompleted tensor (Yuan et al., 2020), formulated as:

$$\min_{\mathcal{L},\{\mathcal{U}^{(n)}\}_{1:N}} \sum_{n=1}^{N} Rank\left(L_{(n)}\right) + \frac{1}{2}\|\mathcal{Y}-\mathcal{L}\|_F^2 \quad \text{s.t. } \mathcal{L} = \mathfrak{M}\left(\{\mathcal{U}^{(n)}\}_{n=1}^{N}\right) \tag{5}$$

Where $\mathcal{Y} = \mathcal{D}_w(\mathcal{T})$ is a high-order tensor of the observed data $\mathcal{T}$ transformed by NCT, $\mathcal{L}$ is the reconstruction component and $L_{(n)}$ is the standard mode-n unfolding of tensor $\mathcal{L}$. The model can identify the data's low-rank structure and approximate the recovered tensor. But the problem of determining the tensor rank is NP-hard. ANTRRM in (Xuegang et al., 2022) is based on mode-{d, l} unfolding with nuclear norm regularization *via* nonlocal tensor ring. Whereas, the local smoothness and consistency of the HSI in this approach is missed and the time-consuming of SVD computation of mode-{d, l} unfolding matrixes is more expensive than unfolding matrixes of low-rank TR factors.

To solve the above issue (Wang et al., 2021), enforced weight low-rankness on each TR factor. The optimization model can be reformulated as follows,

$$\min_{\mathcal{L},\{\mathcal{U}^{(n)}\}_{1:N}} \sum_{n=1}^{N}\sum_{i=1}^{3}\left\|\mathcal{U}_{(i)}^{(n)}\right\|_* + \frac{1}{2}\|\mathcal{Y}-\mathcal{L}\|_F^2 \quad \text{s.t. } \mathcal{L} = \mathfrak{M}\left(\{\mathcal{U}^{(n)}\}_{n=1}^{N}\right) \tag{6}$$

Where $\mathcal{U}_{(i)}^{(n)}$ is the mode-i unfolding matrix of the nth core tensor of $\{\mathcal{U}^{(n)}\}_{n=1:N}$.

Model (6) can significantly reduce computational complexity compared to model (5). But as the decay distributions of singular values of the unfoldings of the TR factors along mode-n diverge. Appropriate weights should be constructed to determine the contributions of different nuclear norms in unfolding the TR tensor components. Therefore, the approach described above still has space for improvement because exploring low-rankness prior is rarely adequate to extract the underlying data by unreasonable weights. Furthermore, smoothness is another important prior that can be found in high-dimensional HSI data.

From Figure 5, the distribution of singular values significantly differs in the different unfolding matrixes. Weights for different unfolding parts should be treated differently. To reflect different contributions, the weight parameters w play an essential role and need to tread

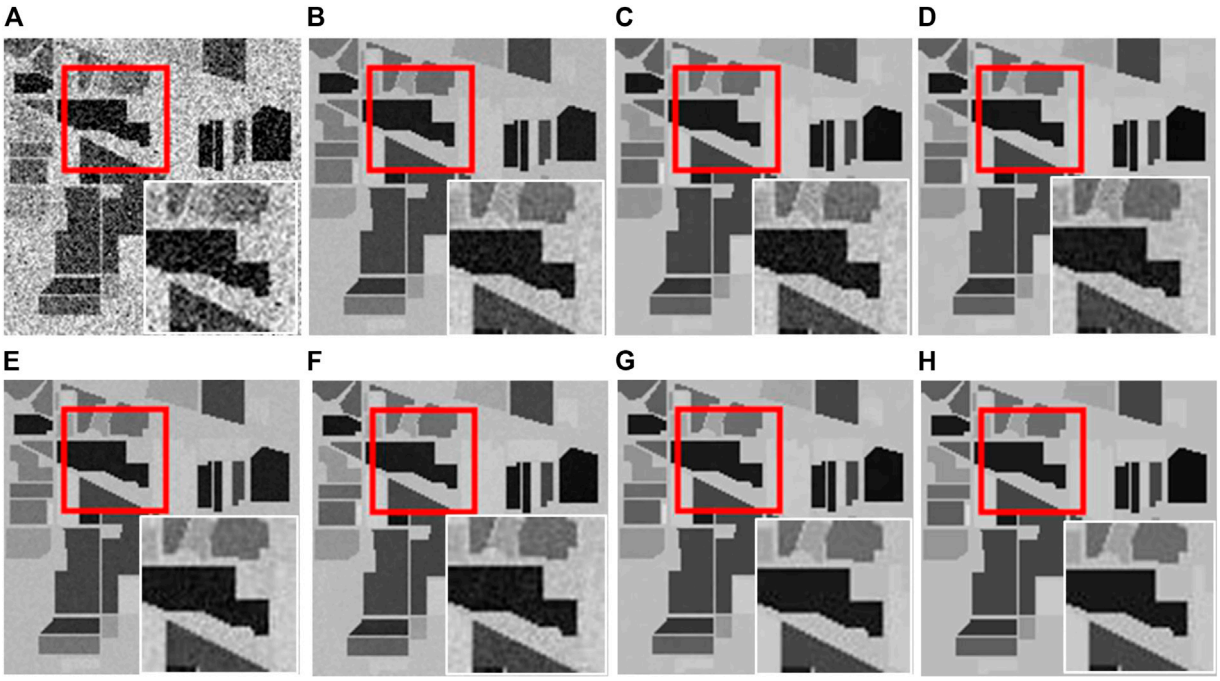


FIGURE 7
Restored results of all comparison methods for band 96 of INDIAN PINES data under Case 3: **(A)** Noisy, **(B)** $L_{1\text{-}2}$ SSTV, **(C)** QRNN3D, **(D)** LRTDGS, **(E)** SBNTRD, **(F)** LRTF-L_0, **(G)** ANTRRM, **(H)**OURS.

carefully. To adapt the TR ranks of different modes, auto-weighted parameter optimization is utilized to measure the importance of different singular values voluntarily, thus minimizing the burden of failure caused by unreasonable weights.

Inspired by this nature, combining the auto-weighted strategy and hybrid smoothness regularization in our work, we can rewrite problem Eq. 6 as the following problem.

$$\min_{\mathcal{L},\boldsymbol{w},\{\mathcal{U}^{(n)}\}_{1:N}} \sum_{n=1}^{N} w_n \sum_{i=1}^{3} \left\|\mathcal{U}_{(i)}^{(n)}\right\|_* + \frac{1}{2}\|\mathcal{Y}-\mathcal{L}\|_F^2 + \gamma\|\boldsymbol{w}\|_F^2$$
$$+\lambda_{TV}\left\|\mathcal{D}_{\omega}^{-1}(\mathcal{L})\right\|_{\mathrm{TV}} + \lambda_{PC}\left\|\mathcal{P}\left(\mathcal{D}_{\omega}^{-1}(\mathcal{L})\right)\right\|_{\mathrm{PC}} \quad \text{s.t.}\mathcal{L} = \mathfrak{M}\left(\{\mathcal{U}^{(n)}\}_{n=1}^{N}\right) \tag{7}$$

Where γ, λ_{TV} and λ_{PC} are regularization parameters, $\{w_n\}_{n=1}^{N}$ are the weight of the nth norm satisfying $w_n \geq 0$ and $\sum_{n=1}^{N} w_n = 1$.

To solve the above problem, auxiliary variables $\mathcal{M}$, $\mathcal{Z}$ and $\{\mathcal{G}_{(i)}^{(n)}\}_{i=1}^{3}$ are introduced, and the equivalent minimization problem is rewritten as

$$\min_{\mathcal{L},\boldsymbol{w},\mathcal{M},\mathcal{Z},\{\mathcal{U}^{(n)}\}_{1:N}} \sum_{n=1}^{N} w_n \sum_{i=1}^{3} \left\|\mathcal{U}_{(i)}^{(n)}\right\|_* + \frac{1}{2}\|\mathcal{Y}-\mathcal{L}\|_F^2$$
$$+\gamma\|\boldsymbol{w}\|_F^2 + \lambda_{TV}\|\mathcal{M}\|_{\mathrm{TV}} + \lambda_{PC}\|\mathcal{Z}\|_{\mathrm{PC}} \tag{8}$$
$$\text{s.t. } \mathcal{L} = \mathfrak{M}\left(\{\mathcal{U}^{(n)}\}_{n=1:N}\right), \mathcal{D}_{\omega}^{-1}(\mathcal{L}) = \mathcal{M}, \mathcal{P}\left(\mathcal{D}_{\omega}^{-1}(\mathcal{L})\right) = \mathcal{Z}$$

The abovementioned problem (8) is divided into two blocks for updating the variables $\mathcal{L}, \boldsymbol{w}, \mathcal{G}, \mathcal{M}, \mathcal{Z}, \{\mathcal{U}^{(n)}\}_{1:N}$. The first block is $\boldsymbol{w}$, which is as follows the problem Eq. 9.

$$\min_{\boldsymbol{w}} \sum_{n=1}^{N} \left(\sum_{i=1}^{3} \left\|\mathcal{U}_{(i)}^{(n)}\right\|_*\right) w_n + \gamma\|\boldsymbol{w}\|_F^2$$
$$s.t.\ w^T\mathbf{1} = 1,\ w_n \geq 0 \tag{9}$$

Then, the second block is the others (such as $\mathcal{L}$ and $\{\mathcal{U}^{(n)}\}_{1:N}$), which is as follows the problem Eq. 10.

$$\min_{\mathcal{L},\mathcal{M},\mathcal{Z},\{\mathcal{U}^{(n)}\}_{1:N}} \sum_{n=1}^{N} w_n \sum_{i=1}^{3} \left\|\mathcal{G}_{(i)}^{(n)}\right\|_* + \frac{1}{2}\|\mathcal{Y}-\mathcal{L}\|_F^2$$
$$+\lambda_{TV}\|\mathcal{M}\|_{\mathrm{TV}} + \lambda_{PC}\|\mathcal{Z}\|_{\mathrm{PC}}$$
$$s.t.\mathcal{L} = \mathfrak{M}\left(\{\mathcal{U}^{(n)}\}_{n=1}^{N}\right), \mathcal{D}_{\omega}^{-1}(\mathcal{L}) = \mathcal{M},$$
$$\mathcal{P}\left(\mathcal{D}_{\omega}^{-1}(\mathcal{L})\right) = \mathcal{Z}\ and\ \mathcal{U}_{(i)}^{(n)} = \mathcal{G}_{(i)}^{(n)} \tag{10}$$

3.2 Optimization for solving the proposed ATRFHS model

3.2.1 Auto-weighted mechanism

Through the problem solver (8), an auto-weighted mechanism can voluntarily balance the importance of different nuclear norms of TR factor matrices. The block

coordinate descent (BCD) optimization framework can optimize the problem Eq. 9. When the variables $\{\mathcal{U}^{(n)}\}_{1:N}$ are fixed, the nuclear norms $\{\sum_{i=1}^{3}\|\mathcal{U}_{(i)}^{(n)}\|_*\}_{n=1}^{N}$ are invariant. Namely, the indicator vector $\boldsymbol{\eta} = [\sum_{i=1}^{3}\|\mathcal{U}_{(i)}^{(1)}\|_*, \sum_{i=1}^{3}\|\mathcal{U}_{(i)}^{(2)}\|_*, ..., \sum_{i=1}^{3}\|\mathcal{U}_{(i)}^{(n)}\|_*]^T$ is fixed. Then, the problem (9) for updating weighting coefficients $\boldsymbol{w}$, automatically weighing the importance of the TR nuclear norm, can be defined as

$$\mathcal{F}(\boldsymbol{w}) = \sum_{k=1}^{N}\eta_k w_k + \gamma\|\boldsymbol{w}\|_F^2 - \mu\left(\boldsymbol{w}^T\mathbf{1} - 1\right) - \boldsymbol{\sigma}^T\boldsymbol{w} \tag{11}$$

where $\mu \geq 0$ and $\boldsymbol{\sigma} = [\sigma_1, \sigma_2 \cdots \sigma_N]^T \geq 0$ are the Lagrangian multipliers. It is a convex Quadratic Programming (QP) problem with equality and non-equality requirements that any QP solver can solve. By taking the derivatives of Eq. 12 to $\boldsymbol{w}$ and setting it as 0, $\partial_w\mathcal{F} = \boldsymbol{\eta} + 2\gamma\boldsymbol{w} - \mu - \boldsymbol{\sigma} = 0$, we can get $w_i = \frac{\mu+\sigma_i-\eta_i}{2\gamma}$, The optimal solution $\boldsymbol{w}$ satisfies the KKT condition. It can be discussed separately in three cases (Chen et al., 2021).

1) if $\eta_i - \mu > 0$, since $\sigma_i > 0$, from the condition $w_i^*\sigma_i = 0$ and $\sigma_i = 0$, then $w_i = \frac{\mu-\eta_i}{2\gamma}$
2) if $\eta_i - \mu = 0$, then $w_i = \frac{\sigma_i}{2\gamma}$. Since $w_i^*\sigma_i = 0$, so $\sigma_i = 0$ and $w_i = 0$ can be inferred
3) if $\eta_i - \mu < 0$ and $\sigma_i > 0$, we can find the positive integer $h = \underset{i}{argmax}(\eta_i - \mu > 0)$ satisfies the nonnegative constraint of w_i

Therefore, the optimal solution to the problem in Eq. 11 is given by

$$w_{+i} = \begin{cases} \dfrac{\mu - \eta_i}{2\gamma} & ,\eta_i > \mu \\ 0 & ,\eta_i \leq \mu \end{cases} \tag{12}$$

Where $\mu = \frac{\sum_{i=1}^{N}\eta_i - 2\gamma}{h}$.

3.2.2 Alternating minimization optimization framework

Problem Eq. 10 is transformed into the following unconstrained augmented Lagrangian function:

$$\begin{aligned}\mathcal{F}(\mathcal{W}) = &\sum_{n=1}^{N}w_n\left(\sum_{i=1}^{3}\|\mathcal{G}_{(i)}^{(n)}\|_* + \langle\mathcal{A}^{(n)}, \mathcal{G}_{(i)}^{(n)} - \mathcal{U}_{(i)}^{(n)}\rangle + \frac{\beta}{2}\|\mathcal{G}_{(i)}^{(n)} - \mathcal{U}_{(i)}^{(n)}\|_F^2\right)\\ &+\frac{1}{2}\left\|\mathcal{Y} - \mathfrak{M}\left(\{\mathcal{U}^{(n)}\}_{n=1}^{N}\right)\right\|_F^2 + \lambda_{TV}\|\mathcal{M}\|_1 + \frac{\beta}{2}\|\mathcal{D}_w^{-1}(\mathcal{L}) - \mathcal{M}\|_F^2 +\\ &\langle\mathcal{B}, \mathcal{D}_w^{-1}(\mathcal{L}) - \mathcal{M}\rangle + \lambda_{PC}\|\mathcal{Z}\|_1 + \frac{\beta}{2}\|\mathcal{P}(\mathcal{D}_w^{-1}(\mathcal{L})) - \mathcal{Z}\|_F^2 + \langle\mathcal{C}, \mathcal{P}(\mathcal{D}_w^{-1}(\mathcal{L})) - \mathcal{Z}\rangle\end{aligned} \tag{13}$$

Where $\mathcal{W} = \{\mathcal{L}, \mathcal{M}, \{\mathcal{G}^{(n)}\}_{n=1}^{N}, \mathcal{Z}, \{\mathcal{U}^{(n)}\}_{n=1}^{N}, \{\mathcal{A}^{(n)}\}_{n=1}^{N}, \mathcal{B}, \mathcal{C}\}$, $\{\mathcal{A}^{(n)}\}_{n=1}^{N}$, $\mathcal{B}$, $\{\mathcal{G}^{(n)}\}_{n=1}^{N}$ and $\mathcal{C}$ are auxiliary variables. Based on the framework of alternating minimization, the updates of $\mathcal{L}$, $\mathcal{M}$, $\{\mathcal{G}^{(n)}\}_{n=1}^{N}$, $\mathcal{Z}$, $\{\mathcal{U}^{(n)}\}_{n=1}^{N}$, $\{\mathcal{A}^{(n)}\}_{n=1}^{N}$, $\mathcal{B}, \mathcal{C}$ are given respectively as follows.

Step 1: Update $\{\mathcal{U}^{(n)}\}_{n=1}^{N}$ and $\mathcal{L}$ with fixing other variables, the $\mathcal{U}^{(n)}$ sub-problem is rewritten as

$$\begin{aligned}\mathcal{F}(\mathcal{U}^{(n)}) = &\sum_{i=1}^{3}\frac{\beta}{2}\left\|\mathcal{G}^{(n,i)} - \mathcal{U}^{(n,i)} + \frac{1}{\beta}\mathcal{A}^{(n)}\right\|_F^2\\ &+\frac{1}{2}\left\|\mathcal{Y} - \mathfrak{M}\left(\{\mathcal{U}^{(n)}\}_{n=1:N}\right)\right\|_F^2\end{aligned}$$

This is a least-squares problem. So for $n = 1, \ldots, N$, $\mathcal{U}^{(n)}$ can be updated by

$$\mathcal{U}_+^{(n)} = \text{fold}_2\left(\frac{\sum_{i=1}^{3}\left(\beta G_{(2)}^{(n,i)} + A_{(2)}^{(n,i)}\right) + T_{<n>}U_{<2>}^{(\neq n)}}{U_{<2>}^{(\neq\neq n)}TU_{<2>}^{(\neq n)} + 3\mathbf{E}}\right) \tag{14}$$

where $\mathbf{E}$ is an identity matrix.

By updated TR factors $\{\mathcal{U}^{(n)}\}_{n=1}^{N}$ for every iteration, then $\mathcal{L}$ is updated as

$$\mathcal{L}_+ = \mathfrak{M}\left(\{\mathcal{U}_+^{(n)}\}_{n=1}^{N}\right) \tag{15}$$

Step 2: Update $\mathcal{G}^{(n)}$ with fixing other variables, by simplifying (13), for $i = 1, 2, 3$, the augmented Lagrangian functions w.r.t. $[\mathcal{G}]$ is expressed as

$$\mathcal{F}(G^{(n)}) = w_n\sum_{i=1}^{3}\|G_{(i)}^{(n)}\|_* + \frac{\beta}{2}\left\|\mathcal{G}^{(n)} - \left(\mathcal{U}^{(n)} - \frac{1}{\beta}\mathcal{A}^{(n)}\right)\right\|_F^2$$

Solving $G_{(i)}^{(n)}$ is a nuclear norm model and has led to a closed form. So for $n = 1, \ldots, N$, $\mathcal{G}^{(n)}$ can be updated by

$$\mathcal{G}_+^{(n)} = \text{fold}_{(i)}\left(\mathbf{S}_{\frac{w_n}{\beta}}\left(\mathcal{U}^{(n)} - \frac{1}{\beta}\mathcal{A}^{(n)}\right)\right) \tag{16}$$

where $\mathbf{S}_{\frac{w_n}{\beta}}$ represents the thresholding SVD operation (Chen et al., 2018).

Step 3: Update $\mathcal{M}$ by fixing other variables. The optimization model can be rewritten as

$$\mathcal{F}(\mathcal{M}) = \lambda_{TV}\|\mathcal{M}\|_1 + \frac{\beta}{2}\left\|\mathcal{M} - \left(\mathcal{D}_w^{-1}(\mathcal{L}) - \mathcal{B}\right)\right\|_F^2 \tag{17}$$

Optimizing (18) can be easily solved by a soft-thresholding operator.

$$\mathcal{M}_+ = \boldsymbol{\Psi}_{\frac{\lambda_{TV}}{\beta}}\left(\mathcal{D}_w^{-1}(\mathcal{L}) - \mathcal{B}\right) \tag{18}$$

where $\boldsymbol{\Psi}_v$ is defined by $\boldsymbol{\Psi}_v(x) = sgn(x).^*max(|x| - v, 0)$.

Step 4: Fixing other variables to update $\mathcal{Z}$, the optimization model can be rewritten as

$$\mathcal{F}(\mathcal{Z}) = \lambda_{PC}\|\mathcal{Z}\|_1 + \frac{\beta}{2}\left\|\mathcal{Z} - \left(\mathcal{P}(\mathcal{D}_w^{-1}(\mathcal{L})) - \mathcal{C}\right)\right\|_F^2 \tag{19}$$

Similarly, the closed-form solution is

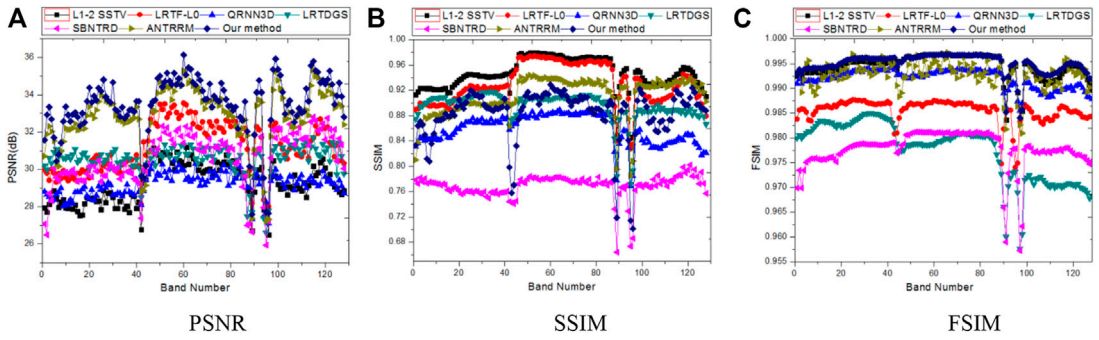


FIGURE 8
The comparative performance of different methods in terms of PSNR, SSIM, and FSIM under Case 2 on WDC Mall. **(A)** PSNR, **(B)** SSIM, **(C)** FSIM.

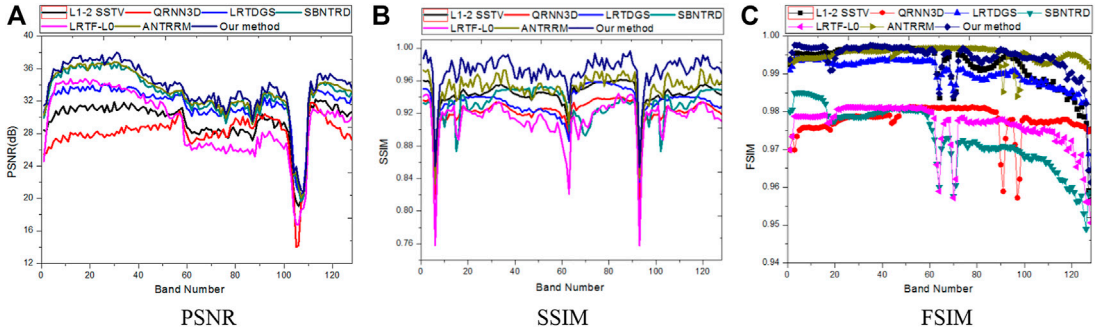


FIGURE 9
The comparative performance of different methods in terms of PSNR, SSIM, and FSIM under Case 4 on *INDIAN PINES*. **(A)** PSNR, **(B)** SSIM, **(C)** FSIM.

TABLE 2 Quantitative comparison of all competing methods on the two GF-5 datasets (time unit: second).

	Datasets	$L_{1\text{-}2}$ SSTV	LRTF-L_0	LRTDGS	SBNTRD	QRNN3D	ANTRRM	OURS
ENL	GF-5 (Shanghai)	84.27	83.98	84.57	85.31	85.37	85.47	**85.98**
EPI		0.9142	0.8934	0.9214	0.9298	0.9276	0.9317	**0.9389**
Time		**102.4**	253.7	534.1	492.7	590.6	357.9	303.4
ENL	GF-5 (Baoqing)	87.45	86.72	85.96	86.74	87.14	87.74	**87.98**
EPI		0.9245	0.9167	0.9047	0.9204	0.9247	0.9278	**0.9327**
Time		**110.7**	196.7	573.1	684.7	610.3	348.7	312.4

The best results for each quality index are shown in bold.

$$\mathcal{Z}_+ = \mathbf{\Psi}_{\frac{\lambda_{PC}}{\beta}}\left(\mathcal{P}\left(\mathcal{D}_w^{-1}(\mathcal{L})\right) - \mathcal{C}\right) \tag{20}$$

$$\begin{cases} \mathcal{A}_{t+1}^{(n)} = \mathcal{A}_t^{(n)} + \beta\left(\mathcal{G}_{t+1}^{(n)} - \mathcal{U}_{t+1}^{(n)}\right) \\ \mathcal{B}_{t+1} = \mathcal{B}_t + \beta\left(\mathcal{M}_{t+1} - \mathcal{D}_w^{-1}(\mathcal{L}_{t+1})\right) \\ \mathcal{C}_{t+1} = \mathcal{C}_t + \beta\left(\mathcal{Z}_{t+1} - \mathcal{P}\left(\mathcal{D}_w^{-1}(\mathcal{L}_{t+1})\right)\right) \end{cases} \tag{21}$$

Step 5: Update $\{\mathcal{A}^{(n)}\}_{n=1}^{N}, \mathcal{B}, \mathcal{C}$: When the $(t+1)$-th iteration begins, the Lagrange multipliers are updated by the following

The specific process of the ADMM-based solver for the ATRFHS HSI reconstruction model and BCD-based solver for auto-weighting is introduced in Algorithm 1.

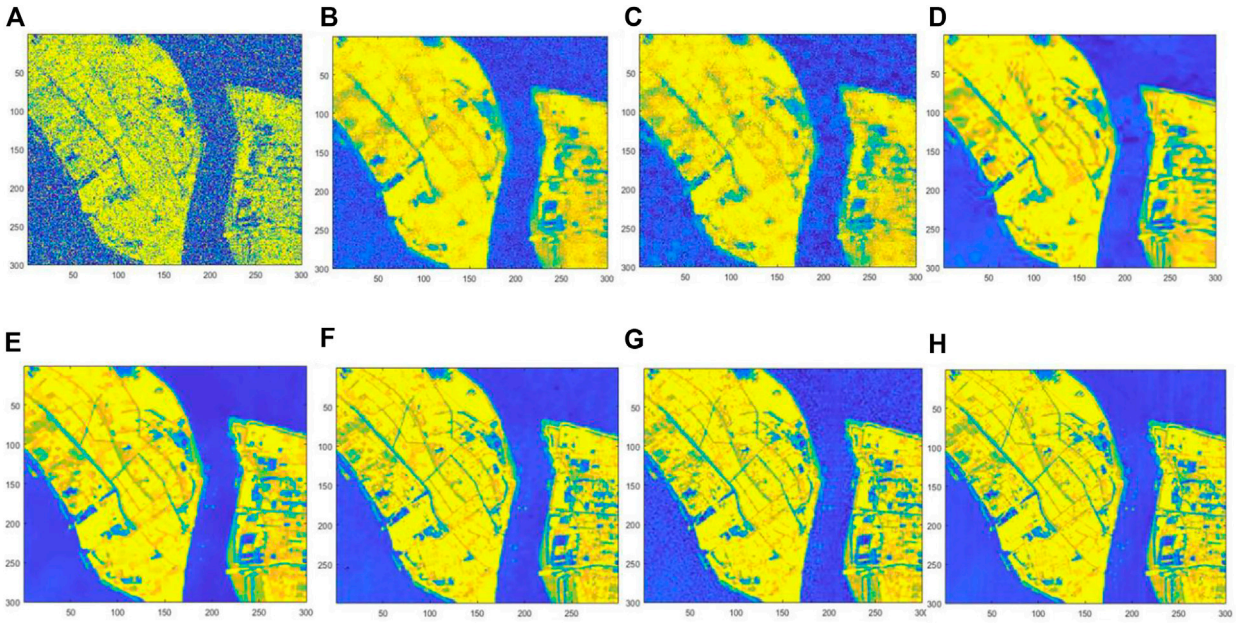


FIGURE 10
Restored results of GF-5 HSI data of Shanghai City: **(A)** Noisy, **(B)** $L_{1\text{-}2}$ SSTV, **(C)** QRNN3D, **(D)** LRTDGS, **(E)** SBNTRD, **(F)** LRTF-L_0, **(G)** ANTRRM, **(H)** OURS.

Input: observed HSI $\mathcal{T} \in \mathbb{R}^{x \times y \times b}$, TR rank $\{R_1, R_2, \ldots R_N\}$, λ_{TV}, λ_{PC}, β
Initialization: $\{\mathcal{A}^{(n)}\}_{n=1}^{N} = 0, \mathcal{M} = \mathcal{Z} = \mathcal{B} = \mathcal{C} = 0$, t=0, $\beta = 0.01$, $\sigma = 1.05$, $\beta_{max} = 5 \times 10^2$, max_it=100, ter=1e-5, $\boldsymbol{w} = \{w_j\}_{j=1}^{N} = \frac{1}{N}$, Grouping nonlocal similar cuboids by NCT to form tensor $\mathcal{Y} = \mathcal{D}_{\omega}(\mathcal{T})$; From $n = 1$ to N, randomly initialize for core tensors $\mathcal{U}^{(n)}$; Compute $\mathcal{L}_{(0)} = \Phi\{\mathcal{U}^{(1)}, \ldots, \mathcal{U}^{(K)}\}$. Calculation MPC by $C(\mathcal{T}(:,:,i))$ *via* (5) i=1, ..., b. While $t \leq$max_it and $||\mathcal{Y} - \mathcal{L}_{\text{last}}||^2 <$ ter Update $\{w_j\}_{j=1}^{N}$ by Eq. 12 Update $\{\mathcal{U}^{(n)}\}_{n=1}^{N}$ by Eq. 14; Update $\mathcal{L}$ by Eq. 15; Update $\{\mathcal{G}^{(n)}\}_{n=1}^{N}$ by Eq. 16; Update $\mathcal{M}$ and $\mathcal{Z}$ by Eqs 18, 20; Update $\{\mathcal{A}^{(n)}\}_{n=1}^{N}$, $\mathcal{B}, \mathcal{C}$ by (21), and the penalty parameter update $\beta = min(\sigma\beta, \beta_{max})$; t=t+1; End Transform $\mathcal{L}$ into a three-order tensor $\mathcal{X} = \mathcal{D}_{\omega}^{-1}(\mathcal{L})$
return: restored HSI $\mathcal{X}$.

Algorithm 1 The whole procedure of the ATRFHS algorithm.

3.3 Computational complexity

The computational complexity of our ATRFHS method is analyzed as follows. For simplicity, we assume to transform HSI data into a high tensor $\mathcal{D} \in \mathbb{R}^{I \times I, \ldots, \times I}$ from by NCT and TR-rank with $R1 = R2 = \cdots = RN = R$. The updating of $\{w_j\}_{j=1}^{N}$, $\{\mathcal{U}^{(n)}\}_{n=1}^{N}$, and $\{\mathcal{G}^{(n)}\}_{n=1}^{N}$ have closed-form solutions, as shown in Algorithm 1. It is obvious to observe that the most time-consuming parts are updating $\{\mathcal{U}^{(n)}\}_{n=1}^{N}$ and SVD operation of $\{\mathcal{G}^{(n)}\}_{n=1}^{N}$. Then the computational complexities of updating $\{\mathcal{U}^{(n)}\}_{n=1}^{N}$ and $\{\mathcal{G}^{(n)}\}_{n=1}^{N}$ are $\mathcal{O}(NI^N R^2)$ and $\mathcal{O}(NI^N R^3)$. The computational complexity of updating weighting $\{w_j\}_{j=1}^{N}$ is $\mathcal{O}(3TN)$. Thus, the overall complexity of our proposed algorithm can be written as $\mathcal{O}(TNI^N R^2(1 + R) + 3TN)$, where T is the number of iterations.

4 Experiments

Two simulated and two real datasets are utilized in the experiments to demonstrate the efficacy of the proposed algorithm with the auto-weight TR rank minimization regular on HSI restoration. Six representative state-of-the-art methods are considered for quantitative and visual comparison; namely, $L_{1\text{-}2}$ SSTV (Zeng et al., 2020) based on 3-D $L_{1\text{-}2}$ spatial-spectral total variation low-rank tensor recovery, LRTF-L_0 (Xiong et al., 2019) based on a spectral-spatial L_0 gradient regularized low-rank tensor factorization, LRTDGS (Chen et al., 2019a) based on weighted group sparsity-regularized low-rank tensor decomposition, SBNTRD (Chen et al., 2020; Oseledets, 2011) based on subspace nonlocal TR decomposition-based method, ANTRRM(Xuegang et al., 2022) based on nonlocal tensor ring rank minimization (Xuegang et al., 2022) and QRNN3D based on 3D Quasi-Recurrent RNN(Wei et al., 2020). All of our experiments are conducted on a Desktop computer with 16 GB of DDR4 RAM

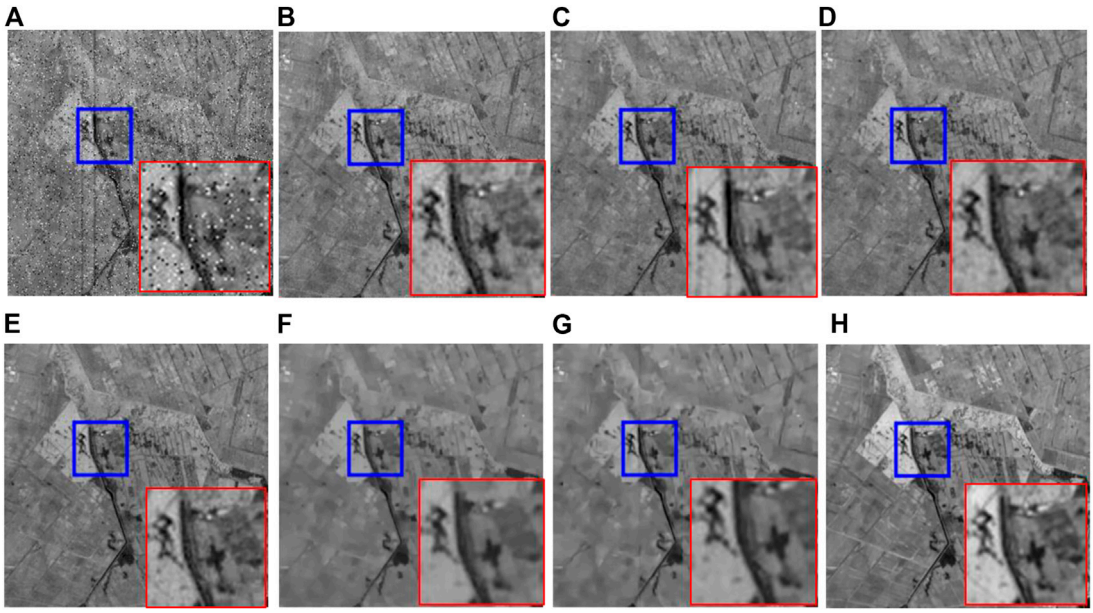


FIGURE 11
Restored results of GF-5 HSI data of Baoqing: **(A)** Noisy, **(B)** $L_{1\text{-}2}$ SSTV, **(C)** QRNN3D, **(D)** LRTDGS, **(E)** SBNTRD, **(F)** LRTF-L_0, **(G)** ANTRRM, **(H)** OURS. Please zoom in for better viewing.

TABLE 3 No-Reference Hyperspectral Image Quality Assessment on the GF-5 Baoqing image.

Approach	$L_{1\text{-}2}$ SSTV	LRTF-L_0	LRTDGS	SBNTRD	QRNN3D	ANTRRM	Ours
Score	17.68	16.85	16.35	17.98	16.98	16.74	15.31

and a 3.2 GHz Intel Core i7-7700K CPU running MATLAB R2018b. All of the competitors' parameters are adjusted following the literature's guidelines.

4.1 Synthetic HSI experiments

Because the ground-truth HSI is provided for the simulated experiments, four quantitative quality indices: peak signal-to-noise ratio (PSNR), structure similarity (SSIM), feature similarity (FSIM), erreur relative global adimensionnelle de synthèse (ERGAS) (Chen et al., 2018) are adopted for validating the performance of the proposed model on two synthetic experiment datasets, namely, the Washington DC Mall and Indian Pines datasets. The MPSNR, MSSIM, and MFSIM, computed by taking the average of all bands, are used to evaluate performance.

The four indices evaluate spatial and spectral information retention, and the PSNR, SSIM, and FSIM values are generated by averaging all bands. The higher the PSNR, SSIM, and FSIM, the lower the ERGAS, and the better the HSI denoising outcome.

1) *The WDC Mall dataset:* The Washington DC Mall dataset was collected by the Hyperspectral Digital Imagery Collection Experiment (HYDICE) with the permission of the Spectral Information Technology Application Center of Virginia. The original size is 1208×307×210. A sub-image of 256×256×128 from this data set is extracted for our experiment.
2) T*he Indian Pines dataset:* The Indian Pines dataset was collected by AVIRIS sensor over the Indian Pines test site in North-western Indiana. It contains 145×145 pixels and 224 spectral reflectance bands with wavelengths ranging from 0.4 to 2.5× 10^{-6} m. The Indian Pines dataset comprises 220 bands with a spatial size of 145×145 pixels. A sub-image of 145×145×128 from this data set is extracted for our experiment.

As for parameter settings, we empirically set the regularization parameter $\lambda_{TV} = 0.02$, $\lambda_{PC} = 0.05, \beta = 0.03$. In NCT, we set s=5, k=7, and p=32. Five different types of noise cases were added to these two clean HSI datasets to simulate

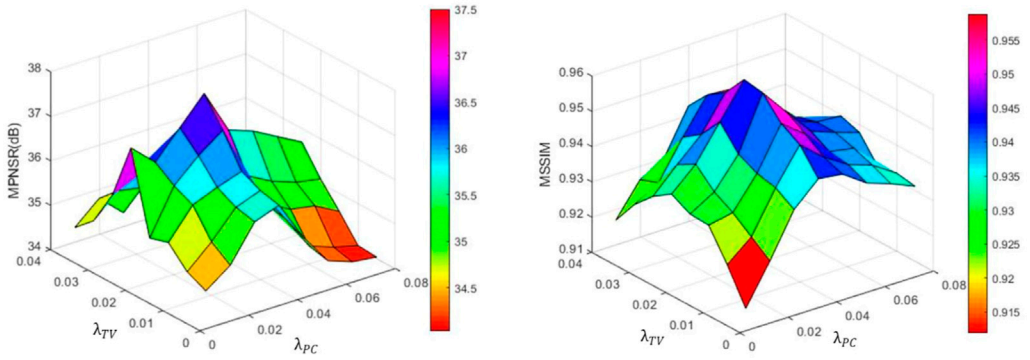


FIGURE 12
Sensitivity analysis of regularization parameter.

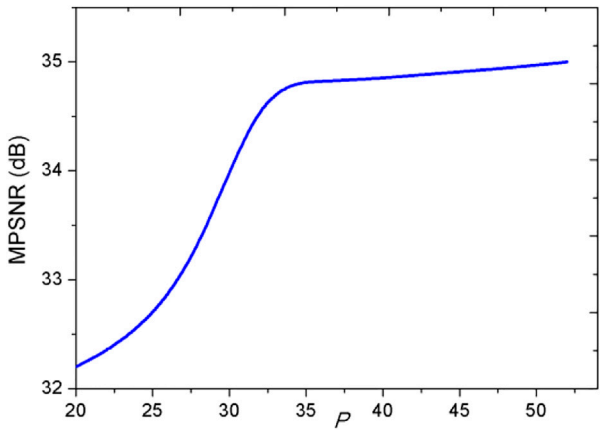


FIGURE 13
Sensitivity analysis of spectral band length *P*

complicated noise cases in a real scene. The following is a detailed description of these cases.

Case 1. (*i.i.d. Gaussian Noise*): Entries in all bands were corrupted by zero-mean i.i.d. Gaussian noise $N(0, \sigma^2)$ with σ=0.05.

Case 2. (*Non-i.i.d. Gaussian Noise*): Entries in all bands were tainted by zero-mean Gaussian noise of different intensities. Each band's signal noise ratio (SNR) is generated by a uniform distribution with a value between (Fan et al., 2017; Yokota et al., 2018)dB.

Case 3. (*Gaussian+Stripe Noise*): Based on Case 2, some stripes randomly selected from 20 to 75 are added from band 10 to band 98 in WDC Mall and Indian Pines datasets.

Case 4. (*Gaussian+Deadline Noise*): Based on Case 2, deadlines are added from band 76 to band 106 in WDC and Indian Pines datasets.

Case 5. (*Gaussian+Impulse Noise*): Based on Case 2, fifty bands in WDC and Indian Pines datasets were randomly chosen to add impulse noise with different intensities, and the percentage of impulse is from 30% to 60%.

Table 1 displays the quantitative results of all comparable approaches in the Washington DC Mall and HYDICE urban data on various cases. The best results for each quality index are shown in bold. From Table 1, it is clear that our proposed approach and SBNTRD obtain the best results over the other compared methods in all cases, confirming our proposed method's advantage over others. It is worth noting that SBNTRD fully exploits the spatial information by nonlocal prior and TR decomposition. Due to the considerations of auto-weight LR properties and efficiently exploiting the structure information of HSI by NCT in our proposed method, the proposed method obtains the best results over the other compared methods except for a small number of indicator cases.

Regarding visual quality, Figure 6 and Figure 7 show the denoised results by seven different methods under Case 5 in the WDC dataset and Case 3 in the Indian Pines dataset, respectively. As shown in the white square from the enlarged red areas of restored images in Figure 6 and Figure 7C, QRNN3D methods can remove noises but fail to retain structure information. Moreover, it is clear to see that low-rank tensor recovery with prior information regularization methods $L_{1\text{-}2}$ SSTV, LRTDGS, SBNTRD LRTF-L_0, and ANTRRM can effectively remove random noise and stripe noise in Figure 5 and Figures 6B,D–G, but the image details cannot be preserved well shown in the enlarged box of Figure 6 and Figure 7. The proposed ATRFHS method, in contrast, can effectively remove all of the mixture noise and preserve more edges and details, as shown in Figure 6 and Figure7H. Because ATRFHS not only considers the more

TABLE 4 Empirical analysis of each regularizer of the ATRFHS model.

Datasets	Index	ATRFHS	No-PC	No-TV	No-TV-PC
WDC mall	MPSNR (dB)	31.89	31.26	31.11	30.37
	MSSIM	0.9217	0.9104	0.9198	0.9031
	MFSIM	0.9674	0.9201	0.9314	0.9178
Indian pines	MPSNR (dB)	34.58	34.14	34.01	33.41
	MSSIM	0.9784	0.9587	0.9647	0.9247
	MFSIM	0.9814	0.9714	0.9624	0.9431

reasonable LR with auto-weight TR rank minimization for Gaussian noise and random noise in the HSI restoration task but the deadlines and stripe noise can be removed shown in Figure 6 and Figure7H by exploring high-order tensors structure, as a higher-order tensor makes it more efficient to exploit the local structures in transformed tensor. The our proposed approach outperforms all the evaluated methods in terms of four quantitative quality indices, eliminating all of the hybrid noise while keeping the detailed edges and texture information in the restored HSI. We further calculate the PSNR, SSIM, and FSIM values of different bands in all simulated data cases and show the curves of evaluation indices.

Figure 8 and Figure 9 show the curves of PSNR and SSIM evaluation indices of each band on WDC Mall under Case 2 and INDIAN PINES under Case 4, respectively. As displayed in Figure 8A and Figure 9A, it is observed that the proposed method performs higher PSNR values than other methods for almost all bands in WDC Mall data and INDIAN PINES. For SSIM indices, the proposed method can outperform other methods in most bands, as demonstrated in Figure 8B and Figure 9B. From Figure 8C and Figure 9C, it can be seen that the proposed ATRFHS method achieves higher FSIM values than other methods in almost all bands, which verifies the robustness of the proposed method using the auto-weighted strategy of low-rank approximation and also demonstrates the superiority of the hybrid regularization compared with others. Our proposed method has obtained the best restoration performance among all competing methods, as evidenced by the distribution of evaluating index of the restoration image in Figures 8 and Figure 9.

In conclusion, the proposed method outperforms the other methods in terms of visual quality and quantitative indices.

4.2 Real data experiments

The two GF-5 real-world hyperspectral data sets acquired by the GaoFen-5 satellite: Shanghai City and Baoqing (available URL: http://hipag.whu.edu.cn/resourcesdownload.html), were used in the real HSI data experiments. GaoFen-5 satellite was developed by the Chinese Aerospace Science and Technology Corporation and launched in 2018. The original size of the GF-5 dataset is 2100 × 2048 × 180, and 25 bands are miss information. This dataset is seriously degraded by the mixture of Gaussian, stripes, and deadlines noises.

The selected GF-5 Shanghai City image is 307 × 307 pixels in size and has 210 bands. The GF-5 Baoqing sub-image has a size of 300×300×305, with some abnormal bands removed. Both GF-5 images are extensively polluted by various stripes, including wide stripe noise that emerges at the same position on the continuous bands as dense stripe noise of varying widths. Furthermore, several of the bands have much-mixed noise. Before denoising, the gray values of authentic HSIs were band-by-band normalized to [0, 1]. After removing the miss bands and extracting a small region, a sub-HSI with the size of 300× 300×156 is chosen for experiments.

Both Equivalent Number of Looks (ENL) (Anfinsen et al., 2009)and Edge Preserving Index (EPI) (Sattar et al., 1997) were employed for performance evaluation. The larger the ENL and EPI values, the better the quality of the restored images.

The quantitative assessment indices ENL and EPI values and the running time of all competing methods are provided in Table 2 on the two GF-5 datasets. The best outcomes for each quality indicator are highlighted in bold. From the table, it is clear that our proposed approach achieves a significantly improved performance in both the ENL and EPI indexes, as compared with other competing methods. Because high-dimension tensor decomposition can capture the global correlation in the spatial-spectral dimensions, ATRFHS obtained better results than the other tensor-based format methods by combining auto-weighted low-rank tensor ring decomposition with total variation and phase congruency regularization. Meanwhile, the effectiveness of the suggested auto-weight TR nuclear standard is shown.

It can be observed from Table 2 that the $L_{1\text{-}2}$ SSTV method is the fastest method among all the compared methods. However, as the previous experimental work demonstrated, it cannot achieve good repair outcomes. Due to the use of updating $\mathcal{U}$ and SVD operation of $\mathcal{G}$ for higher-order data computation, the computational cost of the proposed ATRFHS is relatively higher than $L_{1\text{-}2}$ SSTV, QRNN3D including traning phase and LRTF-L_0 methods but significantly lower than other methods, namely, LRTDGS, ANTRRM and SBNTRD.

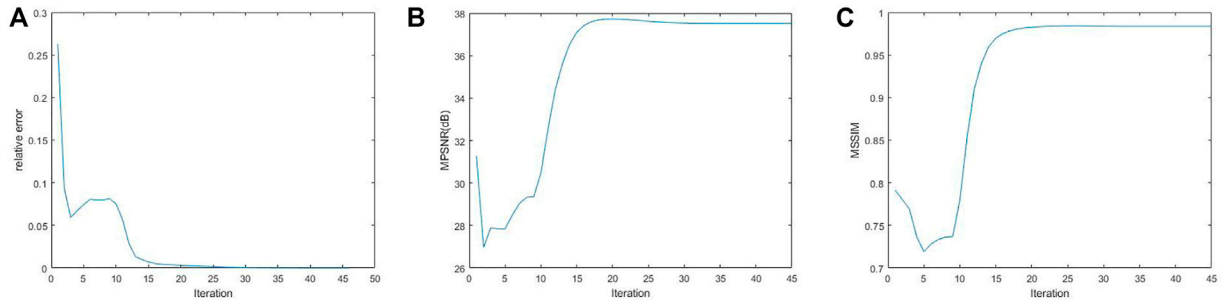


FIGURE 14
Convergence analysis of the algorithm in terms of **(A)** relative error, **(B)** the MPSNR values, and **(C)** the MSSIM values.

TABLE 5 Washington DC Mall-Classification accuracies obtained by different restoration approaches before using RF.

Index	HSI	$L_{1\text{-}2}$ SSTV	LRTF-L_0	LRTDGS	SBNTRD	QRNN3D	Ours
OA	67.54	78.69	80.47	86.74	85.36	86.39	90.58
AA	69.50	84.58	86.78	81.25	84.25	84.12	95.14

The restorations of band 96 in Shanghai City of GF-5 are presented in Figure 10. To clearly illustrate the visualization of the restoration results, a demarcated area in the subfigure is enlarged in the bottom right corner. Figure 10A shows that the image suffers from a mixture of Gaussian and sparse noise. It is straightforward to observe that $L_{1\text{-}2}$ SSTV, QRNN3D, and LRTDGS cannot efficiently maintain edge information to a certain extent. The approaches based on the low-rank prior perform more effectively than other competing methods, as seen in Figure 10. By combining the total variation and phase congruency into a unified TV regularization and utilizing the auto-weighted low-rank tensor ring decomposition to encode the global structure correlation, our proposed ATRFHS method can better remove the complex mixed noise. In particular, compared to other competing methods, our proposed method preserves the most significant detail edge, texture information, and image fidelity.

Figure 11 displays the restoration results of band 109 in Baoqing data of GF-5. From Figure 11A, one can see that the image is wholly contaminated by various noises, including Gaussian, random noise, and heavy structure noise, including stripes and deadlines. After denoising using the different HSI restoration methods, the noise is removed. As shown in Figures 11C,D, the QRNN3D and LRTDGS methods cannot eliminate the stripes in the results, as observed in the enlarged box on the image.

The $L_{1\text{-}2}$ SSTV and SBNTRD can obtain a better visual result than the other methods, but some intrinsic information such as the local smoothness underlying the HSI cube, was not exploited, as shown in Figures 11B,E,F. LRTF-L_0 and the proposed method can remove much noise compared to the TV mentioned above, but LRTF-L_0 does not preserve edges and local detail information, as well as our proposed ATRFHS method. More detailed visual comparison results can be seen in such red boxes. To summarize, the proposed ATRFHS can still achieve the best performance for removing such heavy mixed noise from this dataset.

To further investigate the effect of our method, we provide a no-reference image quality assessment, as presented in (Yang et al., 2017), to evaluate the real-world hyperspectral data before and after denoising. The quality scores are presented in Table 3. A lower no-reference image quality assessment score indicates better denoising quality. The table shows that our proposed ATRFHS method has the lowest score, demonstrating ATRFHS's superiority.

4.3 The impact of parameters

Three parameters in Eq. 9 need to be discussed, including two regularization parameters λ_{TV} and λ_{PC}, and the penalty parameter β .

1) The impact of parameters λ_{TV} , λ_{PC} and β:

TV and PC multichannel images have been widely exploited for their edge-preserving characteristics. To prevent the overfitting of the sharper edge of our proposed approach from influencing the experimental results, we present the MPSNR and MSSIM values

achieved by a function of λ_{TV} and λ_{PC} for the WDC dataset in Case 1 as an example to identify the best parameter values. Figure 12 shows the change in MPSNR values for the proposed algorithm for these two regularization parameters λ_{TV} and λ_{PC}. It is evident that when λ_{TV} is equal to 0.02 and λ_{PC} is set to 0.05, and the proposed method can reach the peak of MPSNR.

2) The impact of parameters spectral band length P:

Furthermore, P also is an important parameter for taking advantage of the spectral local low-rankness properties. As shown from Figure 13 in the simulated WDC data experiments, when P is equal to 32, the MPSNR value tends to be stable. Thus, we suggest the use of p=32.

4.4 effectiveness of hybrid smoothness regularization terms

The proposed ATRFHS is a tensor ring-based method combining TV and PC priors. To verify the effectiveness of the two priors in our model, we further compare our approach with a simplified version of our model without the TV and PC regularization terms, that is, set the parameters $\lambda_{TV} = 0$ and $\lambda_{PC} = 0$ in our model (9). The test is conducted on two simulated datasets by the MPSNR, MFSIM, and MSSIM evaluation indices in Case 3 with a mixture noise. Experimental results are shown in Table 4. ATRFHS is our proposed method, and No-PC is a method using only TV prior without PC prior by $\lambda_{PC} = 0$, No-TV is a method using only PC prior without TV prior by $\lambda_{TV} = 0$ and No-TV-PC is the original weighted tensor ring-based method by $\lambda_{TV} = 0$ and $\lambda_{PC} = 0$. The metric scores listed in Table 3 obtained by ATRFHS are the highest among all the techniques. Hybrid smoothness regularization with TV and PC priors is more suitable for recovering HSIs with more texture information than pure TV methods. The performance of the ATRFHS method demonstrates the effectiveness of hybrid smoothness regularization terms.

4.5 Empirical analysis for convergence of the ATRFHS solver

The convergence behavior of the proposed algorithm is discussed. We present an empirical analysis of the proposed restoration approach convergence on the simulated WDC Mall data set. We offer a numerical experiment to show the convergence behavior in terms of relative error, the MPSNR values, and the MSSIM values. In Figure 14, we can observe that the curves of all assessment indexes come to a stable value when the algorithm reaches a relatively high iteration number, indicating that the proposed algorithm empirically converges well.

4.6 Classification application

In this sub-section, we examine the impact of HSI noise removal procedures as a preprocessing step for HSI classification. We employed Random Forest (RF) classifier (Athey et al., 2019) to make a comparison of the effectiveness of different restoration approaches. The main idea of the RF classifier is to classify an input vector by running down each decision tree in the forest. Each tree outcomes in a unit vote for a specific class, and the forest selects the final classification label based on the most votes. Classification accuracy is utilized to evaluate the effectiveness of different restoration approaches. Two metrics have been applied: Overall Accuracy (OA) and Average accuracy (AA). The percentage value of AA and OA is shown in Table 5. The metrics AA and OA are reported in percentage. Table 5 shows that denoising approaches improve the performance of the subsequent classification technique compared to directly using the raw data after the denoising procedure. The proposed ATRFHS approach achieves the highest OA and AA values among all the classification results achieved by the seven restoration approaches, indicating the best performance in HSI restoration.

5 Conclusion

This article presents an auto-weighted low-rank Tensor Ring Factorization with Hybrid Smoothness regularization (ATRFHS) for HSI restoration. The global spatial structure correlation of HSI was efficiently depicted by the low-rank factorization of TR, which can embody the advantages of both rank approximations and high-dimension structures. An auto-weighted measure of factors rank minimization of TR factorization can more accurately approximate the TR rank and better promote the low-rankness of the solution. Moreover, we employed a hybrid regularization incorporating total variation and phase congruency to smooth the factor and preserve HSI's spatial piecewise constant structure. A well-known alternating minimization framework was developed to solve the ATRFHS model efficiently. Both simulated and real-world datasets were used to demonstrate the performance and superiority of the proposed methods over state-of-the-art HSI denoising methods. In the future, we will try to incorporate more appropriate regularization and nonconvex tensor ring factor rank minimization into our tensor ring model to enhance its HSI restoration capability further.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: http://hipag.whu.edu.cn/resourcesdownload.html.

Author contributions

The project was suggested by JW. JL and XL did the analysis. Data and comments contributed by XL and DL. The article was written by XL, JL, and BW.

Funding

This work was supported in part by the Science and Technology Support Project of Panzhihua City of Sichuan Province of China under Grant 2021CY-S-6, in part by Sichuan Philosophy and Social Science Key Research Base Project under Grant ZGJS 2022-07, in part by Scientific research Project of Panzhihua University under Grant 2020ZD015, in part by Sichuan vanadium and Titanium Materials EngineeringTechnology Research Center Open project under Grant 2022FTGC03 and in part by Sichuan Meteorological Disaster Forecasting and Early Warning and Emergency Management Research center project under Grant 2022FTGC03 ZHYJ21-YB07.

Acknowledgments

The authors thank the Zhang Yanhong Project Group, which provided GaoFen-5 datasets for experiments.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Aggarwal, H. K., and Majumdar, A. (2016). Hyperspectral image denoising using spatio-spectral total variation. *IEEE Geosci. Remote Sens. Lett.* 13 (3), 1–5. doi:10.1109/lgrs.2016.2518218

Anfinsen, S. N., Doulgeris, A. P., and Eltoft, T. (2009). Estimation of the equivalent number of looks in polarimetric synthetic aperture radar imagery. *IEEE Trans. Geosci. Remote Sens.* 47 (11), 3795–3809. doi:10.1109/tgrs.2009.2019269

Athey, S., Tibshirani, J., and Wager, S. (2019). Generalized random forests. *Ann. Stat.* 47 (2), 1148–1178. doi:10.1214/18-aos1709

Bioucas-Dias, J. M., Plaza, A., Camps-Valls, G., Scheunders, P., Nasrabadi, N., and Chanussot, J. (2013). Hyperspectral remote sensing data analysis and future challenges. *IEEE Geosci. Remote Sens. Mag.* 1 (2), 6–36. doi:10.1109/mgrs.2013.2244672

Chen, C., Wu, Z. B., Chen, Z. T., Zheng, Z. B., and Zhang, X. J. (2021). Auto-weighted robust low-rank tensor completion via tensor-train. *Inf. Sci.* 567, 100–115. doi:10.1016/j.ins.2021.03.025

Chen, Y., Guo, Y., Wang, Y., Wang, D., Peng, C., and He, G. (2017). Denoising of hyperspectral images using nonconvex low rank matrix approximation. *IEEE Trans. Geosci. Remote Sens.* 55 (9), 5366–5380. doi:10.1109/tgrs.2017.2706326

Chen, Y., He, W., Yokoya, N., and Huang, T. Z. (2019). Hyperspectral image restoration using weighted group sparsity-regularized low-rank tensor decomposition. *IEEE Trans. Cybern.* 50 (8), 3556–3570. doi:10.1109/tcyb.2019.2936042

Chen, Y., He, W., Yokoya, N., Huang, T. Z., and Zhao, X. L. (2019). Nonlocal tensor-ring decomposition for hyperspectral image denoising. *IEEE Trans. Geosci. Remote Sens.* 58 (2), 1348–1362. doi:10.1109/tgrs.2019.2946050

Chen, Y., He, W., Zhao, X. -L., Huang, T. -Z., Zeng, J., and Lin, H. (2022). Exploring nonlocal group sparsity under transform learning for hyperspectral image denoising. *IEEE Trans. Geosci. Remote Sens.* 60 (5537518), 1–18. doi:10.1109/TGRS.2022.3202359(

Chen, Y., Huang, T. -Z., He, W., Zhao, X. -L., Zhang, H., and Zeng, J. (2022). Hyperspectral image denoising using factor group sparsity-regularized nonconvex low-rank approximation. *IEEE Trans. Geosci. Remote Sens.* 60 (5515916), 1–16. doi:10.1109/TGRS.2021.3110769

Chen, Y., Huang, T. Z., He, W., Yokoya, N., and Zhao, X. L. (2020). Hyperspectral image compressive sensing reconstruction using subspace-based nonlocal tensor ring decomposition. *IEEE Trans. Image Process.* 29, 6813–6828. doi:10.1109/tip.2020.2994411

Chen, Y., Wang, S., and Zhou, Y. (2018). Tensor nuclear norm-based low-rank approximation with total variation regularization. *IEEE J. Sel. Top. Signal Process.* 12 (6), 1364–1377. doi:10.1109/jstsp.2018.2873148

Dabov, K., Foi, A., Katkovnik, V., and Egiazarian, K. (2007). Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Trans. Image Process.* 16 (8), 2080–2095. doi:10.1109/tip.2007.901238

Dian, R., Li, S., and Fang, L. (2019). Learning a low tensor-train rank representation for hyperspectral image super-resolution. *IEEE Trans. Neural Netw. Learn. Syst.* 30 (9), 2672–2683. doi:10.1109/tnnls.2018.2885616

Dian, R., and Li, S. (2019). Hyperspectral image super-resolution via subspace-based low tensor multi-rank regularization. *IEEE Trans. Image Process.* 28 (10), 5135–5146. doi:10.1109/tip.2019.2916734

Ding, M., Huang, T. Z., Zhao, X. L., and Ma, T. H. (2022). Tensor completion via nonconvex tensor ring rank minimization with guaranteed convergence. *Signal Process.* 194, 108425. doi:10.1016/j.sigpro.2021.108425

Fan, H., Chen, Y., Guo, Y., Zhang, H., and Kuang, G. (2017). Hyperspectral image restoration using low-rank tensor recovery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 10 (10), 4589–4604. doi:10.1109/jstars.2017.2714338

He, W., Yao, Q., Li, C., Yokoya, N., Zhao, Q., Zhang, H., et al. (2022). Non-local meets global: An iterative paradigm for hyperspectral image restoration. *IEEE Trans. Pattern Anal. Mach. Intell.* 44 (4), 2089–2107. doi:10.1109/TPAMI.2020.3027563

He, W., Yokoya, N., Yuan, L., and Zhao, Q. (2019). Remote sensing image reconstruction using tensor ring completion and total variation. *IEEE Trans. Geosci. Remote Sens.* 57 (11), 8998–9009. doi:10.1109/tgrs.2019.2924017

Huang, H., Liu, Y., Liu, J., and Zhu, C. (2020). Provable tensor ring completion. *Signal Process.* 171, 107486. doi:10.1016/j.sigpro.2020.107486

Huang, Z., Li, S., Fang, L., Li, H., and Benediktsson, J. A. (2017). Hyperspectral image denoising with group sparse and low-rank tensor decomposition. *IEEE Access* 6, 1380–1390. doi:10.1109/access.2017.2778947

Liu, X., Bourennane, S., and Fossati, C. (2012). Denoising of hyperspectral images using the PARAFAC model and statistical performance analysis. *IEEE Trans. Geosci. Remote Sens.* 50 (10), 3717–3724. doi:10.1109/tgrs.2012.2187063

Liu, Y., Shang, F., Jiao, L., Cheng, J., and Cheng, H. (2014). Trace norm regularized CANDECOMP/PARAFAC decomposition with missing data. *IEEE Trans. Cybern.* 45 (11), 2437–2448. doi:10.1109/tcyb.2014.2374695

Luo, X. G., Wang, H. J., and Wang, S. (2015). Monogenic signal theory based feature similarity index for image quality assessment. *AEU - Int. J. Electron. Commun.* 69 (1), 75–81. doi:10.1016/j.aeue.2014.07.015

Morrone, M. C., and Owens, R. A. (1987). Feature detection from local energy. *Pattern Recognit. Lett.* 6 (5), 303–313. doi:10.1016/0167-8655(87)90013-4

Oseledets, I. V. (2011). Tensor-train decomposition. *SIAM J. Sci. Comput.* 33 (5), 2295–2317. doi:10.1137/090752286

Sattar, F., Floreby, L., Salomonsson, G., and Lovstrom, B. (1997). Image enhancement based on a nonlinear multiscale method. *IEEE Trans. Image Process.* 6 (6), 888–895. doi:10.1109/83.585239

Wang, M., Wang, Q., Chanussot, J., and Hong, D. (2021). Total variation regularized weighted tensor ring decomposition for missing data recovery in high-dimensional optical remote sensing images. *IEEE Geosci. Remote Sens. Lett.* 19, 1–5. doi:10.1109/lgrs.2021.3069895

Wang, W., Sun, Y., Eriksson, B., Wang, W., and Aggarwal, V. (2018). "Wide compression: Tensor ring nets," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 9329–9338.

Wang, Y., Peng, J., Zhao, Q., Leung, Y., Zhao, X. L., and Meng, D. (2017). Hyperspectral image restoration via total variation regularized low-rank tensor decomposition. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 11 (4), 1227–1243. doi:10.1109/jstars.2017.2779539

Wei, K., Fu, Y., and Huang, H. (2020). 3-D quasi-recurrent neural network for hyperspectral image denoising. *IEEE Trans. Neural Netw. Learn. Syst.* 32 (1), 363–375. doi:10.1109/tnnls.2020.2978756

Wu, Z., Wang, Q., Jin, J., and Shen, Y. (2017). Structure tensor total variation-regularized weighted nuclear norm minimization for hyperspectral image mixed denoising. *Signal Process.* 131, 202–219. doi:10.1016/j.sigpro.2016.07.031

Xiong, F., Zhou, J., and Qian, Y. (2019). Hyperspectral restoration via L_0 gradient regularized low-rank tensor factorization. *IEEE Trans. Geosci. Remote Sens.* 57 (12), 10410–10425. doi:10.1109/tgrs.2019.2935150

Xue, J., Zhao, Y., Bu, Y., Chan, J. C. W., and Kong, S. G. (2022). When laplacian scale mixture meets three-layer transform: A parametric tensor sparsity for tensor completion. *IEEE Trans. Cybern.*, 1–15. doi:10.1109/TCYB.2021.3140148

Xue, J., Zhao, Y., Huang, S., Liao, W., Chan, J. C. W., and Kong, S. G. (2021). Multilayer sparsity-based tensor decomposition for low-rank tensor completion. *IEEE Trans. Neural Netw. Learn. Syst.* 33, 6916–6930. doi:10.1109/TNNLS.2021.3083931

Xue, J., Zhao, Y., Liao, W., Chan, J. C. W., and Kong, S. G. (2019). Enhanced sparsity prior model for low-rank tensor completion. *IEEE Trans. Neural Netw. Learn. Syst.* 31 (11), 4567–4581. doi:10.1109/tnnls.2019.2956153

Xue, J., Zhao, Y., Liao, W., and Chan, J. C. W. (2019). Nonlocal low-rank regularized tensor decomposition for hyperspectral image denoising. *IEEE Trans. Geosci. Remote Sens.* 57 (7), 5174–5189. doi:10.1109/tgrs.2019.2897316

Xuegang, L., Lv, J., and Wang, J. (2022). Hyperspectral image restoration via auto-weighted nonlocal tensor ring rank minimization. *IEEE Geosci. Remote Sens. Lett.* 19, 1–5. doi:10.1109/LGRS.2022.3199820

Yang, J., Zhao, Y. Q., Yi, C., and Chan, J. C. W. (2017). No-reference hyperspectral image quality assessment via quality-sensitive features learning. *Remote Sens.* 9 (4), 305. doi:10.3390/rs9040305

Yokota, T., Erem, B., Guler, S., Warfield, S. K., and Hontani, H. (2018). "Missing slice recovery for tensors using a low-rank model in embedded space," in Proceedings of the IEEE conference on computer vision and pattern recognition, 8251–8259.

Yokota, T., Zhao, Q., and Cichocki, A. (2016). Smooth PARAFAC decomposition for tensor completion. *IEEE Trans. Signal Process.* 64 (20), 5423–5436. doi:10.1109/tsp.2016.2586759

Yu, J., Li, C., Zhao, Q., and Zhao, G. (2019). "Tensor-ring nuclear norm minimization and application for visual: Data completion," in ICASSP 2019-2019 IEEE international conference on acoustics, speech and signal processing (ICASSP), 3142–3146.

Yuan, L., Li, C., Cao, J., and Zhao, Q. (2020). Rank minimization on tensor ring: An efficient approach for tensor decomposition and completion. *Mach. Learn.* 109 (3), 603–622. doi:10.1007/s10994-019-05846-7

Yuan, L., Zhao, Q., Gui, L., and Cao, J. (2019). High-order tensor completion via gradient-based optimization under tensor train format. *Signal Process. Image Commun.* 73, 53–61. doi:10.1016/j.image.2018.11.012

Zeng, H., Xie, X., Cui, H., Yin, H., and Ning, J. (2020). Hyperspectral image restoration via global $L_{1\text{-}2}$ spatial-spectral total variation regularized local low-rank tensor recovery. *IEEE Trans. Geosci. Remote Sens.* 59 (4), 3309–3325. doi:10.1109/tgrs.2020.3007945

Zhang, H., He, W., Zhang, L., Shen, H., and Yuan, Q. (2013). Hyperspectral image restoration using low-rank matrix recovery. *IEEE Trans. geoscience remote Sens.* 52 (8), 4729–4743.

Zhang, Z., Ely, G., Aeron, S., Hao, N., and Kilmer, M. (2014). "Novel methods for multilinear data completion and denoising based on tensor-SVD," in Proceedings of the IEEE conference on computer vision and pattern recognition, 3842–3849.

Zhao, Q., Zhou, G., Xie, S., Zhang, L., and Cichocki, A. (2016). Tensor ring decomposition. arXiv preprint arXiv:1606.05535.

frontiers | Frontiers in Earth Science

TYPE Original Research
PUBLISHED 11 January 2023
DOI 10.3389/feart.2022.1025635

OPEN ACCESS

EDITED BY
Xiaolan Qiu,
Aerospace Information Research Institute (CAS), China

REVIEWED BY
Zhiqi Guo,
Jilin University, China
Jiachun You,
Chengdu University of Technology, China

*CORRESPONDENCE
Gen-Yang Tang,
tanggenyang@163.com

SPECIALTY SECTION
This article was submitted to Environmental Informatics and Remote Sensing, a section of the journal Frontiers in Earth Science

RECEIVED 23 August 2022
ACCEPTED 20 October 2022
PUBLISHED 11 January 2023

CITATION
Feng G, Zeng H-H, Xu X-R, Tang G-Y and Wang Y-X (2023), Shear wave velocity prediction based on deep neural network and theoretical rock physics modeling.
Front. Earth Sci. 10:1025635.
doi: 10.3389/feart.2022.1025635

Shear wave velocity prediction based on deep neural network and theoretical rock physics modeling

Gang Feng[1], Hua-Hui Zeng[1], Xing-Rong Xu[1], Gen-Yang Tang[2]* and Yan-Xiang Wang[1]

[1]Research Institute of Petroleum Exploration and Development-Northwest (NWGI), PetroChina, Lanzhou, China, [2]China University of Petroleum, Beijing, China

Shear wave velocity plays an important role in both reservoir prediction and pre-stack inversion. However, the current deep learning-based shear wave velocity prediction methods have certain limitations, including lack of training dataset, poor model generalization, and poor physical interpretability. In this study, the theoretical rock physics models are introduced into the construction of the labeled dataset for deep learning algorithms, and a forward simulation of the theoretical rock physics models is utilized to supplement the dataset that incorporates geological and geophysical knowledge. This markedly increases the physical interpretability of the deep learning algorithm. Theoretical rock physics models for two different types of reservoirs, i.e., conventional sandstone and tight sandstone reservoirs, are first established. Then, a full-sample labeled dataset is constructed using these two types of theoretical rock physics models to traverse the elasticity parameter space of the two types of reservoirs through random variation and combination of parameters in the theoretical models. Finally, based on the constructed full-sample labeled dataset, four parameters (P-wave velocity, clay content, porosity, and density) that are highly correlated with the shear wave velocity are selected and combined with a deep neural network to build a deep shear wave velocity prediction network with good generalization and robustness, which can be directly applied to field data. The errors between the predicted shear wave velocity using the deep neural network and the measured shear wave velocity data in the laboratory and the logging data in three real field work areas are less than 5%, which are much smaller than the errors predicted by both Han's and Castagna's empirical formula. Furthermore, the prediction accuracy and generalization performance are better than those of these two common empirical formulas. The forward simulation based on theoretical models supplements the training dataset and provides high-quality labels for machine learning. This can considerably improve the interpretability and generalization of models in real applications of a machine learning algorithm.

KEYWORDS

deep neural network, rock physics modeling, theoretical rock physics model, full-sample labeled dataset, shear wave velocity prediction, empirical formula

1 Introduction

Shear wave (S-wave) velocity plays an important role in reservoir prediction (Du, 2014). S-wave velocity data are required for pre-stack inversion and pre-stack attribute analysis. However, in real field work areas, especially in old wells, the high cost of acquiring shear wave velocity data leads to lack of S-wave velocity data (Bagheripour et al., 2015). Therefore, predicting the S-wave velocity in wells where these values has not been measured is essential. The conventional S-wave velocity prediction methods can be divided into three categories, empirical formula method, theoretical rock physics model method and machine learning prediction method.

The empirical formula method utilizes the existing logging data from the target area to statistically analyze the relationship between these data and the S-wave velocity. The formula is generally obtained by fitting data point pairs based on some kind of mathematical expression. There is no need to have a complete theoretical derivation process, and this method is only applicable to specific geological environments (Castagna et al., 1985; Han et al., 1986; Eberhart-Phillips et al., 1989; Ameen et al., 2009). The rock physics model prediction method is to establish the relationship between elastic parameters and reservoir parameters based on theoretical models. Therefore, the S-wave velocity prediction is often more accurate than the empirical formula (Gassmann, 1951; Biot, 1956; Xu and White, 1995, 1996; Xu and Payne, 2009; Sun et al., 2012). Theoretically, the rock physical model is not specifically limited to a particular region, but there is a lot of noise in the real field data, and the predicted results have great uncertainty. In addition, the application of the rock physics model to predict S-wave velocity needs to consider the influence of skeleton composition, fluid distribution and pore shape, which make the application of the rock physics model to predict shear wave velocity difficult since there parameters are not easily accessible.

Neural networks have great advantages in dealing with nonlinear problems, and S-wave velocity prediction is a typical nonlinear problem. In recent years, S-wave velocity prediction using well log data and back-propagation neural network (BPNN) has been widely applied in practical field areas (Eskandari et al., 2004; Alimoradi et al., 2011; Maleki et al., 2014). Each hidden layer of the recurrent neural networks (RNNs) has a feedback to a previous layer, and the subsequent behavior can be shaped by the response of the previous layer. Thus, RNNs are well suited for processing sequential data, and since logging data are connected in-depth, RNNs and their variants long short-term memory (LSTM) networks and gated recurrent units (GRU) networks have been introduced into the S-wave velocity prediction (Mehrgini et al., 2017; Zhang et al., 2020) and other rock parameters (Yuan et al., 2022). Moreover, convolutional neural networks (CNNs) have tremendous advantages in feature extraction, thus the CNNs were widely developed and applied in many research fields (Yuan et al., 2018; Hu et al., 2020; Hu et al., 2021), and a combination of RNNs and CNNs for S-wave velocity prediction has been proposed recently (Wang et al., 2022; Zhang et al., 2022). However, the neural network-based S-wave velocity prediction method has poor generalization and limited labels for establishing S-wave velocity prediction networks, which brings many difficulties to real applications.

To overcome these limitations, we combine theoretical rock physics models and deep neural networks (DNNs) for S-wave velocity prediction. Synthetic datasets can be used when building labeled datasets, if the synthetic datasets are sufficiently complicated, that is, if the most important factors are considered when generating the datasets, the trained network may be able to process realistic datasets directly (Wu et al., 2019; Yu and Ma, 2021; Gao et al., 2022). Therefore, a rich and complete labeled dataset is first constructed using the theoretical rock physics models, and then a deep S-wave velocity prediction network is established using the DNN and the data, such as the P-wave velocity and porosity in the full-sample labeled dataset. Instead of using the data of a certain area to train the neural network, the data generated by the theoretical rock physics models are used for the training, and then the established network is directly applied to the real target work area for S-wave velocity prediction.

2 Theoretical rock physics modeling for multi-type reservoirs

The rock physics model can link the elastic parameters to physical parameters, fluid and lithology (Guo et al., 2022), and specific theoretical rock physic model needs to be established for different types of reservoirs due to different composition, texture and pore microstructure of the reservoir rocks.

2.1 Theoretical rock physics modeling of conventional sandstone reservoir

The porosity and permeability of the conventional sandstone are quite high with relatively simple pore geometry, so the conventional sandstone reservoir is high-quality reservoir. In this study, the Voigt-Ruess-Hill (VRH; Hill, 1952) model is used to calculate the moduli of the rock matrix, and then the Kuster-Toksöz model (Kuster and Toksőz, 1974) is utilized to add stiff and compliant pores to the rock matrix to calculate the moduli of the dry skeleton, which are expressed as follows:

$$(K^{*}_{KT}- K_m)\frac{K_m + 4\mu_m/3}{K^{*}_{KT} + 4\mu_m/3} = \sum_{i=1}^{N} x_i (K_i - K_m)P^{mi} \quad (1)$$

$$(\mu^{*}_{KT} - \mu_m)\frac{\mu_m + \zeta_m}{\mu^{*}_{KT} + \zeta_m} = \sum_{i=1}^{N} x_i (\mu_i - \mu_m)Q^{mi} \quad (2)$$

TABLE 1 Sampling range of conventional sandstone model parameters.

Parameter	Lower bound	Upper bound	Description
Fluid saturation (oil,gas,water)	0	1	the sum of the three is 1
Feldspar, calcite content	0	0.1	
Clay content	0	1	with the sum of other three minerals is 1
Quartz content	0	1	with the sum of other three minerals is 1
Porosity	0	0.3	greater than compliant porosity
Compliant porosity	0.0001	0.01	less than porosity
Compliant porosity aspect ratio	0.0001	0.001	

TABLE 2 Sampling range of tight sandstone model parameters.

Parameter	Lower bound	Upper bound	Description
Fluid saturation (gas,water)	0	1	the sum of the two is 1
Feldspar, calcite content	0	0.1	
Clay content	0	0.4	
Quartz content	0.4	1	with the sum of other three minerals is 1
Porosity	0	0.1	greater than compliant porosity
compliant porosity	0.0001	0.01	less than porosity
compliant porosity aspect ratio	0.0001	0.001	

where x_i denotes the volume concentration for eachinclusion; K_m and μ_m are the bulk modulus and shear modulus of the rock matrix; K^*_{KT} and μ^*_{KT} are the bulk modulus and shear modulus of the dry skeleton, The coefficients P^{mi} and Q^{mi} describe the effect of the inclusion material i in the background medium m.

Then the Wood model is utilized to calculate the bulk moduli of the mixed fluid, and finally, the Gassmann equation (Gassmann, 1951) is used to calculate the saturated rock moduli.

$$\frac{K_{sat}}{K_{sat}-K_{\mathrm{m}}}=\frac{K^*_{KT}}{K^*_{KT}-K_m}+\frac{K_{fl}}{\phi\left(K_{fl}-K_m\right)} \tag{3}$$

$$\mu_{sat}=\mu^*_{KT} \tag{4}$$

where K_{sat} and μ_{sat} are the bulk modulus and shear modulus of the saturated rock.

2.2 Theoretical rock physics modeling of tight sandstone reservoir

For tight sandstone reservoirs, the heterogeneity, microscopic pore structure and pore fluid distribution of rocks are quite complex (Guo et al., 2021). When saturated with different fluids, the fluid flow caused by wave propagation makes the overall elastic responses of rocks more complex. For tight sandstone reservoirs, firstly, the moduli of the rock skeleton are also calculated using the VRH model and the Kuster-Toksöz model, and then the squirt flow effect is considered to account for the velocity dispersion and attenuation (White, 1975; Dvorkin et al., 1995). A simple squirt flow model (Gurevich et al., 2010) can be used to characterize the wave-induced flow effects occurring at microscopic scales in tight sandstones. The idea of a simple squirt flow model is to modify the dry skeleton of the rock as if the compliant pores are saturated with fluid and the stiff pores remain dry, which are expressed as follows:

$$\frac{1}{K_{mf}(P,\omega)}=\frac{1}{K_h}+\frac{1}{\frac{1}{\frac{1}{K_{dry}(P)}-\frac{1}{K_h}}+\left(\frac{1}{K^*_f(P,\omega)}-\frac{1}{K_{\mathrm{m}}}\right)\phi_c(P)} \tag{5}$$

$$\frac{1}{\mu_{mf}(P,\omega)}=\frac{1}{\mu_{dry}(P)}-\frac{4}{15}\left(\frac{1}{K_{dry}(P)}-\frac{1}{K_{mf}(P,\omega)}\right) \tag{6}$$

where K_{mf} is the bulk modulus of the modified skeleton at different frequency and pressures. μ_{mf} is the shear modulus of the corresponding modified skeleton. K_h is the bulk modulus of dry rock under high effective pressure, which can be estimated with the Kuster-Toksöz model, K_{dry} and μ_{dry} are the bulk modulus and shear modulus in the dry condition, K_m is the bulk modulus of the rock matrix, ϕ_c is the compliant porosity and K^*_f is the modified fluid bulk modulus.

After obtaining the modified dry skeleton moduli, the saturated rock elastic moduli are calculated by the Gassmann fluid substitution equations (Han et al., 2021) as follows:

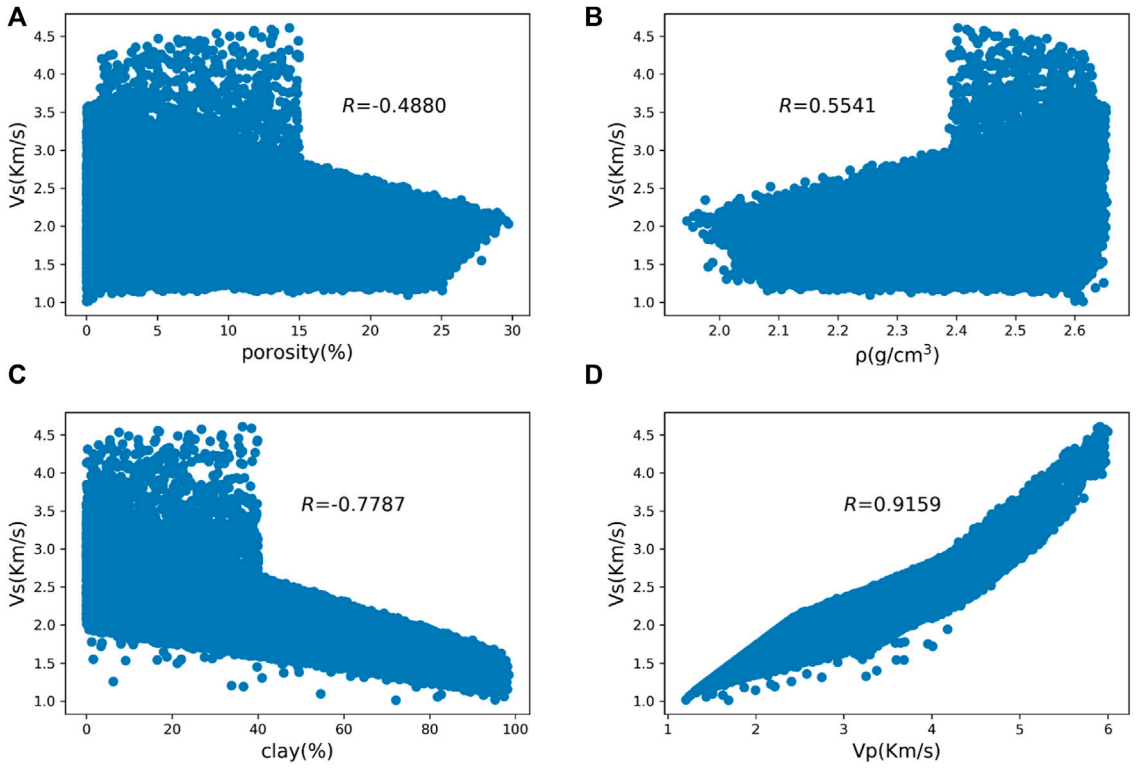


FIGURE 1
Correlation between reservoir parameters and S-wave velocity. **(A)** Porosity versus S-wave velocity. **(B)** Density versus S-wave velocity. **(C)** Clay content versus S-wave velocity. **(D)** P-wave velocity versus S-wave velocity.

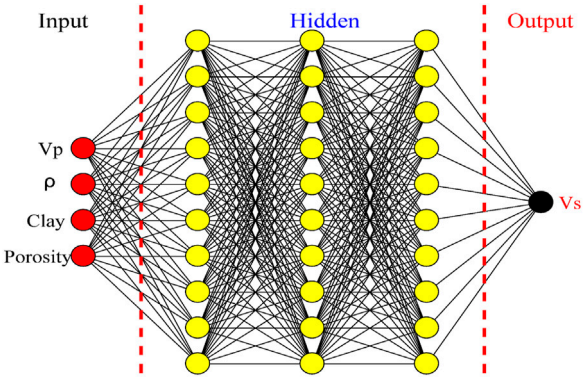


FIGURE 2
The structure of the deep S-wave velocity prediction network.

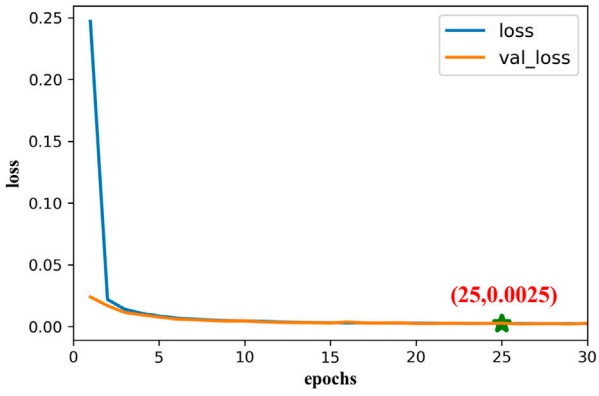


FIGURE 3
Learning curve of neural network.

$$\frac{K_{sat}}{K_{sat}-K_{\mathrm{m}}}=\frac{K_{mf}}{K_{mf}-K_{m}}+\frac{K_{fl}}{\phi_{s}\left(K_{fl}-K_{m}\right)} \tag{7}$$

$$\mu_{sat}=\mu_{mf} \tag{8}$$

where K_{sat} and μ_{sat} are the bulk modulus and shear modulus of the saturated rock.

According to the established theoretical rock physics models, the bulk moduli, the shear moduli can be obtained, and the

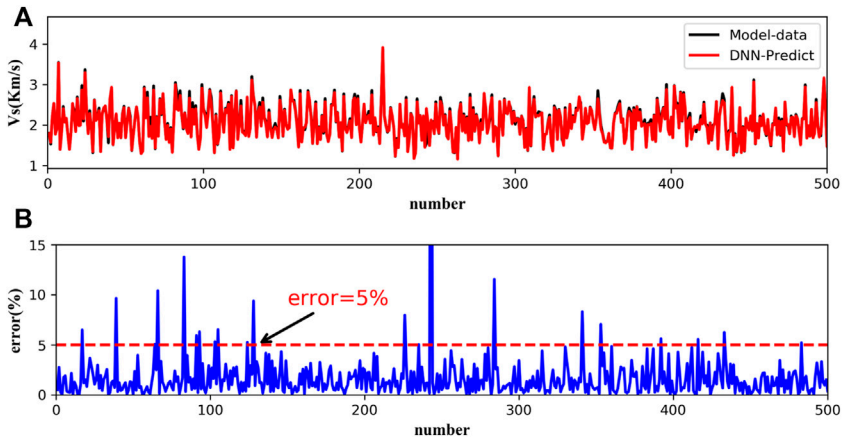


FIGURE 4
Test results of 500 synthetic data. **(A)** Test results. **(B)** Relative errors.

P-wave velocity and S-wave velocity are calculated using the relationship between moduli and density (Eqs 9, 10) as follows:

$$V_p = \sqrt{(K_{sat} + 4\mu_{sat}/3)/\rho} \tag{9}$$

$$V_s = \sqrt{\mu_{sat}/\rho} \tag{10}$$

where K_{sat} and μ_{sat} are the bulk modulus and shear modulus of the saturated rock.

3 S-wave velocity prediction method

In this study, a combination of DNNs and rock physics model is used for S-wave velocity prediction.

3.1 Data preparation

Two types of theoretical rock physics models from the previous section are used to generate 128,000 synthetic data. To ensure the generality and richness of the synthetic data, the sampling ranges of these model parameters cover all possible values, as shown in Tables 1, 2, and the sampling range of the parameters is determined from the real field areas and experimental measurements. Random values in the parameter's sampling space and random combinations of different parameters are used to obtain corresponding S-wave velocity dataset.

Since real field data normally contain noise from the data acquisition, processing and interpretation procedures, we add 10% Gaussian noise to the synthetic data to construct a full-sample labeled dataset that mimics the real data, which helps enhance the robustness of the neural network.

3.2 Feature parameter selection

The reservoir parameters reflect the characteristics of the reservoir, and there is a certain connection between them and the S-wave velocity. Since the trained neural network is to be directly applied to the real field work area, four reservoir parameters, such as porosity, density, clay content and P-wave velocity, which are easily accessible in real field areas, are selected. The correlation between the four parameters and the S-wave velocity is as follows (see Figure 1), where R is the correlation coefficient.

According to the correlation analysis in Figure 1, it can be found that these four parameters have a good correlation with the S-wave velocity. The P-wave velocity and density are positively correlated with the S-wave velocity, and the porosity and clay content are negatively correlated with the S-wave velocity, among which the P-wave velocity has the strongest correlation with the S-wave velocity, and the absolute values of the four correlation coefficients are greater than 0.4. Thus, these four parameters will be used as the input features of the S-wave velocity prediction network.

3.3 Feature parameter normalization

Feature normalization is an important step in deep learning. Since different features always have different amplitudes, units, and ranges, the features with high magnitudes will impose higher impact on networks. If the data is not processed to the same range, the network may not converge when it is trained, and the training time is long, giving more weight to features with larger values, which will limit the prediction accuracy of the regression equation. In order to eliminate this effect, it is necessary to normalize the

TABLE 3 petrophysical parameters of five tight sandstone samples.

Sample	Density (g/cm^3)	Porosity (percent)	Permeability (md)	Clay (percent)	Quartz (percent)
S1	2.65	2.37	0.014	3	41
S2	2.64	3.77	0.023	5	40
S3	2.50	6.48	0.023	5.7	58.8
S4	2.47	6.71	0.069	5.5	68.6
S5	2.41	7.22	0.131	4.7	65.2

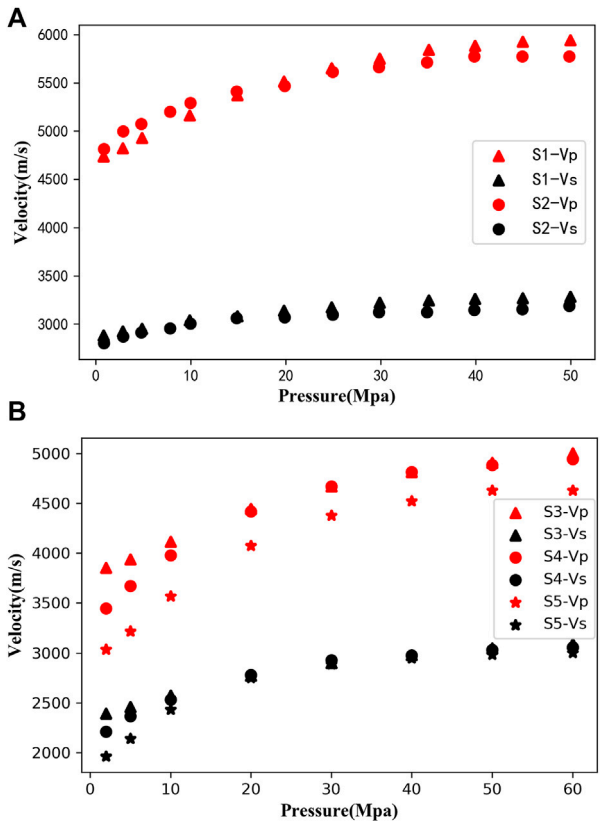


FIGURE 5
P- and S-wave velocities of samples from published literatures. **(A)** Han et al. (2021). **(B)** Li et al. (2018).

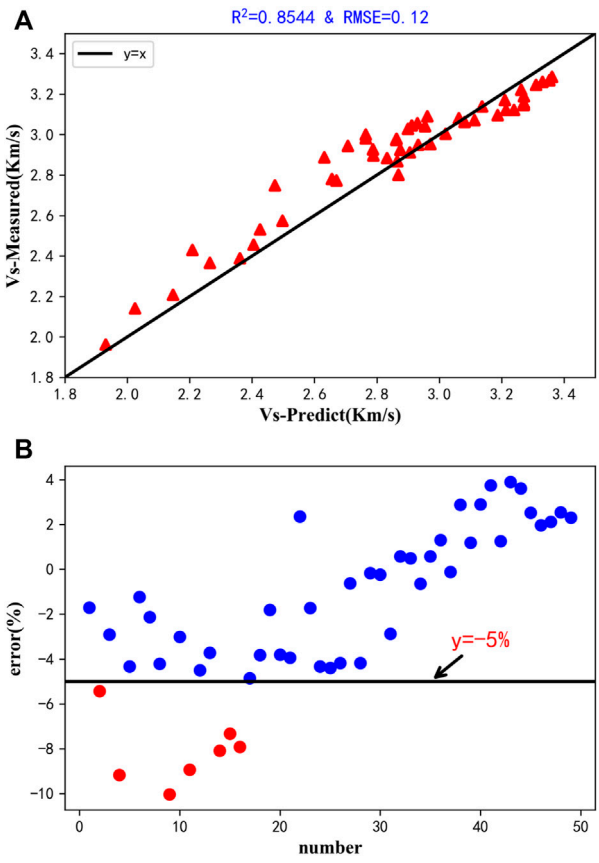


FIGURE 6
Comparison of S-wave velocity measured in laboratory with predicted value of DNN. **(A)** Prediction results. **(B)** Relative errors.

features to the same scale. This study uses the min-max method to normalize the features, and the normalized data are all between 0 and 1, the expressions are as follows:

$$X_{\text{norm}} = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} \tag{11}$$

where X_{norm} is the normalized value, X is the original value, X_{max} is the maximum value of the features, X_{min} is the minimum value of the features.

3.4 Deep shear wave velocity prediction network building and training

In this study, a fully connected neural network with three hidden layers (the number of hidden layer neurons is 10), four inputs and one output is constructed using the P-wave velocity, density, porosity and clay content as input features and the S-wave velocity as the label (see Figure 2). The neural

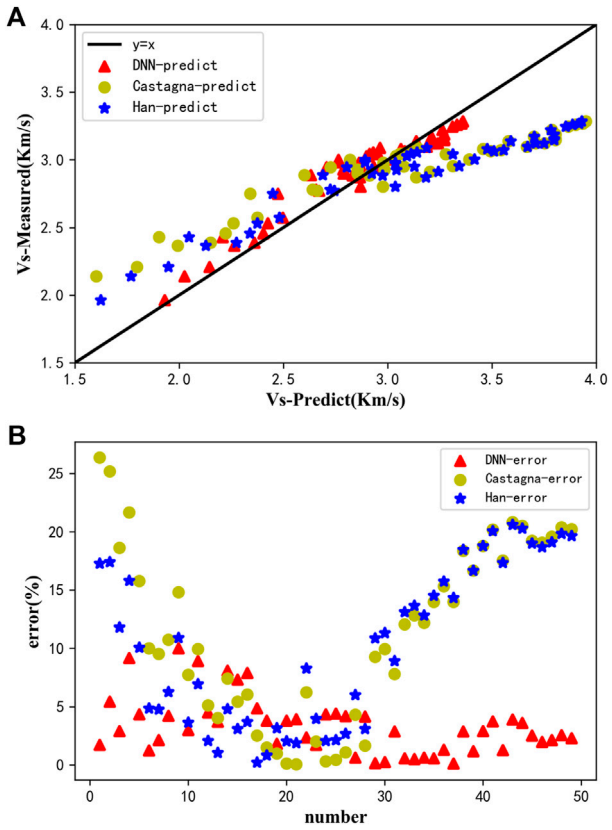


FIGURE 7
Comparison of three S-wave velocity prediction methods. **(A)** Prediction results. **(B)** Relative errors.

network minimizes the mean square error between the label and the output by back propagation using the gradient descent method. The activation function chosen for this network is the Rectified linear unit (ReLU) function to increase the nonlinear characterization ability of the neural network, the optimization algorithm is Adam, and the loss function is the mean square error (MSE) function.

We use the full-sample labeled dataset constructed in Section 3.1 to train the deep shear velocity prediction network. Firstly, all the input features are normalized by the max-min method so that all the features fall between 0 and 1. Then 80% of the labeled dataset is used for training, 10% for validation and 10% for testing. As shown in Figure 3, the training and validation losses decrease simultaneously and converge to relatively low values after 14 epochs of training, which means that the neural network has been fitted well. The validation loss reaches a global minimum after 25 epochs and is as low as the training loss, indicating that the network has been completely fitted and the trained neural network can be generalized to new data for S-wave velocity prediction.

For the trained network, the synthetic data with noise were first tested, and the test results are shown in Figure 4. From Figure 4A, it can be seen that the predicted S-wave velocity using the neural network can match the real S-wave velocity well both in terms of variation trend and values, except for individual data points where the error can reach more than 10%, the error in all other data points is below 5% (see Figure 4B).

4 Deep S-wave velocity prediction network application

In this section we present the results of applying the neural network to laboratory data and real field data.

4.1 Application to the laboratory data

We obtained the clay content, porosity, density, P-wave velocity and S-wave velocity (dry conditions) of five tight

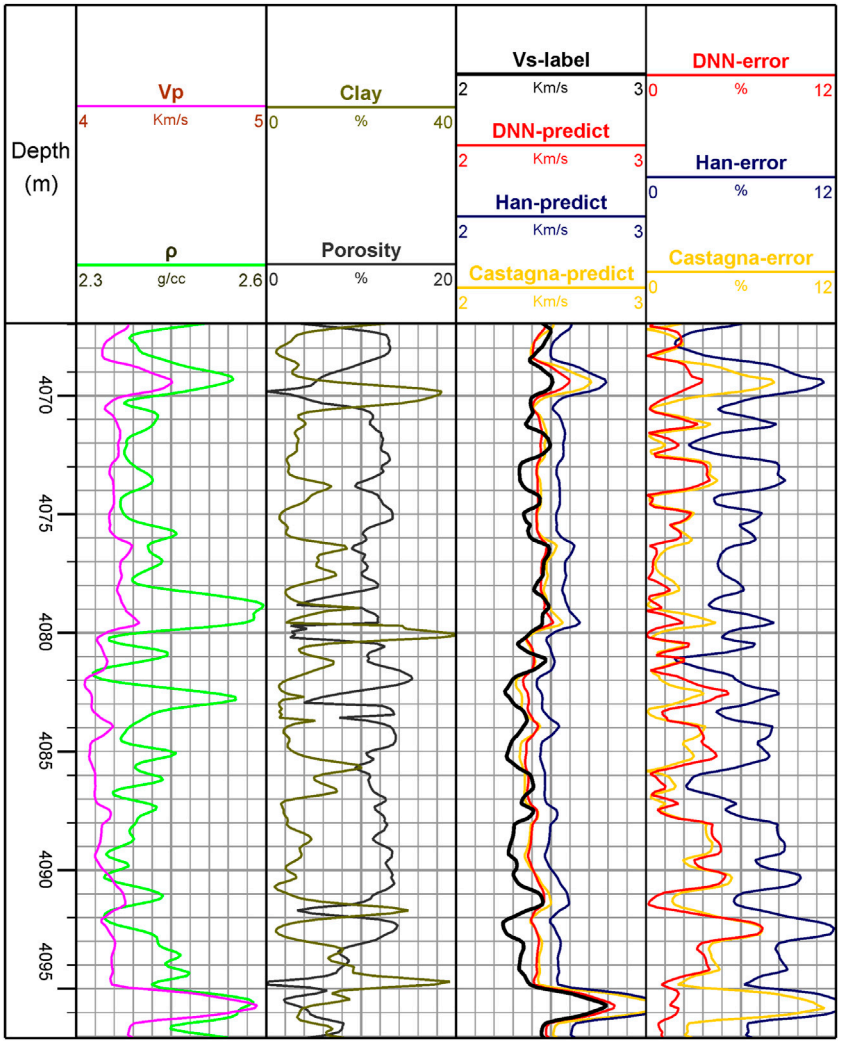


FIGURE 8
Data information of well 1 and S-wave velocity prediction results (Vs-label is the logging S-wave velocity, xx-predict indicates the result predicted using a certain method, and xx-error indicates the corresponding relative error).

sandstone samples from the published literature (Li et al., 2018; Han et al., 2021). The physical parameters of the five sandstone samples are shown in Table 3, and the P- and S-wave velocities are shown in Figure 5. Four features data from the published literature are introduced into the previously trained network and then output the predicted S-wave velocities.

As shown in Figure 6A, the data are concentrated around the line y = x. The coefficient of certainty R2 of the prediction results is above 0.85, and the root mean square error (RMSE) is 0.12, indicating that the predicted S-wave velocities of the neural network are in strong agreement with the laboratory measurements. Most of the relative errors between the predicted results and experimental measurements are within 5% (see Figure 6B), which also indicates that the constructed deep S-wave velocity prediction network has a very good prediction performance, while the large deviation of individual points may be due to some errors generated by the experimental measurement process, resulting in low or high measured values.

To illustrate the superiority of the constructed deep S-wave velocity prediction network, the predicted results of the network were compared with those predicted by the empirical formula proposed by Han et al. (1986), Eq. 12 and by Castagna et al. (1985), Eq. 13.

$$V_s = 0.794V_p - 0.787 \quad (12)$$

$$V_s = 0.862V_p - 1.172 \quad (13)$$

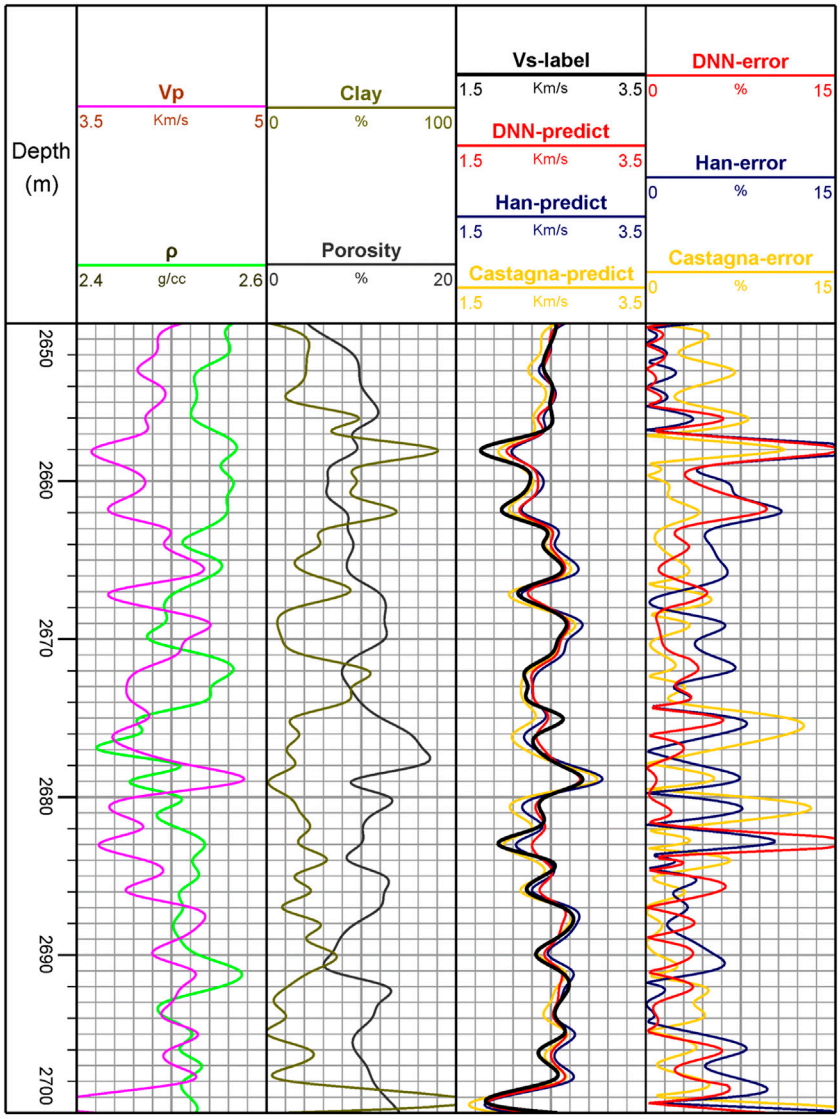


FIGURE 9
Data information of well 2 and S-wave velocity prediction results (Vs-label is the logging S-wave velocity, xx-predict indicates the result predicted using a certain method, and xx-error indicates the corresponding relative error).

The comparison results and the average relative errors are shown in Figure 7. Figure 7A shows the two-dimensional intersection of the S-wave velocities predicted by the three methods and the laboratory measured S-wave velocities, and Figure 7B shows the relative errors. From Figure 7A, we can see that the intersection analysis results of the predicted and true values of the three methods are all distributed around the line y = x, which indicates that the prediction results have certain accuracy. However, the value predicted by the deep S-wave velocity prediction network is closer to the line y = x, which indicates that the network is the most accurate among the three methods. The errors between the predicted and true values of the network are the smallest, as can be seen in Figure 7B, which also confirm this conclusion. Also, it can be found from the figure that Han's empirical formula is more applicable than Castagna's empirical formula at the ultrasonic frequency band in the laboratory.

For tight sandstone, the P-wave velocity increases with increasing water saturation under high pressure conditions, and the S-wave velocity basically does not change, while both P-wave velocity and S-wave velocity increase with increasing water saturation under low pressure conditions (Li et al., 2018).

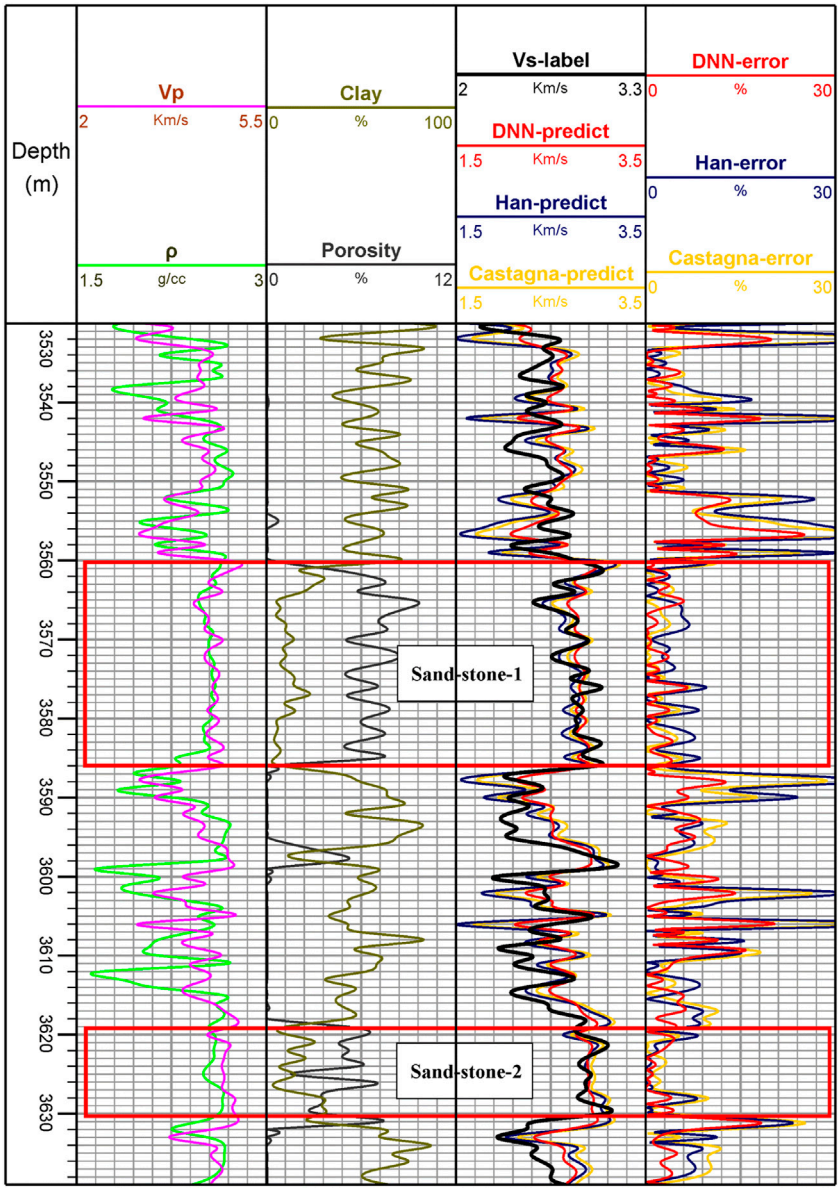


FIGURE 10
Data information of well 3 and S-wave velocity prediction results (Vs-label is the logging S-wave velocity, xx-predict indicates the result predicted using a certain method, and xx-error indicates the corresponding relative error).

TABLE 4 Average relative error of the three S-wave velocity prediction methods in different wells.

Well	DNN-average relative error (%)	Han-average relative error (%)	Castagna-average relative error (%)
Well 1	2.24	6.68	2.86
Well 2	3.33	4.48	3.73
Well 3	4.62	8.24	7.33
Sandstone1&2 of well 3	2.56	4.24	3.47

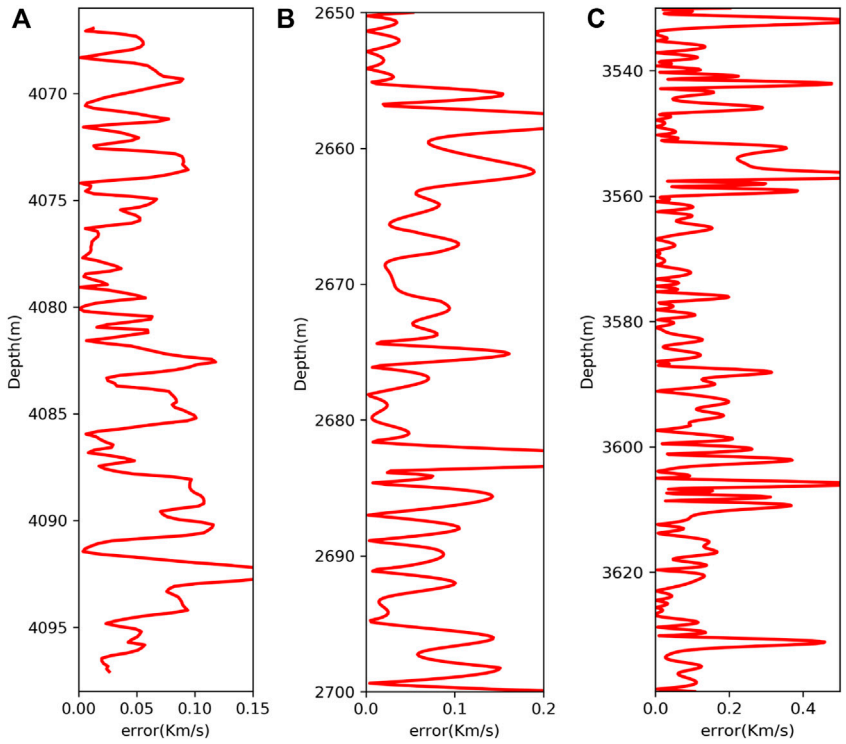


FIGURE 11
Absolute error curve of S-wave velocity predicted by DNN. **(A)** Well 1. **(B)** Well 2. **(C)** Well 3.

Han's empirical formula is obtained by fitting the saturated water sandstone data, while the data obtained in this study are measured under dry conditions, therefore, Han's empirical formula results in a large prediction of the S-wave velocity under high pressure conditions (Figure 7A), while the prediction results of the S-wave velocity under low pressure conditions are very close to the true values.

4.2 Application to the field well log data

The trained neural network was applied to the well log data from three real field areas for S-wave velocity prediction, where well 1 and well 2 were tight sandstone reservoirs and conventional sandstone reservoirs, while well 3 included both sandstone reservoirs and mudstone layers, and the results predicted by the deep S-wave velocity prediction neural network were compared with the real well log data and the results predicted by empirical formulas. The prediction results are shown in Figures 8–10, which show the logging data used in the prediction task as well as the prediction results and relative errors for the three methods. In Figures 8–10, Vs-label indicates real log shear wave velocity (black line), DNN-predict indicates the DNN prediction result (red line), Han-predict indicates the prediction result of Han's empirical formula (blue line), Castagna-predict indicates the prediction result of Castagna's empirical formula (yellow line), and Table 4 shows the average relative errors of the prediction results.

From Figures 8, 9 and Table 4, we can see that for both Well one and Well 2, the S-wave velocity prediction results of the deep neural network and the real log data have the same general trend and small error, which has a good match. Compared with the other two prediction methods, the error of the DNN prediction results is smaller (2.24%, 3.33%) and the trend is closer to the real S-wave velocity, which indicates that the established deep S-wave velocity prediction network has good application in the real field work areas.

For well 3, the deep neural network prediction results are relatively poor, but from Figure 10 and Table 4, we can see that the deep neural network still performs well in the two sandstone reservoir sections (2.56%), while the prediction results of the mudstone section deviate greatly from the true values. This is owing to the fact that the rock physic responses of mudstone are different from that of both tight sandstone and conventional sandstone, and the labeled dataset constructed by the sandstone model is less applicable to mudstone, so there exists a large prediction error. In addition, the prediction accuracy of the deep neural network

is also superior to the two empirical formulations, the same as the results of wells 1 and 2, because this high clay content is taken into account in the sandstone modeling process. Figures 8–10 and Table 4 also show that Castagna's empirical formula is more applicable than Han's empirical formula at the well-logging frequency band.

Figure 11 shows the absolute error (the absolute value of the difference between the measured and predicted S-wave velocities) of the S-wave velocities predicted based on the deep neural network for the three wells. As can be seen from the figure, the absolute error is basically below 0.15 km/s for well 1, and below 0.2 km/s for well 2 as well as the sandstone section of well 3, which indicates that the practicality of the method proposed in this study is fairly good.

From the application of the laboratory data and the well log data, the prediction accuracy of the deep S-wave velocity prediction network established in this study is higher than the common empirical formulas. Han's empirical formula is more applicable to the ultrasonic frequency band, while Castagna's proposed empirical formula is more applicable to the well-logging frequency band, which may be because Han's empirical formula is statistically based on the data at ultrasonic frequency band, while Castagna's empirical formula is based on the well-logging data. Compared with the two empirical formulations, the deep S-wave velocity prediction network proposed in this study is applicable to the full frequency band S-wave velocity prediction and has better generalization.

5 Conclusion

In this study, we proposed a shear wave velocity prediction method based on DNNs and rock physics modeling. We have applied the established deep S-wave velocity prediction network to real field data directly. Theoretical rock physics models are developed for the properties of conventional and tight sandstone reservoirs. The geological and geophysical knowledge is incorporated into the data set of the deep neural network by means of forwarding simulation of the theoretical rock physics models to construct a full-sample labeled dataset that traverses the entire S-wave velocity space. A robust and generalizable deep S-wave velocity prediction network without multiple training is built by combining the full-sample labeled dataset and the deep neural network. When the established network is applied to the real field data, the errors of S-wave velocity prediction are very small, all within 200 m/s, and the average relative errors are below 5%. In addition, the prediction errors of the deep S-wave velocity prediction network constructed in this study applied to laboratory data (3.32%) and well log data (2.24, 3.33, and 4.62%) are much smaller than that of Han's empirical formula (10.30, 6.68, 4.48, and 8.24%) and Castagna's empirical formula (11.40, 2.86, 3.73, and 7.33%). Compared with the two common empirical formulations, the deep S-wave velocity prediction network established has better prediction ability and generalization ability. The network is applicable to the S-wave velocity prediction in the full frequency band of sandstone reservoirs, and can provide S-wave velocity information for reservoir prediction work. The use of theoretical model forward simulation supplements the training dataset for machine learning, improving the interpretability of machine learning algorithms and generalization of models in real applications. Furthermore, we provide a new idea for the construction of labeled datasets in machine learning tasks.

Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

Author contributions

GF: acquisition and processing of data, methodology, manuscript writing and revising; H-HZ: manuscript reviewing and editing; X-RX: manuscript editing; G-YT: the conception and design of the study; Y-XW: provide some suggestions.

Funding

This study was supported by the Research on Full-Frequency Processing Method of Thin Reservoir and Research on Target Fine Characterization Technology Research (2022KT1503), the Research on Geophysical Description Technology of Continental Clastic Reservoir and Field Test (2022KT1505), the NSFC program (41930425) and the CNPC-CUPB Strategic Corporation Science and Technology Program (ZLZX2020-03).

Conflict of interest

Authors GF, H-HZ, X-RX, and Y-XW were employed by PetroChina.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Alimoradi, A., Shahsavani, H., and Rouhani, A. K. (2011). Prediction of shear wave velocity in underground layers using SASW and artificial neural networks. *Engineering* 3 (03), 266–275. doi:10.4236/eng.2011.33031

Ameen, M. S., Smart, B., Somerville, J. M., Hammilton, S., and Naji, N. A. (2009). Predicting rock mechanical properties of carbonates from wireline logs (A case study: Arab-D reservoir, Ghawar field, Saudi Arabia). *Mar. Pet. Geol.* 26 (4), 430–444. doi:10.1016/j.marpetgeo.2009.01.017

Bagheripour, P., Gholami, A., Asoodeh, M., and Vaezzadeh-Asadi, M. (2015). Support vector regression based determination of shear wave velocity. *J. Petroleum Sci. Eng.* 125, 95–99. doi:10.1016/j.petrol.2014.11.025

Biot, M. A. (1956). Theory of propagation of elastic waves in a fluid-saturated porous solid. I. Low-Frequency range. *J. Acoust. Soc. Am.* 28 (2), 168–178. doi:10.1121/1.1908239

Castagna, J. P., Batzle, M. L., and Eastwood, R. L. (1985). Relationships between compressional-wave and shear-wave velocities in clastic silicate rocks. *Geophysics* 50 (4), 571–581. doi:10.1190/1.1441933

Du, X. (2014). Shear wave velocity delicate estimation based on Trivariate Cauchy constraint AVO inversion. *Prog. Geophys.* 29 (2), 681–688. doi:10.6038/pg20140228

Dvorkin, J., Mavko, G., and Nur, A. (1995). Squirt Flow in fully saturated rocks. *Geophysics* 60 (1), 97–107. doi:10.1190/1.1443767

Eberhart-Phillips, D., Han, D. H., and Zoback, M. D. (1989). Empirical relationships among seismic velocity, effective pressure, porosity, and clay content in sandstone. *Geophysics* 54 (1), 82–89. doi:10.1190/1.1442580

Eskandari, H., Rezaee, M. R., and Mohammadnia, M. (2004). Application of multiple regression and artificial neural network techniques to predict shear wave velocity from wireline log data for a carbonate reservoir South-West Iran. *CSEG Rec.* 42, 48.

Gao, J., Song, Z., Gui, J., and Yuan, S. (2022). Gas-bearing prediction using transfer learning and CNNs: An application to a deep tight dolomite reservoir. *IEEE Geosci. Remote Sens. Lett.* 19, 1–5. doi:10.1109/LGRS.2020.3035568

Gassmann, F. (1951). Elastic waves through a packing of spheres. *Geophysics* 16 (4), 673–685. doi:10.1190/1.1437718

Guo, W., Dong, C., Lin, C., Wu, Y., Zhang, X., and Liu, J. (2022). Rock physical modeling of tight sandstones based on digital rocks and reservoir porosity prediction from seismic data. *Front. Earth Sci. (Lausanne).* 10, 932929. doi:10.3389/feart.2022.932929

Guo, Z., Qin, X., Zhang, Y., Niu, C., Wang, D., and Ling, Y. (2021). Numerical investigation of the effect of heterogeneous pore structures on elastic properties of tight gas sandstones. *Front. Earth Sci. (Lausanne).* 9, 641637. doi:10.3389/feart.2021.641637

Gurevich, B., Makarynska, D., De Paula, O. B., and Pervukhina, M. (2010). A simple model for squirt-flow dispersion and attenuation in fluid-saturated granular rocks. *Geophysics* 75 (6), N109–N120. doi:10.1190/1.3509782

Han, D., Nur, A., and Morgan, D. (1986). Effects of porosity and clay content on wave velocities in sandstones. *Geophysics* 51 (11), 2093–2107. doi:10.1190/1.1442062

Han, X., Wang, S., Tang, G., Dong, C., He, Y., Liu, T., et al. (2021). Coupled effects of pressure and frequency on velocities of tight sandstones saturated with fluids: Measurements and rock physics modelling. *Geophys. J. Int.* 226 (2), 1308–1321. doi:10.1093/gji/ggab157

Hill, R. (1952). The elastic behaviour of a crystalline aggregate. *Proc. Phys. Soc. A* 65 (5), 349–354. doi:10.1088/0370-1298/65/5/307

Hu, C., Wang, F., and Ai, C. (2021). Calculation of average reservoir pore pressure based on surface displacement using image-to-image convolutional neural network model. *Front. Earth Sci. (Lausanne).* 9, 712681. doi:10.3389/feart.2021.712681

Hu, R., Peng, Z., Ma, J., and Li, W. (2020). CNN-based vehicle target recognition with residual compensation for circular SAR imaging. *Electronics* 9 (4), 555. doi:10.3390/electronics9040555

Kuster, G. T., and Toksőz, M. N. (1974). Velocity and attenuation of seismic waves in two-phase media:Part I. Theoretical formulations. *Geophysics* 39 (5), 587–606. doi:10.1190/1.1440450

Li, D., Wei, J., Di, B., Ding, P., Huang, S., and Shuai, D. (2018). Experimental study and theoretical interpretation of saturation effect on ultrasonic velocity in tight sandstones under different pressure conditions. *Geophys. J. Int.* 212 (3), 2226–2237. doi:10.1093/gji/ggx536

Maleki, S., Moradzadeh, A., Riabi, R. G., Gholami, R., and Sadeghzadeh, F. (2014). Prediction of shear wave velocity using empirical correlations and artificial intelligence methods. *NRIAG J. Astronomy Geophys.* 3 (1), 70–81. doi:10.1016/j.nrjag.2014.05.001

Mehrgini, B., Izadi, H., and Memarian, H. (2017). Shear wave velocity prediction using Elman artificial neural network. *Carbonates Evaporites* 34 (4), 1281–1291. doi:10.1007/s13146-017-0406-x

Sun, S. Z., Wang, H., Liu, Z., Li, Y., Zhou, X., and Wang, Z. (2012). The theory and application of DEM-Gassmann rock physics model for complex carbonate reservoirs. *Lead. Edge* 31 (2), 152–158. doi:10.1190/1.3686912

Wang, J., Cao, J., Zhao, S., and Qi, Q. (2022). S-wave velocity inversion and prediction using a deep hybrid neural network. *Sci. China Earth Sci.* 65 (4), 724–741. doi:10.1007/s11430-021-9870-8

White, J. E. (1975). Computed seismic speeds and attenuation in rocks with partial gas saturation. *Geophysics* 40 (2), 224–232. doi:10.1190/1.1440520

Wu, X., Geng, Z., Shi, Y., Pham, N., Fomel, S., and Caumon, G. (2019). Building realistic structure models to train convolutional neural networks for seismic structural interpretation. *Geophysics* 85 (4), WA27–WA39. doi:10.1190/geo2019-0375.1

Xu, S., and Payne, M. A. (2009). Modeling elastic properties in carbonate rocks. *Lead. Edge* 28 (1), 66–74. doi:10.1190/1.3064148

Xu, S., and White, R. E. (1995). A new velocity model for clay-sand mixtures [1]. *Geophys. Prospect.* 43 (1), 91–118. doi:10.1111/j.1365-2478.1995.tb00126.x

Xu, S., and White, R. E. (1996). A physical model for shear-wave velocity prediction. *Geophys. Prospect.* 44 (4), 687–717. doi:10.1111/j.1365-2478.1996.tb00170.x

Yu, S., and Ma, J. (2021). Deep learning for geophysics: Current and future trends. *Rev. Geophys.* 59 (3), e2021RG000742. doi:10.1029/2021RG000742

Yuan, S., Jiao, X., Luo, Y., Sang, W., and Wang, S. (2022). Double-scale supervised inversion with a data-driven forward model for low-frequency impedance recovery. *Geophysics* 87 (2), R165–R181. doi:10.1190/geo2020-0421.1

Yuan, S., Liu, J., Wang, S., Wang, T., and Shi, P. (2018). Seismic waveform classification and first-break picking using convolution neural networks. *IEEE Geosci. Remote Sens. Lett.* 15 (2), 272–276. doi:10.1109/LGRS.2017.2785834

Zhang, Y., Zhang, C., Ma, Q., Zhang, X., and Zhou, H. (2022). Automatic prediction of shear wave velocity using convolutional neural networks for different reservoirs in Ordos Basin. *J. Petroleum Sci. Eng.* 208, 109252. doi:10.1016/j.petrol.2021.109252

Zhang, Y., Zhong, H. R., Wu, Z. Y., Zhou, H., and Ma, Q. Y. (2020). Improvement of petrophysical workflow for shear wave velocity prediction based on machine learning methods for complex carbonate reservoirs. *J. Petroleum Sci. Eng.* 192, 107234. doi:10.1016/j.petrol.2020.107234

frontiers | Frontiers in Earth Science

TYPE Original Research
PUBLISHED 17 January 2023
DOI 10.3389/feart.2023.1039683

OPEN ACCESS

EDITED BY
Peng Zhenming,
University of Electronic Science and Technology of China, China

REVIEWED BY
Siwei Yu Siwei,
Harbin Institute of Technology, China
Ramin Ranjbarzadeh,
Dublin City University, Ireland

*CORRESPONDENCE
Li-De Wang,
wld2402364497@163.com

SPECIALTY SECTION
This article was submitted to Environmental Informatics and Remote Sensing, a section of the journal Frontiers in Earth Science

RECEIVED 08 September 2022
ACCEPTED 05 January 2023
PUBLISHED 17 January 2023

CITATION
Wang L-D, Wu J, Xu X-R, Zeng H-H, Gao Y and Liu W-Q (2023), Intelligent velocity picking considering an expert experience based on the Chan–Vese model and mean-shift clustering.
Front. Earth Sci. 11:1039683.
doi: 10.3389/feart.2023.1039683

Intelligent velocity picking considering an expert experience based on the Chan–Vese model and mean-shift clustering

Li-De Wang[1]*, Jie Wu[1], Xing-Rong Xu[1], Hua-Hui Zeng[1], Yang Gao[2] and Wen-Qing Liu[1]

[1]Research Institute of Petroleum Exploration & Development-Northwest (NWGI), PetroChina, Lanzhou, China, [2]China University of Petroleum, Beijing, China

The velocity of seismic data can initially be established by identifying energy clusters on velocity spectra at different moments, which is crucial to the migration imaging and the stacking of common midpoint (CMP) gathers in the seismic data processing. However, the identification of energy clusters currently relies on manual work, with low efficiency and different standards. With the increasing application of wide-frequency, wide-azimuth, and high-density seismic exploration technology, the amount of seismic data has increased significantly, greatly increasing the cost of manual labor and time. In this paper, an intelligent velocity picking method based on the Chan–Vese (CV) model and mean-shift clustering algorithm was proposed. It can be divided into three steps. First, a velocity trend band is set up on the velocity spectrum by experts to avoid multiples and other noises. Then, the velocity trend band is applied to the Chan–Vese model as the initial time condition to segment the velocity spectrum and obtain the velocity candidate region. Finally, mean-shift clustering is adopted to cluster the useful energy clusters retained in the candidate region derived from the Chan–Vese model. When implementing the mean-shift clustering algorithm, the Gaussian kernel function and the energy of the velocity spectrum are utilized to control the efficiency and accuracy of the cluster. The tests of the model and real data prove that the proposed method can dramatically improve the accuracy and efficiency of velocity picking compared with the K-means and manual picking method.

KEYWORDS

velocity spectrum, intelligent velocity picking, Chan–Vese, mean-shift clustering, expert experience constraint

1 Introduction

Velocity analysis of a seismic wave is a critical step in seismic data processing and also the basis for subsequent data processing procedures and interpretation. For example, the normal moveout (NMO) correction relies on stacking velocity (Wang et al., 2021a), the migration imaging relies on migration velocity (Jones et al., 1998; Nemeth et al., 1999; Hou & Marfurt, 2002), and time–depth conversion relies on time-domain velocity (Cameron et al., 2008). If the velocity field is accurate, the seismic profiles obtained by migration can reflect the underground structure more clearly. Currently, the velocity is mainly obtained by manually picking the velocity energy clusters. Although manual picking makes full use of expert experience, it is labor-intensive and repetitive. Moreover, the manual way is generally of low density in the picking of a velocity spectrum and has different views on the characterization of complex

geological structures. Therefore, it is imperative to establish an efficient automatic velocity picking method to relieve the labor and improve the accuracy of velocity imaging in structures, following the main principle of the manual method and making full use of expert knowledge, especially with the increase in seismic data.

In order to improve efficiency and accuracy, in the early days, model-driven automatic velocity picking algorithms were developed rapidly. From the fundamentals, considering the mathematical and physical relationship between velocity and seismic data is the main thought behind these model-driven algorithms. A large number of scholars used an iterative optimization method to search for the optimal velocity by establishing an objective function that could represent waveform consistency or stacked energy. Toldi (1989) was the first to suggest the automatic velocity picking method. He designed an objective function by maximizing the sum of stacked energy, and the optimal velocity is obtained by iterative updating. However, this method assumes that the model is linear but not in practice, and this brings about inapplicability in low signal-to-noise seismic data. Moreover, this method considers a lot of constraints and complicates the model, which makes the practical performance of the method very poor. Zhang et al. (2015) established a non-linear objective function by analyzing the selected velocity distribution rules, which can achieve automatic velocity picking; however, this still does not solve the large computational effort and low noise immunity of the model-driven approach. Moreover, Wilson and Gross (2019) used a particle swarm optimization method to find the optimal time difference to flatten the hyperbolic curve which solves the local minimum problems partially. Velis (2021) gave a non-hyperbolic and anisotropic velocity analysis algorithm. The objective function of non-hyperbolic energy involved the anisotropy parameters, and the velocity was established by the simulated annealing-based iterative search. In this method, static and dynamic boundaries were used to avoid multiple noises. In addition, Yuan et al. (2019) proposed that full waveform inversion can also retrieve the background velocity structure, and the low-frequency full-waveform inversion result was considered as an *a priori* model, which can be applied to reservoir prediction. However, the full waveform inversion is a strong non-linear inversion. When the accuracy of the initial velocity model is not enough, there will be cycle jumps and it will result in failure. All in all, the model-driven method requires the hypothesis that the mathematical physical model used can accurately express the relationship between seismic data and velocity, causing the problems of local minima and large costs of computation.

In recent years, due to the significant improvement in computer performance, machine learning has been applied to various fields, such as tumor and liver segmentation in CT images (Aghamohammadi et al., 2021), brain tumor segmentation (Ranjbarzadeh et al., 2021), and breast tumor segmentation in mammograms (Ranjbarzadeh et al., 2022). Also, for geophysics, underground structure segmentation, automatic velocity picking can be achieved by deep learning methods or unsupervised clustering methods.

Compared to traditional model-driven methods, deep learning can be regarded as a data-driven method, which can establish an optimal non-linear mapping relationship between seismic data (e.g., common middle point gathers, velocity spectra, or shot gathers) and velocity. The deep learning method updates the network parameters mainly through a certain depth of neural network model and a back-propagation algorithm (Rumelhart et al., 1986), as well as automatically learning the effective features in data and the establishment of multi-domain mapping (LeCun et al., 2015; Goodfellow et al., 2016). Park and Sacchi (2020) proposed the velocity automatic picking method based on the convolutional neural network (CNN); this class of methods transforms the identification of energy clusters on a velocity spectrum into a classification problem in the field of image recognition; hence, this method requires a high signal-to-noise ratio of the velocity spectrum. Biswas et al. (2019) and Zhang et al. (2019) proposed the recurrent neural network (RNN)-based automatic velocity picking method, which considers the temporal order of seismic data and treats the velocity picking as a normalization problem, resulting in higher accuracy. In addition, Fabien-Ouellet and Sarkar (2020) combined the CNN and long- and short-term memory (LSTM) network to estimate the root mean square velocity and interval velocity of seismic data. The combination of these two networks can simultaneously learn the characteristics of CMP gathers and velocity spectra to more accurately predict the velocity. Wang et al. (2020) contrasted the velocity picking algorithms of the regression-based neural network and the classification-based neural network and claimed that both methods could reasonably predict the velocity field. Then, Yuan et al. (2022) proposed a double-scale gated recurrent unit neural network method, which uses data-driven methods to learn forward, and inversion simulations to establish the non-linear relationship between post-stack data and velocity or impedance; this method recovers the low-frequency impedance so as to make the velocity field more accurate, and geological laws and blind wells can verify its rationality. Recently, Cao et al. (2022) proposed a seismic velocity inversion method based on the CNN-LSTM fusion deep neural network, which can simultaneously estimate the root mean square velocity and interval velocity from the CMP gather. In the proposed method, a CNN encoder and two LSTMs are used to extract spatial and temporal features from seismic data, and a CNN decoder is used to recover the velocity, improving the accuracy of imaging. As a whole, all the deep learning methods mentioned above that get rid of the ill-posed inverse problem of the traditional model-driven method use artificial neural networks to establish a non-linear mapping relationship between seismic data or a velocity spectrum (input) and a velocity model (output). However, such supervised neural network intelligent velocity picking methods require professionals to manually pick up a rich and large amount of labeled data for training, which is time-consuming and has weak generalization ability. When predicting seismic data with different features, the labeled data need to be reconstructed with a retrained network. In addition, it is impossible to interpret the process of training since the prediction process of seismic data is a huge composite function.

For the attractive unsupervised clustering method, it can search for energy cluster features of velocity without constantly establishing labels and training according to different data. This type of intelligent picking method groups the energy clusters of velocity with the same features used to obtain the approximate distribution of the data. It has a simple algorithm and is easy to implement. In addition, it is highly interpretable and generalizable, adapting to seismic data of any features. Therefore, it is more extensible in industrial applications. Agudelo et al. (2017) and Araya-Polo et al. (2017) use the K-means clustering algorithm that uses the distance of samples as a similarity indicator to process the seismic data, and the class center is regarded as the picking location. However, it poorly clusters non-spherical energy clusters of velocity, and the K values

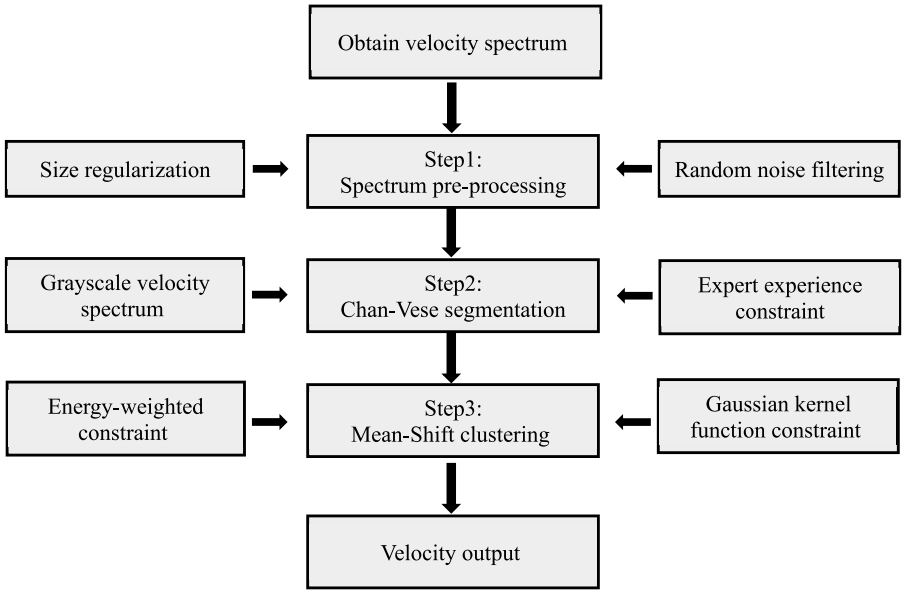


FIGURE 1
Intelligent velocity picking workflow.

need to be fixed manually. Chen et al. (2018) proposed a bottom-up strategy, considering the problem of different K values for different velocity spectra, which solves the drawback of fixed K values to some extent. Waheed et al. (2019) also applied the density clustering method to pick up the energy clusters of velocity to achieve automatic velocity picking, which avoids the problem of choosing K values manually in K-means clustering. In addition, Wang et al. (2021b) proposed an approach based on adaptive threshold-constrained K-means; this method can improve the identification of the energy cluster of velocity which was weak on the velocity spectrum. Wang et al. (2022) also suggested a Gaussian mixture clustering method to achieve automatic velocity picking, as an extreme case of the Gaussian mixture model, and K-means is difficult to characterize velocity energy clusters with low focus ability, while the Gaussian mixture model can accurately fit and provide uncertainty analysis at the same time. These unsupervised clustering algorithms, however, only consider the ability to identify energy clusters of the velocity, without taking into account the complexity of the actual seismic data and expert experiences, resulting in the velocity picking methods affected by the multiples and some other noises on the velocity spectrum.

In this article, we proposed an automatic velocity picking method based on the Chan–Vese (CV) model and mean-shift clustering algorithm, which can effectively improve the accuracy and efficiency of velocity picking and solve the interference of multiples. Meanwhile, the method we proposed can also reduce the manual labor and improve the adaptability of unsupervised clustering methods in actual seismic data. First, a velocity trend band is set up on the velocity spectrum by experts to avoid multiples and noises. Then, the velocity trend band is applied to the CV model as the initial time condition to segment the valid energy clusters of velocity on the velocity spectrum. Finally, the mean-shift clustering algorithm is adopted to cluster the valid energy clusters in the candidate region derived from the CV model. In order to improve the accuracy and efficiency of mean-shift clustering, the Gaussian kernel function and the energy value of velocity are applied for weighting, which allows the identification of not only the isolated and highly focused energy clusters of velocity but also the interconnected energy clusters.

2 Methods

We started with a brief analysis of how seismic data processors perform velocity picking. After obtaining a CMP velocity spectrum, the processor first analyzes the trend and range of velocity for that CMP using empirical and geological knowledge; then, the energy clusters of effective reflected waves within the correct trend are identified with the naked eye and cluster centers are picked up as the velocity for that location. The entire picking is the process by which processors translate geophysical and geological theories into the geometry of energy clusters on the velocity spectrum (Wang et al., 2022). Our method follows this process to achieve automatic velocity picking. Corresponding to the first step of manual picking, the valid velocity trend and range are identified using the CV model with expert experience constraints, and the second step of manual picking is replaced by the mean-shift clustering method.

The main body of our method in this article can be divided into three parts, as shown in Figure 1. The first step is the pre-processing of the velocity spectrum, including the size regularization of the velocity spectrum and the random noise filtering. The size regularization of the velocity spectrum is to make the time dimension of each velocity spectrum consistent so as to ensure that the velocity curve picked up in every velocity spectrum, subsequently, has the same sampling time for stacking; generally, this time dimension is consistent with the seismic records. Random noise filtering is used to filter some salt-and-pepper noises on the velocity spectrum by using the median filtering method so as to improve the accuracy of energy cluster identification of the subsequent CV model. The second step is the output of the effective

candidate region of velocity by the CV model, the expert experience constraints, and the grayscale based on velocity spectrum energy, which is the normalization of velocity spectrum energy that is performed in this step. The third step is the use of the mean-shift clustering method to pick up the centers of the effective reflected energy clusters, and the Gaussian kernel function and energy weighting constraints are applied in this step. These steps are discussed in the following sections in detail.

2.1 CV model in velocity spectrum segmentation

Image segmentation is an important image analysis and processing technology that has been widely used in medical image analysis, intelligent traffic management, remote sensing image processing, and other fields (Li et al., 2021). Image segmentation methods include threshold segmentation, region segmentation, edge segmentation, and the active contour model and level set methods (Osher & Sethian, 1988). Getreuer (2012) shows that based on level set methods, Chan and Vese proposed the CV model which solves the problem of computational difficulty caused by the primary length term and the secondary area term in the expression. Keegan et al. (2017) proposed that the CV model be the basis of multi-phase image segmentation, and the model does not need to define the boundary by gradient, which significantly reduces the complexity and improves the efficiency of segmentation.

2.1.1 Theory of the CV model

The differences in location, size, and shape of velocity spectrum energy clusters make the level-set-based CV model more effective. The energy function of the CV model can be expressed as follows:

$$E(C) = \mu Length(C) + vArea(C) + \lambda_1 E_1(C) + \lambda_2 E_2(C), \quad (1)$$

where the first term of $E(C)$ is the length constraint of the evolution curve, in which the curve (C) is originally provided by an effective velocity trend band on the velocity spectrum; the second term is the shape constraint of C; and the third and fourth terms are the pixel losses based on the grayscale map of the velocity spectrum inside and outside C, respectively, which are calculated by Eqs 2, 3. μ, v, λ_1, and λ_2 are the constant coefficients, which are generally taken as 1.

$$E_1(C) = \int_{inside(C)} |I_0(x,y) - C_1|^2 dxdy, \quad (2)$$

$$E_2(C) = \int_{outside(C)} |I_0(x,y) - C_2|^2 dxdy, \quad (3)$$

where $I_0(x,y)$ denotes each pixel value of the inner region and outer region of the evolution curve C. C_1 and C_2 are the average of the pixels in the inner and outer regions of curve C, respectively. $E_1(C)$ and $E_2(C)$ are minimized when the evolution curve C can correctly partition the effective reflected energy clusters of the velocity spectrum.

To correctly solve the evolution of C, the CV model uses the level set method when the energy function is minimized, and the level set method replaces the evolution of C with the evolution of a curved surface $\phi(x,y)$. The specific programs are as follows:

$$\begin{cases} C = \partial\omega = \{(x,y) \in \Omega: \phi(x,y) = 0\} \\ inside(C) = \omega = \{(x,y) \in \Omega: \phi(x,y) > 0\} \\ outside(C) = \Omega\backslash\bar{\omega} = \{(x,y) \in \Omega: \phi(x,y) < 0\} \end{cases}. \quad (4)$$

C is denoted by $\phi(x,y) = 0$. The inner region of C is denoted by $\phi(x,y) > 0$, and the outer region of C is denoted by $\phi(x,y) < 0$. Then, we define a step function and its derivative as

$$H(z) = \begin{cases} 1, & z > 0, \\ 0, & z < 0, \end{cases} \quad (5)$$

$$\delta_0(z) = \frac{d}{dz} H(z). \quad (6)$$

Substituting Eqs 4–6 into Eq 1, the energy function of the curved surface based on the level set is obtained as follows:

$$\begin{aligned} E(\phi(x,y)) &= \mu \int_\Omega |\nabla H(\phi(x,y))| dxdy + v \int_\Omega H(\phi(x,y)) dxdy \\ &+ \lambda_1 \int_\Omega |I_0(x,y) - C_1|^2 H(\phi(x,y)) dxdy \\ &+ \lambda_2 \int_\Omega |I_0(x,y) - C_2|^2 (1 - H(\phi(x,y)) dxdy \end{aligned}. \quad (7)$$

Using the energy function E and its corresponding Euler–Lagrange to minimize Eq 7, the result is

$$\frac{\partial\phi}{\partial t} = \delta_0(\phi)\left[\mu div\left(\frac{\nabla\phi}{|\nabla\phi|}\right) - v - \lambda_1 (I - C_1)^2 + \lambda_2 (I - C_2)^2\right], \quad (8)$$

$$C_1 = \frac{\int_\Omega I_0(x,y) H(\phi(x,y)) dxdy}{\int_\Omega H(\phi(x,y)) dxdy}, \quad (9)$$

$$C_2 = \frac{\int_\Omega I_0(x,y) (1 - H(\phi(x,y)) dxdy}{\int_\Omega (1 - H(\phi(x,y)) dxdy}. \quad (10)$$

Therefore, the final evolution partial differential equation of the CV model is

$$\begin{cases} \frac{\partial\phi}{\partial t} = \delta_0(\phi)\left[\mu div\left(\frac{\nabla\phi}{|\nabla\phi|}\right) - v - \lambda_1 (I - C_1)^2 + \lambda_2 (I - C_2)^2\right] \\ \phi|_{t=0} = u(x,y) \end{cases}, \quad (11)$$

where $u(x,y)$ denotes C at the initial moment of the evolution equation. In our method, to obtain faster and more accurate segmentation results, experts are required to combine geophysical and geological theories to give the original range of velocity, which is treated as $u(x,y)$, and the energy clusters outside the $u(x,y)$ are not involved in the subsequent clustering. Then, the following advantages can be obtained:

1) The curve evolution speed can be accelerated
2) The interference of multiples and other noises can be avoided
3) The valid velocity candidate region can be provided for mean-shift clustering

Also, the process of segmenting the velocity spectrum using the CV model can be summarized as follows:

1) Size regularization and random noise filtering for the velocity spectrum
2) Energy-oriented grayscale of the velocity spectrum

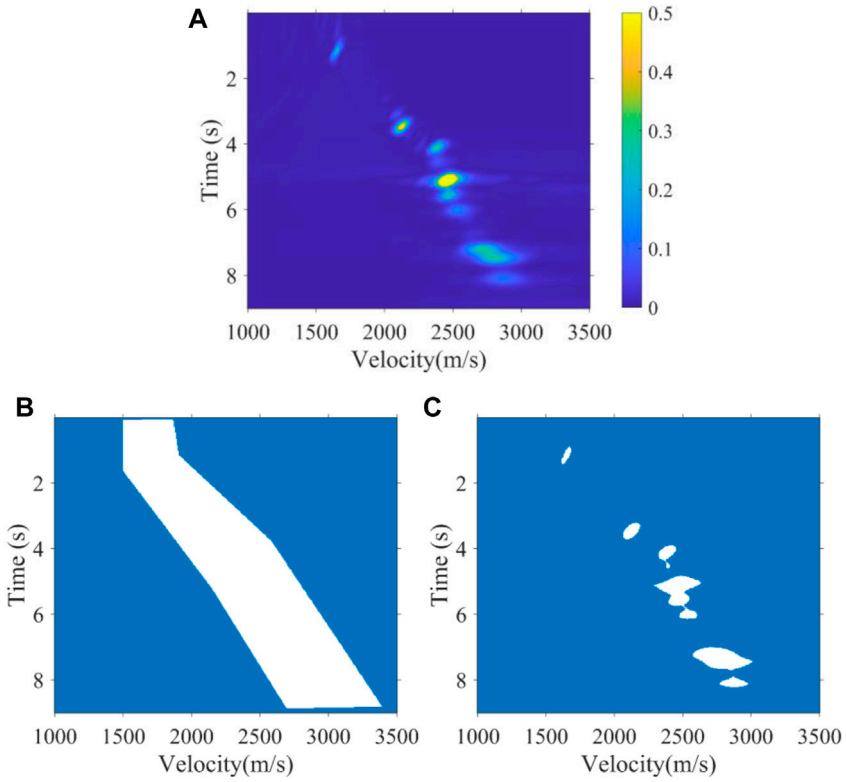


FIGURE 2
(A) 40th velocity spectrum with a high signal-to-noise ratio of the Marmuosi model, **(B)** expert experience constraint band, and **(C)** segmentation result.

3) Provide an initial evolutionary curve by experts
4) Evolve segmentation curves to obtain an effective velocity candidate region for mean-shift clustering

2.1.2 Velocity spectrum segmentation examples

Our aim is that the CV model can segment the velocity energy clusters that are truly effective as shown in Figure 2 and Figure 3, just as the manual pickup of energy clusters. We applied the CV model with expert constraints previously to the velocity spectra of one synthetic data and one real data. Depending on the lateral variation of the velocity, the expert constraint bands were implemented at intervals of 10 and 500 velocity spectra. Figure 2 shows the segmentation result of the 40th velocity spectrum of the Marmuosi model (Martin et al., 2006). For high signal-to-noise ratio synthetic data, the CV model can well identify the effective energy clusters of the velocity. Figure 3 shows the segmentation results of the 3900th velocity spectrum of the real data, which develops multiples in 6–9 s. The energy clusters of multiples velocity are interference signals, which generally appear in the high-time low-velocity region in the velocity spectrum as shown in Figure 3, and the CV model can avoid the effect of multiples energy clusters after the expert experience constraint and just obtaining the valid energy clusters of the velocity.

2.2 Mean-shift clustering for energy cluster picking

Mean-shift clustering based on density does not need to artificially determine the number of clusters and the initial cluster center locations like other algorithms, such as K-means. It can automatically select the number of clusters based on the density distribution of the data (Wang et al., 2018). Therefore, mean-shift clustering can be well adapted to a case of a continuous distribution of energy clusters on the velocity spectrum. At the same time, mean-shift clustering is less computational, faster, and more stable, so we use mean-shift clustering to improve efficiency.

2.2.1 Theory of mean-shift clustering

The mean-shift algorithm (Comaniciu and Meer, 2002) is an iterative process in which the mean value position of the energy cluster is calculated. In each iteration, the mean value position is updated and then the updated position is used as a new start to calculate the shift value until it reaches the threshold. The calculation of the shift value can be expressed as

$$\mathbf{M}_h = \frac{1}{k} \sum_{\mathbf{x}_i \in D_h} (\mathbf{x}_i - \mathbf{x}), \tag{12}$$

where $\mathbf{x}$ is the mean value position of the circle D_h with $\mathbf{x}$ as the center and h as the radius, and $\mathbf{x}_i$ is the position of every energy point within D_h. k is the number of energy points within D_h. Then, the $\mathbf{x}$ is updated by

$$\mathbf{x}^{\star} = \mathbf{x} + \mathbf{M}_h. \tag{13}$$

Eqs 12, 13 allow the mean value position to continuously move toward the center of energy clusters, and the update stops when the shift value $\mathbf{M}_h$ is less than the threshold value.

As shown in Eq. 12, the contribution of energy points within D_h are the same when calculating the shift value; however, the energy of

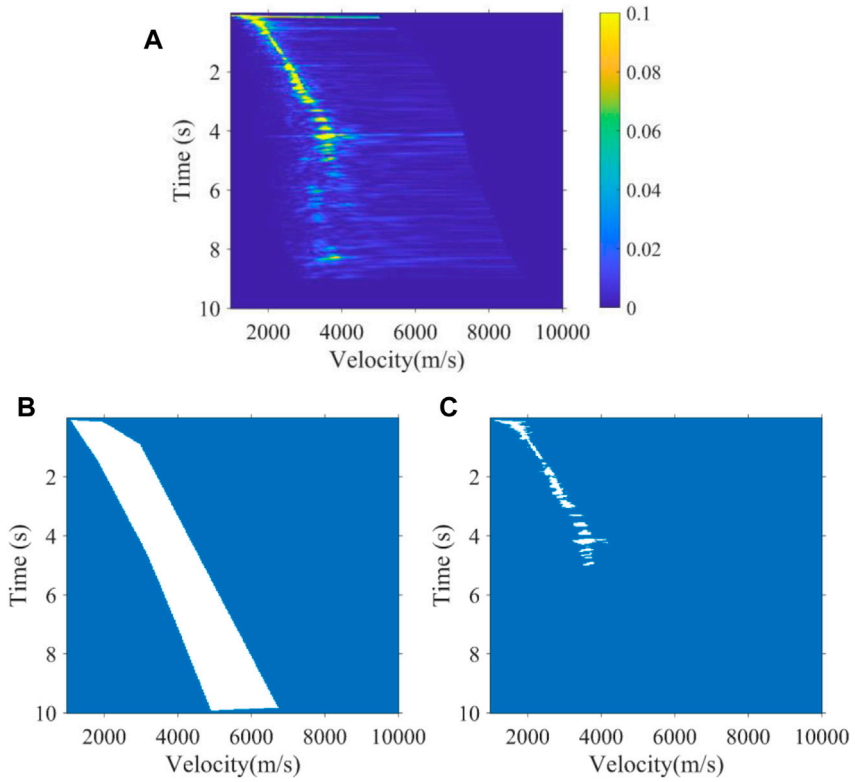


FIGURE 3
(A) 3900th velocity spectrum with multiple developments in 6–9 s of real data, **(B)** expert experience constraint band, and **(C)** segmentation result.

each point on the velocity spectrum is different, leading to different contributions. The larger the energy, the more the contributions will be. Therefore, we applied an improved calculation of the shift value based on energy weighting to improve the accuracy of energy cluster picking:

$$\mathbf{M}_h = \frac{\sum_{i=1}^{n} G\left(\left\|\frac{\mathbf{x}_i-\mathbf{x}}{h}\right\|^2\right)\omega(\mathbf{x}_i)(\mathbf{x}_i-\mathbf{x})}{\sum_{i=1}^{n} G\left(\left\|\frac{\mathbf{x}_i-\mathbf{x}}{h}\right\|^2\right)\omega(\mathbf{x}_i)}, \tag{14}$$

where G is the Gaussian kernel function, h denotes the radius of the region D_h, and $\omega(\mathbf{x}_i)$ is the energy of $\mathbf{x}_i$ within the region D_h. Substituting Eq. 14 into Eq. 13 to obtain the new update equation, we get

$$\mathbf{x}^{\star} = \frac{\sum_{i=1}^{n} G\left(\left\|\frac{\mathbf{x}_i-\mathbf{x}}{h}\right\|^2\right)\omega(\mathbf{x}_i)\mathbf{x}_i}{\sum_{i=1}^{n} G\left(\left\|\frac{\mathbf{x}_i-\mathbf{x}}{h}\right\|^2\right)\omega(\mathbf{x}_i)}. \tag{15}$$

Eq. 15 is used to iterate the mean value position of the energy cluster, until $\mathbf{x}^{\star}$ moves to the position where the energy of the cluster is maximum, which is regarded as the picking position.

2.2.2 Examples of velocity picking using mean-shift clustering

We applied the mean-shift clustering method proposed previously to identify the segmentation results of the CV model in subsection 2.2. Due to the constraints of expert experience, the CV model only selects the valid energy clusters of the velocity of effective reflection seismic waves and abandons the incorrect energy clusters of the velocity of disturbing multiples in deep regions. In fact, the most classic manual method is based on personal experience to roughly pick up the valid velocity energy clusters, while there is no valid energy cluster in the deep region of a velocity spectrum, and they obtain the velocity by using velocity curve fitting. So, when using the mean-shift clustering method, only the energy clusters segmented by the CV model several times are picked up. At other sampling times without valid energy clusters, we obtain the velocity by fitting the velocity curve according to the expert experience trend and accurate velocity obtained intelligently. To illustrate the correctness of our proposed method more intuitively, we place the real velocity of the Marmousi model and most classic manual picking results of real data in Figure 4 and Figure 5 for comparison, respectively. Furthermore, Figure 4 shows the picking results of the 40th velocity spectrum of the Marmuosi model, and Figure 5 shows the picking results of the 3900th velocity spectrum of the real data. These tests show whether the velocity spectrum has a high signal-to-noise ratio or interfering multiples development, and the intelligent picking results we proposed in this article always conform to the manual picking result; this means that our automatic picking method can replace manual picking for high accuracy and efficiency, with each velocity spectrum being picked up in just 1 s.

3 Examples

3.1 Synthetic data

To test the feasibility of the proposed method, we tested the 2D Marmousi model shown in Figure 6A with 60 velocity spectra

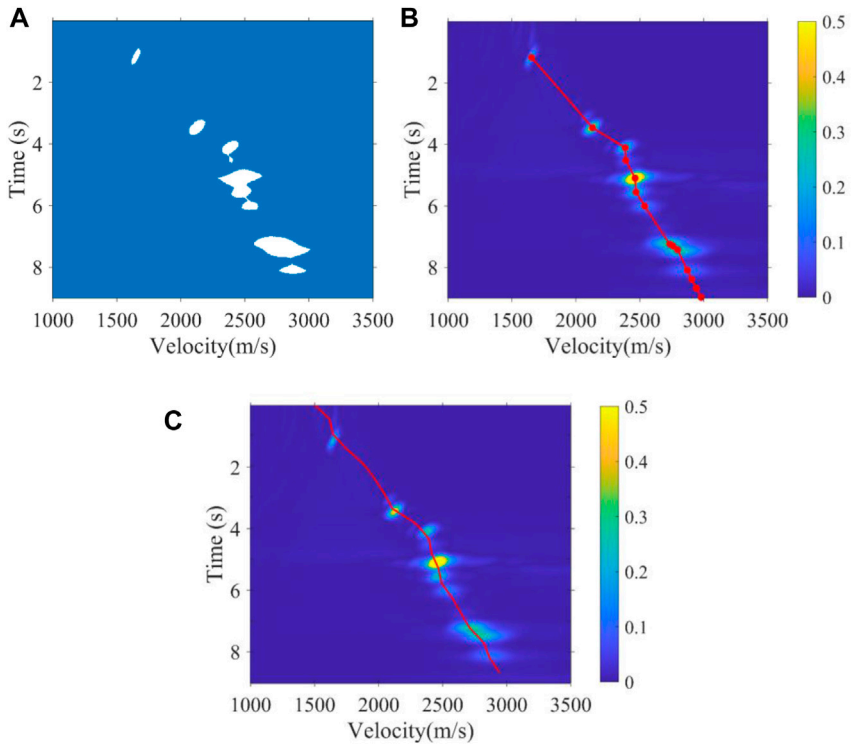


FIGURE 4
40th velocity spectrum of the Marmuosi model. **(A)** Segmentation result, **(B)** intelligent picking result by using the method proposed, and **(C)** real velocity curve.

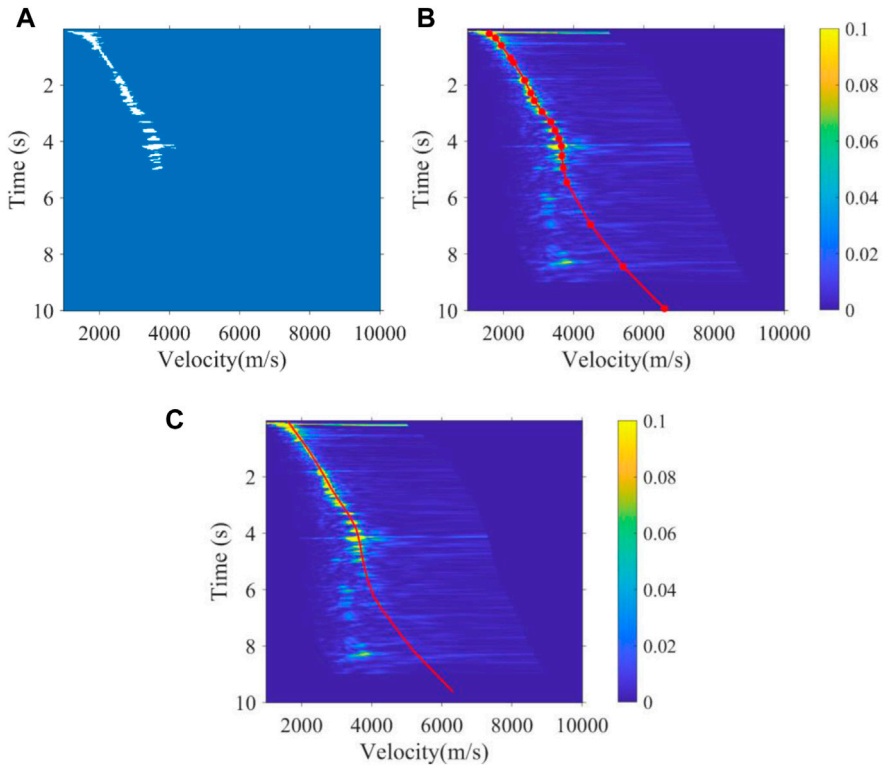


FIGURE 5
3900th velocity spectrum with multiple developments in 6–9 s of real data. **(A)** Segmentation result, **(B)** intelligent picking result by using the method proposed, and **(C)** manual picking result by semblance analysis.

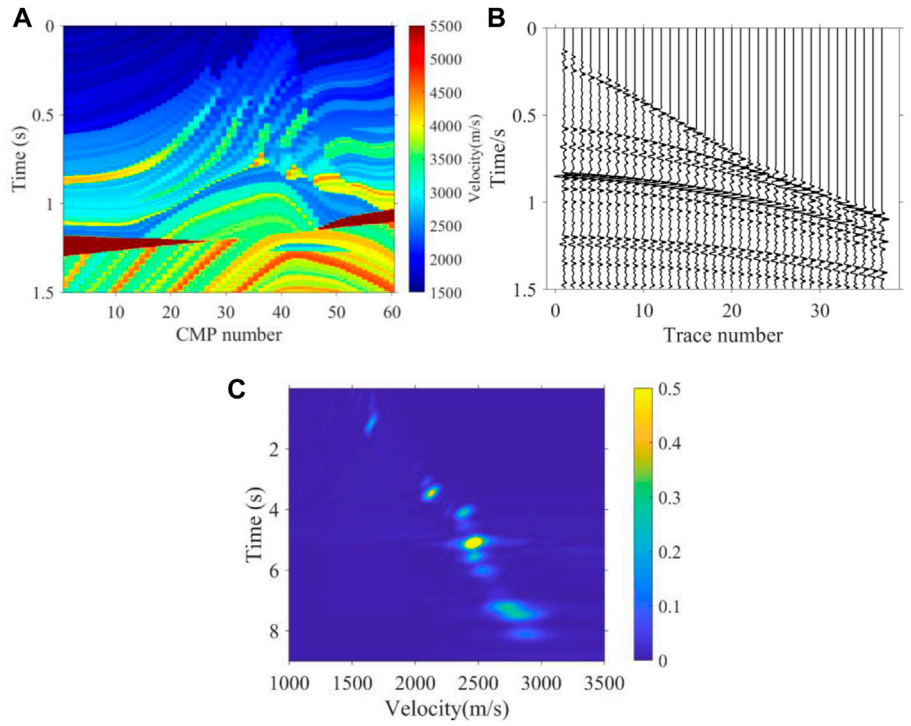


FIGURE 6
(A) True interval velocity field, **(B)** 40th CMP seismic records of synthetic data without noise, and **(C)** 40th velocity spectrum of synthetic data without noise.

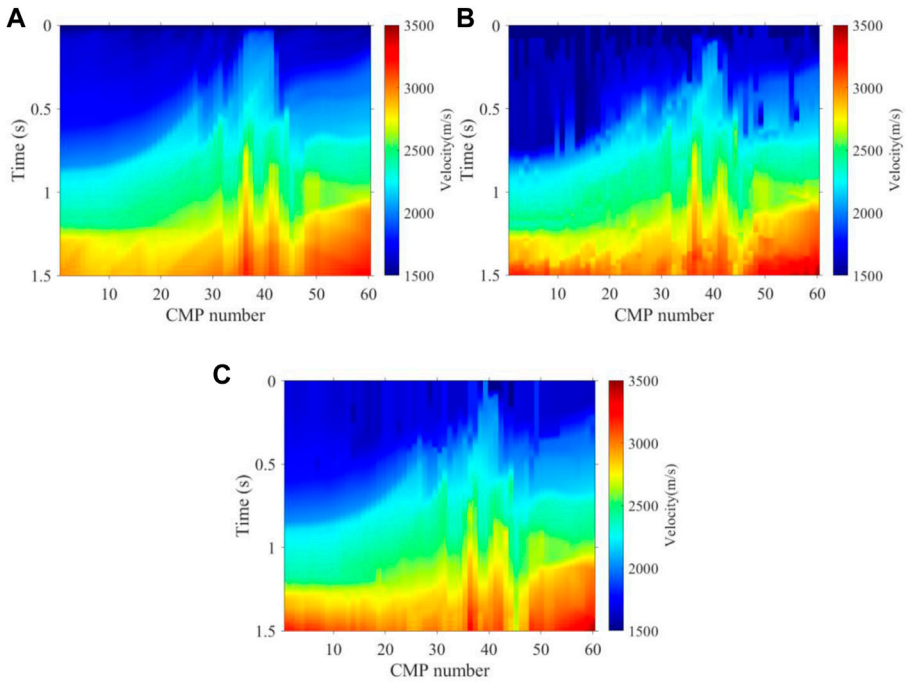


FIGURE 7
Stacking velocity field of synthetic data without noise. **(A)** True velocity, **(B)** K-means intelligent picking velocity, and **(C)** intelligent picking velocity by the method proposed.

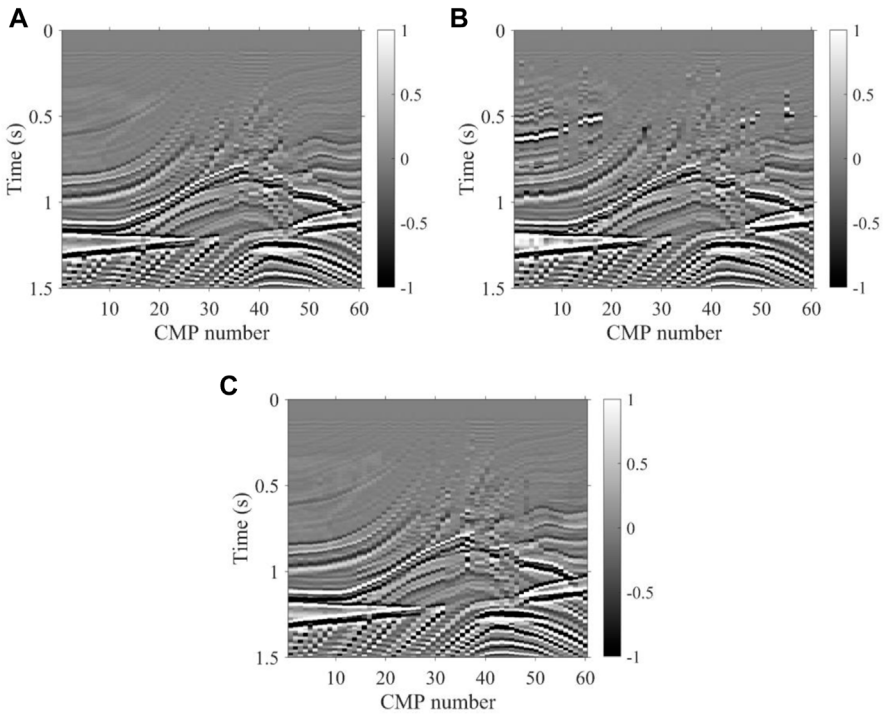


FIGURE 8
Stacking profiles of synthetic data without noise. **(A)** True velocity, **(B)** velocity of K-means, and **(C)** velocity of the method proposed.

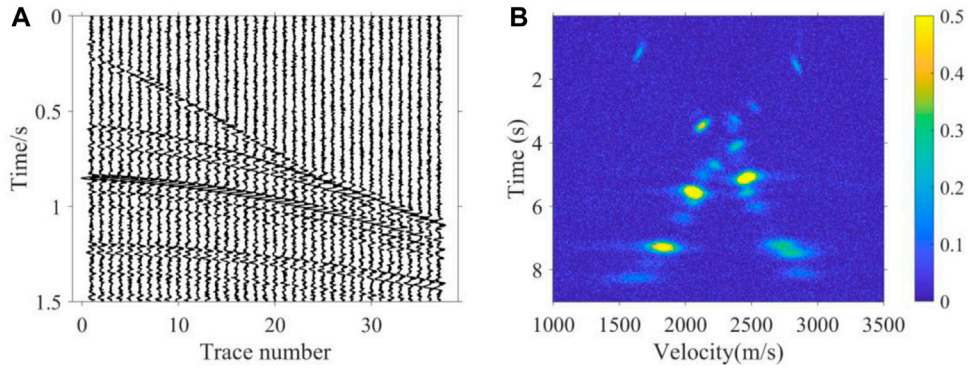


FIGURE 9
(A) 40th CMP seismic records of synthetic data with random noise. **(B)** 40th velocity spectrum of synthetic data with random noise and strong energy regular noise.

responding to 60 CMPs, while the distribution of energy clusters of velocity is different for each velocity spectrum. We selected a representative part of P-wave velocity in the Marmousi model and directly replaced the depth domain with the time domain. Through Dix's equation, we calculated the stacking velocity as a reference, as shown in Figure 7A. It has a complex structure, in which the lateral velocity changes sharply and the fault dip angle is large. There are 60 velocity spectra (CMPs) in the horizontal direction and 749 time sampling points in the vertical direction, and the time interval is 2 ms. All the velocity curves picked up on the 60 velocity spectra form a complete two-dimensional velocity field. In fact, every velocity spectrum is produced from seismic records through a series of mathematical calculations; hence, the signal-to-noise ratio of the seismic record determines the signal-to-noise ratio of the corresponding velocity spectrum. First, we tested the velocity spectra without noise which come from the seismic records, as shown in Figure 6B and Figure 6C.

Figure 7B and Figure 7C show the two-dimensional stacking velocity of traditional K-means and the method we proposed, respectively, which are combined with the 60 picked curves of the

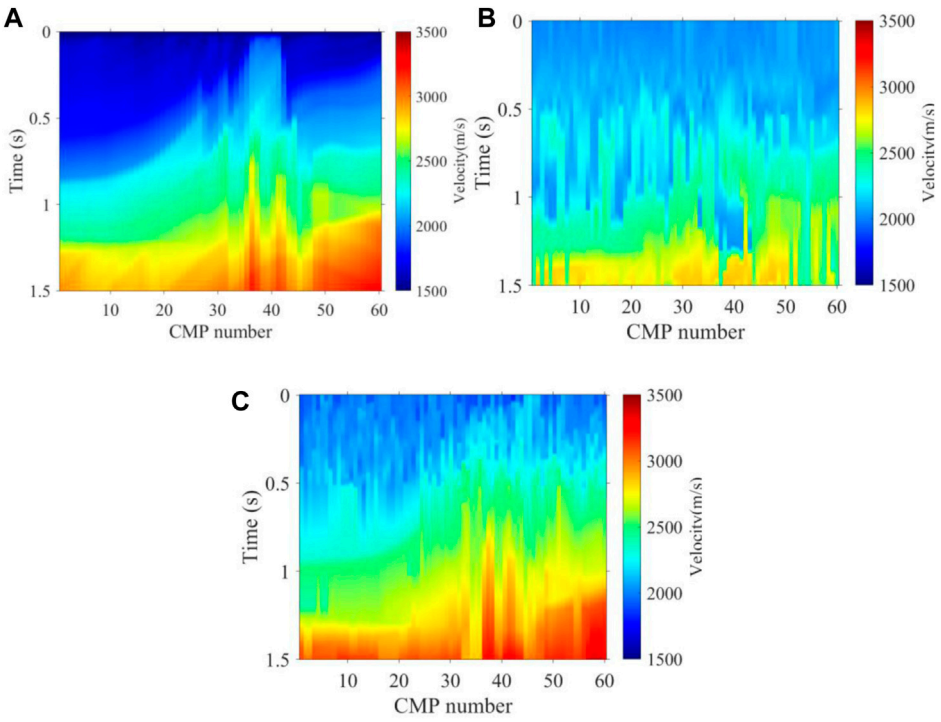


FIGURE 10
Stacking velocity field of synthetic data with random noise and strong energy regular noise. **(A)** True velocity, **(B)** K-means intelligent picking velocity, and **(C)** intelligent picking velocity by the method proposed.

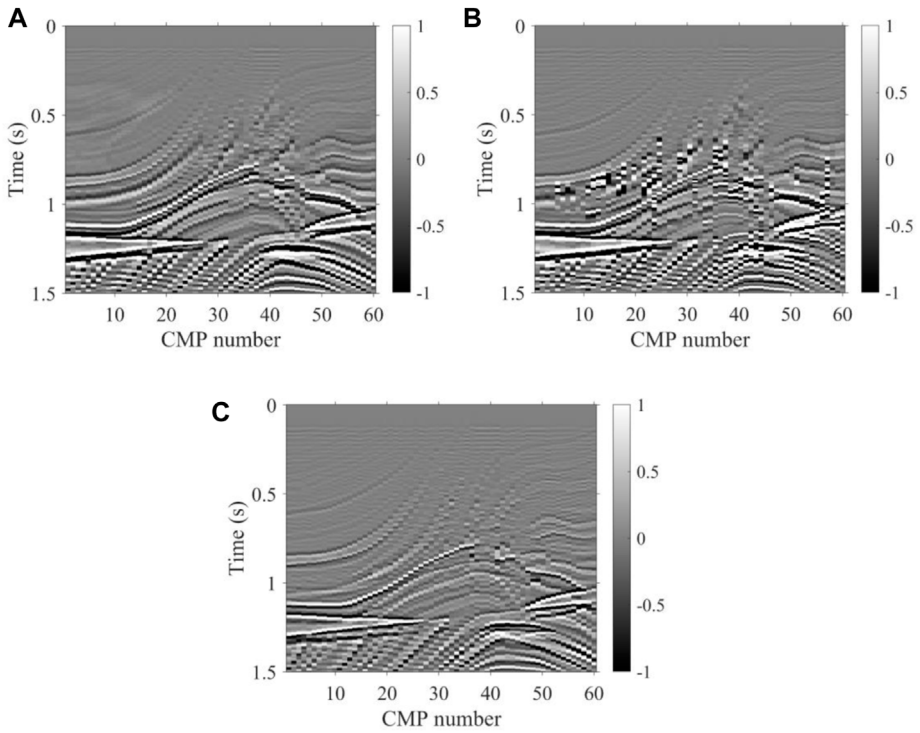


FIGURE 11
Stacking profiles of synthetic data with random noise and strong energy regular noise. **(A)** True velocity, **(B)** velocity of K-means, and **(C)** velocity of the method proposed.

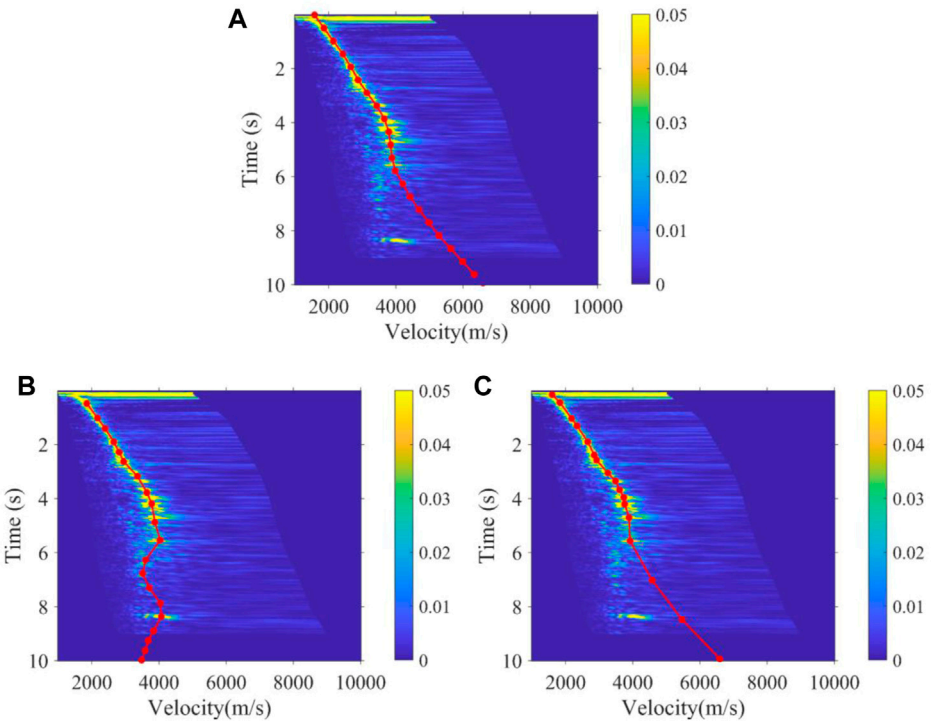


FIGURE 12
Velocity function of the 3500th velocity spectrum obtained by **(A)** manual, **(B)** K-means, and **(C)** the method proposed.

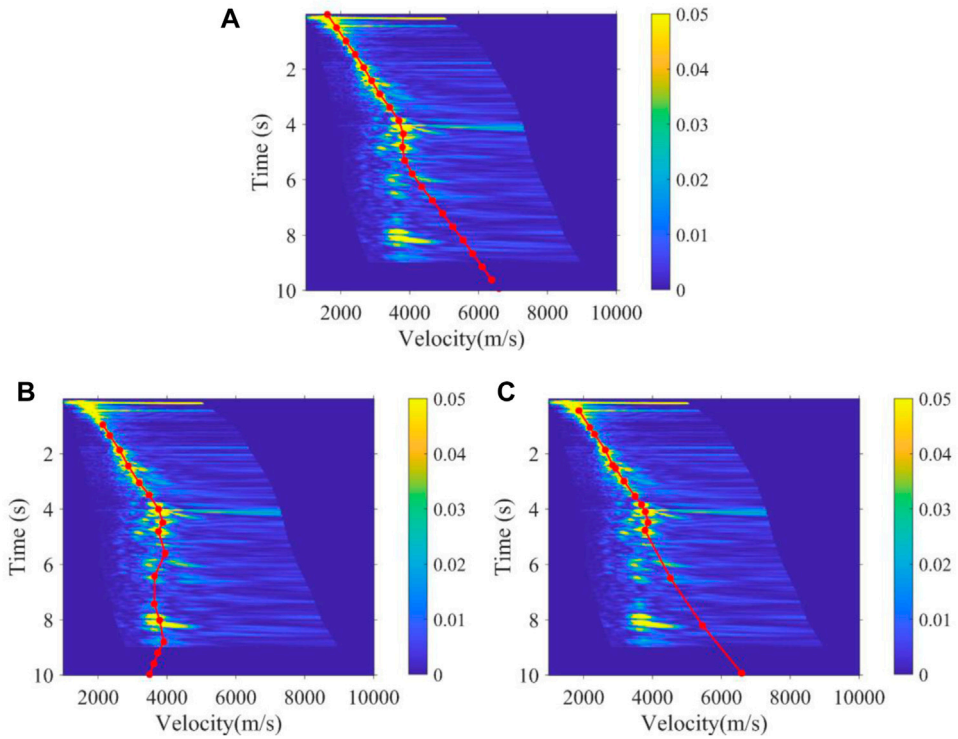


FIGURE 13
Velocity function of the 4300th velocity spectrum obtained by **(A)** manual, **(B)** K-means, and **(C)** the method proposed.

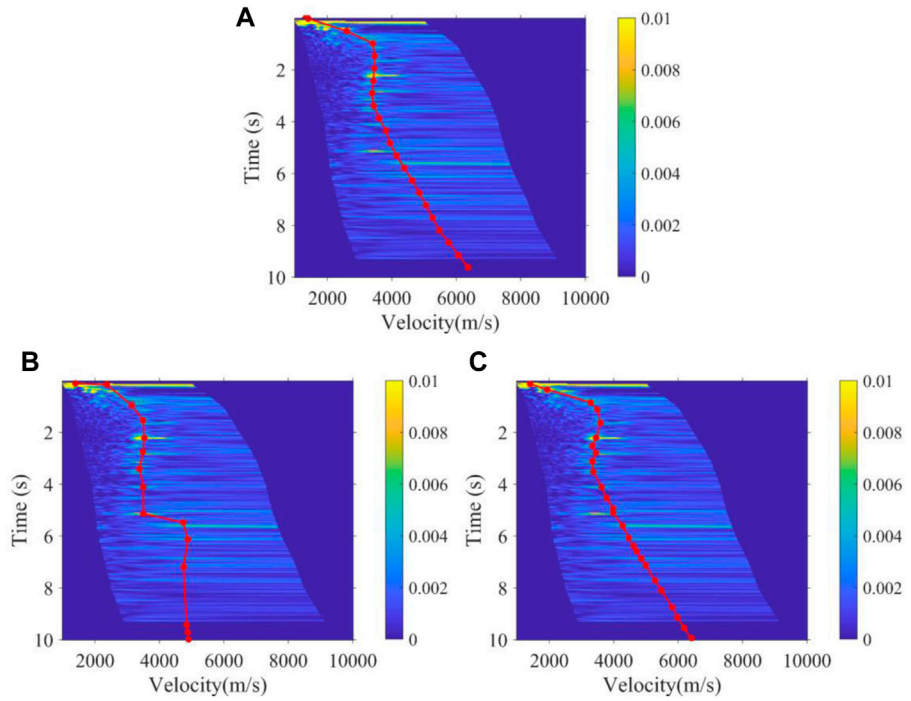


FIGURE 14
Velocity function of the 100th velocity spectrum obtained by **(A)** manual, **(B)** K-means, and **(C)** the method proposed.

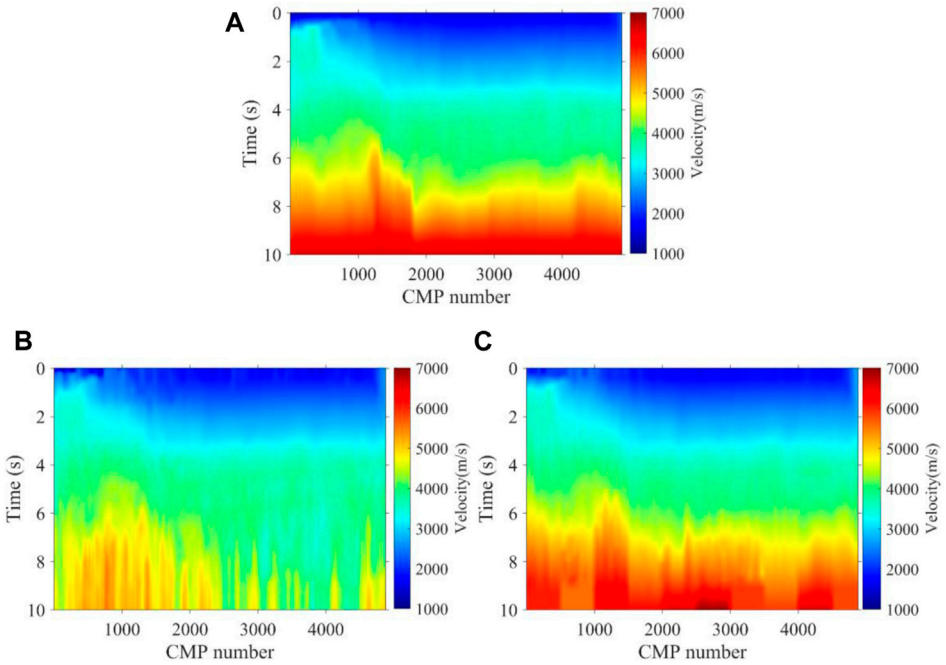


FIGURE 15
Velocity field of a 2D real data established by **(A)** manual, **(B)** K-means, and **(C)** the method proposed.

60 velocity spectra. Every 10 velocity spectra are constrained by an expert experience. In general, the two-dimensional stacking velocity field obtained by our method is basically consistent with the real velocity field shown in Figure 7A on the tectonic trend, while the one with K-means contains a lot of background noise and incorrect construction. Compared to the two-dimensional stacking velocity

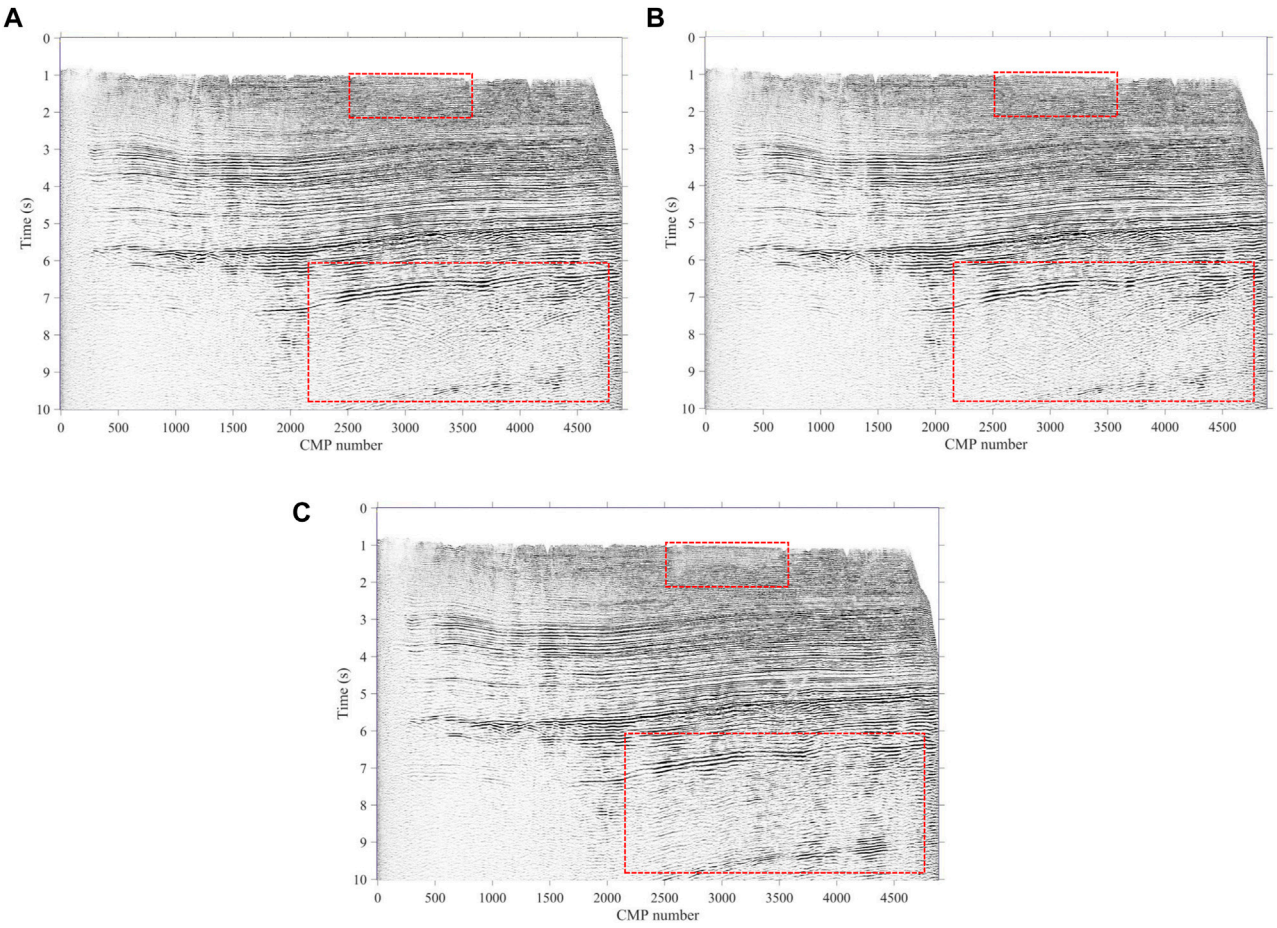


FIGURE 16
Stacking profiles of **(A)** manual, **(B)** method proposed, and **(C)** K-means.

field of K-means, our method is more continuous and stable, since almost all the velocity spectra energy clusters are picked correctly just like the real velocity spectra. In detail, local structures such as a deep high-velocity body and shallow weak reflection interface in the velocity field are also well portrayed by using our method. In seismic data processing, the higher the accuracy of the velocity field, the more realistic the stacked tectonic profile obtained is, and then the subsurface structure is reflected more accurately. In addition, we made an NMO stack using seismic records and the three velocity fields in Figure 7 for further comparison in the tectonic profile, and the stacked tectonic profiles are shown in Figure 8. Compared to the K-means profile, the profile obtained by our method is closer to the real structure shown in Figure 8A, and hence both the shallow and deep structures' imaging accuracy of our method is higher than the K-means. With regards to efficiency, compared to the time required for manual picking, our method is very efficient. It takes only approximately 1 s to pick up one velocity spectrum, which significantly improves the efficiency similar to the K-means method, but our method has a highly accurate structure imaging result.

In order to test the noise immunity of our proposed method, we tested the 2D Marmousi model, from which the seismic records have 30% random noise, and the velocity spectra from the seismic records had 30% random noise and, additionally, strong energy regular noise rotated by valid energy clusters, as shown in Figure 9A, B. Compared to the velocity field of K-means, as shown in Figure 10B, the velocity field obtained by our method shown in Figure 10C is more consistent with the real velocity shown in Figure 10A on the overall trend; therefore, we can see that the K-means velocity has a large error both in the shallow and deep regions of the velocity field. In addition, we made an NMO stack using seismic records and the three velocity fields of Figure 10 for further comparison from the tectonic profile. As shown in Figure 11, the stack profile obtained by our method still has a higher accuracy in structure imaging than the K-means, especially for the deep region. The test results show that our method could obtain better noise immunity. This means that our method is better adapted to real noise-bearing seismic data in the field.

3.2 Real data

To further verify the applicability of the method we proposed, 2D land data with 4,880 velocity spectra (CMPs) were first picked up manually and then picked up intelligently by using K-means and our method. There are often many energy clusters in the CMP velocity spectrum of seismic data. In fact, these energy clusters are generated by effective primary reflection seismic waves and interference multiple reflection waves, each energy cluster corresponding to one velocity

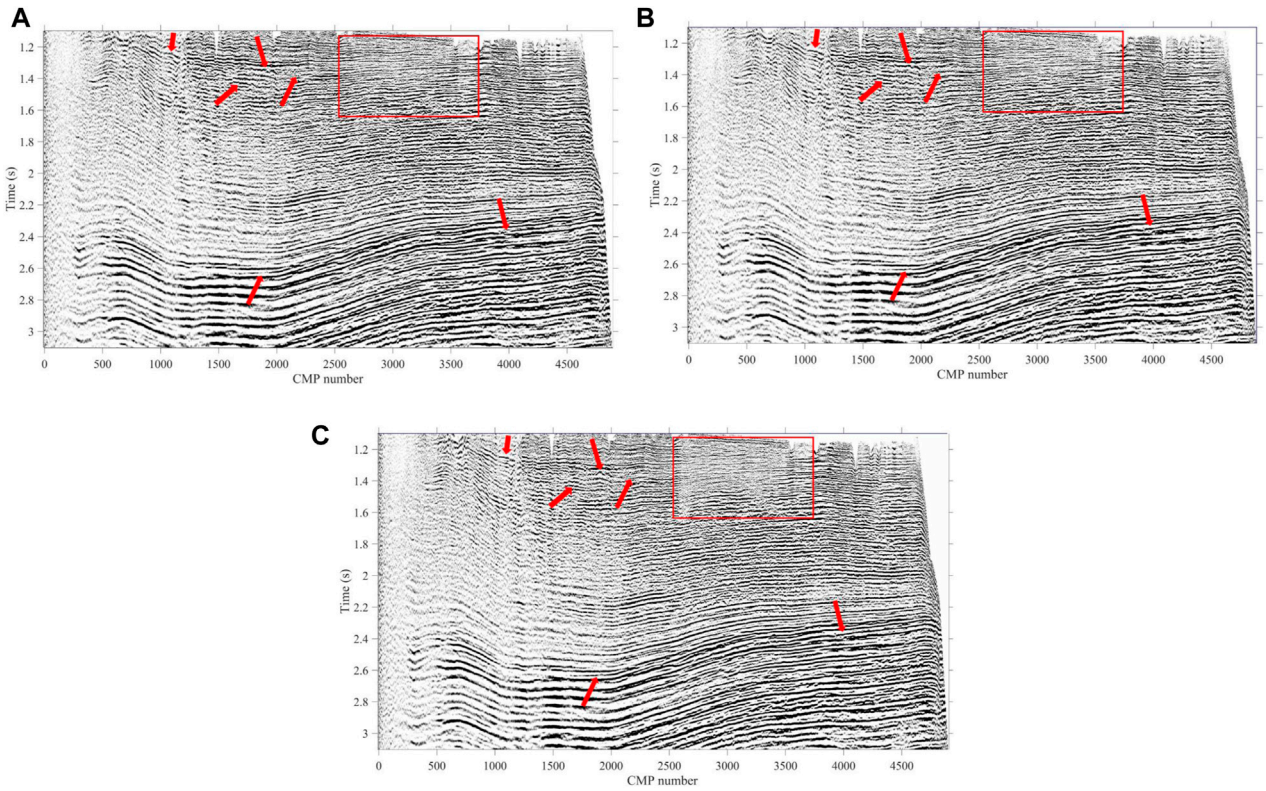


FIGURE 17
Local enlarged stacking profiles of **(A)** manual, **(B)** method proposed, and **(C)** K-means.

value at that time. Compared to the energy clusters of primary reflection waves, the interference energy clusters generated by these multiple waves are often located at a lower velocity (near the left side of the velocity spectrum). In the process of manual velocity picking by the processors, it is necessary to identify and discard the energy clusters generated by multiple reflection waves to avoid false structures in the subsequent velocity field and stack profile. Compared to traditional unsupervised clustering methods, such as K-means, our method introduces expert experience to evade the energy clusters in the deep region of the velocity spectrum generated by the multiples in the real data so as to achieve the accuracy of manual picking, while the K-means incorrectly picks up the energy clusters of multiples and makes some false structures on the subsequent velocity field and stack tectonic profile. The 3900th velocity spectrum with multiples developing at 6–9 s shown in Figure 4A is one of the 4,880 velocity spectra. In fact, after the 3500th velocity spectrum, this real data has obvious multiple interference signals. Expert experience constraints are performed every 500 velocity spectrums when we performed the intelligent velocity picking method we proposed.

As shown in Figure 12 and Figure 13, the 3500th and 4300th velocity spectrum of the land data are processed by experts, the K-means method, and the method we proposed. Obviously, our method can effectively avoid the interference of multiples, which conforms to the trend of manual picking, while the K-means regard the multiples as effective signals to be picked up. Figure 14 shows the picking results of the 100th velocity spectrum (CMP). The signal-to-noise ratio of the velocity spectrum is low, and it is difficult to see effective energy clusters. However, our method can maintain consistency with the trend of manual results, as shown in Figure 14A, due to the expert experience constraints, while the K-means results are abnormal. Figure 15 shows the velocity field of this 2D real data constructed by manual picking, K-means, and our method. It is found that the velocity field established by K-means is significantly different from the manual velocity field, and its accuracy is seriously affected by the interference multiples. However, the velocity field established by our method has the same trend and structure as the manual one both in the shallow and deep regions, and at the same time, the efficiency of our method is much faster than that of manual picking, as it only takes about 1 s to pick up one velocity spectrum while manual picking takes at least 30 s. As a whole, compared to unsupervised clustering methods, such as K-means, our method can better replace experts to pick up the velocity spectrum, which improves efficiency and frees manpower, while meeting the picking accuracy.

In addition, for a more intuitive comparison, we made an NMO stack with the three velocity fields in Figure 15 for further comparison in the tectonic profile, and as shown in Figure 16C, the subsurface structure imaging profile of K-means has fuzzy structures in the shallow red frame and false structures caused by multiples in the deep red frame. However, the profiles of our method and manual are the same as each other, both in the shallow and mid-deep layers, and the subsurface imaging results of them can provide references for subsequent geological understanding. Moreover, since our method can pick up velocity spectra one by one owing to the significantly decreased picking time, the profile of

our method even has a better imaging performance in the local positions than the manual one at the red arrows in Figure 17A, B, where the weak reflection events are strengthened and are more continuous. For the K-means, its stack tectonic profile loses some important structures, especially in the deep region, as shown in the red frame of Figure 17C. All in all, from the tectonic profile, our method has better image results than those of the manual and K-means; hence, the method we proposed improves efficiency, frees up manpower, and can better replace experts to pick up the velocity spectrum automatically.

4 Conclusion

Up to now, manual velocity picking in seismic data processing was the primary way; however, it is labor-intensive and repetitive. An intelligent velocity picking method considering the expert experience based on the Chan–Vese model and mean-shift clustering is proposed by imitating the process of manual velocity picking so as to improve efficiency and free up manpower. In our method, the valid velocity energy clusters are identified by using the CV model with expert experience constraints, which corresponds to the first step of manual picking. Meanwhile, the clustering of the valid energy clusters corresponds to the second step of manual picking by using the mean-shift clustering method. These two steps translate geophysical and geological theories into the geometry of energy clusters on the velocity spectrum. The theoretical model and actual data test prove that our method has several advantages as can be seen in the following paragraph.

Compared to the manual and K-means, the method we proposed can obtain a highly accurate velocity field and subsurface tectonic imaging profile, which can provide a better reference for subsequent geological understanding, while the K-means method always falls into the wrong picking result. In terms of efficiency, our method takes only approximately 1s to pick up one velocity spectrum similar to other automatic methods such as K-means, while manual picking takes at least 30 s. All in all, our method can replace manual velocity picking by experts, which improves efficiency, frees up manpower, and enhances picking accuracy.

Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

Author contributions

L-DW is the main author of the research results; JW and X-RX provided auxiliary support; and H-HZ, YG, and W-QL applied and tested the results.

Funding

This study was supported by the science and technology project of PetroChina: Research and Field Tackling Test of Artificial Intelligence Seismic Reservoir Carving Technology (2022KT1501) and Subsalt Seismic Imaging Tackling and Target Implementation in the Cambrian Subsalt System at the Periphery of the Hotan River and Tarim Basin (2022KT0506).

Conflict of interest

The authors L-DW, JW, X-RX, H-HZ and W-QL were employed by the Research Institute of Petroleum Exploration & Development-Northwest (NWGI), PetroChina.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Aghamohammadi, A., Ranjbarzadeh, R., Naiemi, F., Mogharrebi, M., Dorosti, S., and Bendechache, M. (2021). Tpcnn: Two-path convolutional neural network for tumor and liver segmentation in CT images using a novel encoding approach. *Expert Syst. Appl* 183, 115406. doi:10.1016/j.eswa.2021.115406

Agudelo, W., Villa, Y., Duarte, C., and Sierra, D. (2017). Seismic attribute selection and clustering to detect and classify surface waves in multicomponent seismic data by using k-means algorithm. *Lead. edge* 36 (3), 239–248. doi:10.1190/tle36030239.1

Araya-Polo, M., Dahlke, T., Frogner, C., Zhang, C., Poggio, T., and Hohl, D. (2017). Automated fault detection without seismic processing. *Lead. edge* 36 (3), 208–214. doi:10.1190/tle36030208.1

Biswas, R., Vassiliou, A., Stomberg, R., and Sen, M. K. (2019). Estimating normal moveout velocity using the recurrent neural network. *Interpretation* 7 (4), T819–T827. doi:10.1190/int-2018-0243.1

Cameron, M., Fomel, S., and Sethian, J. (2008). Time-to-depth conversion and seismic velocity estimation using time-migration velocity. *Geophysics* 73 (5), 205–210. doi:10.1190/1.2967501

Cao, W., Guo, X. B., Tian, F., Ying, S., Wei-Hong, W., and Hong-Ri, S., (2022). Seismic velocity inversion based on CNN-LSTM fusion deep neural network. *Appl. Geophys* 18 (4), 499–514. doi:10.1007/S11770-021-0913-3

Chen, Y. Q., (2018). Automatic semblance picking by a bottom-up clustering method. *SEG Maximizing Asset Value Through Artif. Intell. Mach. Learn*, 44–48. doi:10.1190/AIML2018-12.1

Comaniciu, D., and Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. *IEEE Trans Pattern Analysis Mach. Intell* 24 (5), 603–619. doi:10.1109/34.1000236

Fabien-Ouellet, G., and Sarkar, R. (2020). Seismic velocity estimation: A deep recurrent neural-network approach. *Geophysics* 85 (1), U21–U29. doi:10.1190/geo2018-0786.1

Getreuer, P. (2012). Chan-vese segmentation. *Image Process. Line* 2, 214–224. doi:10.5201/ipol.2012.g-cv

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. The MIT Press. Cambridge, MA, USA, doi:10.1007/s10710-017-9314-z

Hou, A., and Marfurt, K. J. (2002). Multicomponent prestack depth migration by scalar wavefield extrapolation. *Geophysics* 67 (6), 1886–1894. doi:10.1190/1.1527088

Jones, I. F., Ibbotson, K., Grimshaw, M., and Plasterie, P. (1998). 3-D prestack depth migration and velocity model building. *Lead. Edge* 17 (7), 897–906. doi:10.1190/1.1438063

Keegan, M. S., Sandberg, B., and Chan, T. (2017). A multiphase logic framework for multichannel image segmentation. *Inverse Problems Imaging* 6 (1), 95–110. doi:10.1109/34.537343

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521 (7553), 436–444. doi:10.1038/nature14539

Li, Q., Pan, Z. K., Wei, W. B., and Song, T. T. (2021). Fast segmentation methods for convex relaxation chan-vese model. *Comput. Simul* 38 (06), 226–232. doi:10.3969/j.issn.1006-9348.2021.06.047

Martin, G. S., Wiley, R., and Marfurt, K. J. (2006). Marmousi2: An elastic upgrade for Marmousi. *Lead. Edge* 25 (2), 156–166. doi:10.1190/1.2172306

Nemeth, T., Wu, C. J., and Schuster, G. T. (1999). Least-squares migration of incomplete reflection data. *Geophysics* 64 (1), 208–221. doi:10.1190/1.1444517

Osher, S., and Sethian, J. A. (1988). Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations. *J. Comput. Phys* 79 (1), 12–49. doi:10.1016/0021-9991(88)90002-2

Park, M. J., and Sacchi, M. D. (2020). Automatic velocity analysis using convolutional neural network and transfer learning. *Geophysics* 85 (1), V33–V43. doi:10.1190/geo2018-0870.1

Ranjbarzadeh, R., Bagherian, K. A., Jafarzadeh, G. S., Anari, S., Naseri, M., and Bendechache, M. (2021). Brain tumor segmentation based on deep learning and an attention mechanism using MRI multi-modalities brain images. *Sci. Rep* 11, 10930. doi:10.1038/s41598-021-90428-8

Ranjbarzadeh, R., Tataei, S. N., Jafarzadeh, G. S., Saleh Esfahani, M., Parhizkar, M., and Pourasad, Y., (2022). MRFE-CNN: Multi-route feature extraction model for breast tumor segmentation in Mammograms using a convolutional neural network. *Ann. Oper. Res.* doi:10.1007/s10479-022-04755-8

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature* 323 (6088), 533–536. doi:10.1038/323533a0

Toldi, J. L. (1989). Velocity analysis without picking. *Geophysics* 54 (2), 191–199. doi:10.1190/1.1442643

Velis, D. (2021). Simulated annealing velocity analysis: Automating the picking process. *Geophysics* 86 (2), V119–V130. doi:10.1190/geo2020-0323.1

Waheed, U. B., Al-Zahrani, S., and Hanafy, S. M. (September, 2019). Machine learning algorithms for automatic velocity picking: K-Means vs. DBSCAN. Proceedings of the SEG Technical Program Expanded Abstracts. San Antonio, Tx, USA,

Wang, D., Yuan, S. Y., Liu, T., Li, S. J., and Wang, S. X. (2021a). Inversion-based non-stationary normal moveout correction along with prestack high-resolution processing. *J. Appl. Geophys* 191, 104379. doi:10.1016/j.jappgeo.2021.104379

Wang, D., Yuan, S. Y., Yuan, H., Zeng, H. H., and Wang, S. X. (2021b). Intelligent velocity picking based on unsupervised clustering with the adaptive threshold constraint. *Chin. J. Geophys* 64 (3), 1048–1060. doi:10.6038/cjg2021O0305

Wang, W. L., McMechan, G. A., Ma, J. W., and Xie, F. (2020). Automatic velocity picking from semblances with a new deep-learning regression strategy: Comparison with a classification approach. *Geophysics* 86 (2), U1–U13. doi:10.1190/geo2020-0423.1

Wang, X., Liu, H., and Ma, W. L. (2018). Sparse least squares support vector machines based on Meanshift clustering method. *IFAC-PapersOnLine* 51 (18), 292–296. doi:10.1016/j.ifacol.2018.09.315

Wang, X. W., Gao, Y., Chen, C., Yuan, H., and Yuan, S. (2022). Intelligent velocity picking and uncertainty analysis based on the Gaussian Mixture Model. *Acta Geophys* 70, 2659–2673. doi:10.1007/s11600-022-00859-8

Wilson, H., and Gross, L. (2019). Reflection-constrained 2D and 3D non-hyperbolic moveout analysis using particle swarm optimization. *Geophys. Prospect.* 67, 550–571. doi:10.1111/1365-2478.12758

Yuan, S. Y., Jiao, X. Q., Luo, Y. N., Sang, W. J., and Wang, S. X. (2022). Double-scale supervised inversion with a data-driven forward model for low-frequency impedance recovery. *Geophysics* 87 (2), R165–R181. doi:10.1190/geo2020-0421.1

Yuan, S. Y., Wang, S. X., Luo, Y. N., Wei, W. W., and Wang, G. C. (2019). Impedance inversion by using the low-frequency full-waveform inversion result as an a priori model. *Geophysics* 84 (2), R149–R164. doi:10.1190/geo2017-0643.1

Zhang, H., Zhu, P. M., Gu, Y., and Li, X. Z. (September, 2019). Automatic velocity picking based on deep learning. Proceedings of the SEG Technical Program Expanded Abstracts, San Antonio, Tx, USA, 2604–2608. doi:10.1190/segam2019-3215633.1

Zhang, P., Lu, W., and Zhang, Y. (2015). Velocity analysis with local event slopes related probability density function. *J. Appl. Geophys* 123, 177–187. doi:10.1016/j.jappgeo.2015.10.010

frontiers | Frontiers in Earth Science

TYPE Original Research
PUBLISHED 25 January 2023
DOI 10.3389/feart.2022.1037877

OPEN ACCESS

EDITED BY
Sanyi Yuan,
China University of Petroleum, Beijing, China

REVIEWED BY
Naihao Liu,
Xi'an Jiaotong University, China
Lihua Fu,
China University of Geosciences, Wuhan, China

*CORRESPONDENCE
Xing-Rong Xu,
xu_xr@petrochina.com.cn

SPECIALTY SECTION
This article was submitted to Environmental Informatics and Remote Sensing, a section of the journal Frontiers in Earth Science

RECEIVED 06 September 2022
ACCEPTED 25 November 2022
PUBLISHED 25 January 2023

CITATION
Wang D-Y, Xu X-R, Zeng H-H, Sun J-Q, Xu X and Zhang Y-K (2023), Self-adaptive seismic data reconstruction and denoising using dictionary learning based on morphological component analysis.
Front. Earth Sci. 10:1037877.
doi: 10.3389/feart.2022.1037877

Self-adaptive seismic data reconstruction and denoising using dictionary learning based on morphological component analysis

De-Ying Wang[1], Xing-Rong Xu[1]*, Hua-Hui Zeng[1], Jia-Qing Sun[1], Xin Xu[1] and Yi-Kui Zhang[2]

[1]Research Institute of Petroleum Exploration & Development-Northwest (NWGI), PetroChina, Lanzhou, China, [2]Wuhua Energy Technology Co., Ltd., Xi An, China

Data reconstruction and data denoising are two critical preliminary steps in seismic data processing. Compressed Sensing states that a signal can be recovered by a series of solving algorithms if it is sparse in a transform domain, and has been well applied in the field of reconstruction, when, sparse representation of seismic data is the key point. Considering the complexity and diversity of seismic data, a single mathematical transformation will lead to incomplete sparse expression and bad restoration effects. Morphological Component Analysis (MCA) decomposes a signal into several components with outstanding morphological features to approximate the complex internal data structure. However, the representation ability of combined dictionaries is constrained by the number of dictionaries, and cannot be self-adaptively matched with the data features. Dictionary learning overcomes the limitation of fixed base function by training dictionaries that are fully suitable for processed data, but requires huge amount of time and considerable hardware cost. To solve the above problems, a new dictionary library (K-Singluar Value Decomposition learning dictionary and Discrete Cosine Transform dictionary) is hereby proposed based on the efficiency of fixed base dictionary and the high precision of learning dictionary. The self-adaptive sparse representation is achieved under the Morphological Component Analysis framework and is successfully applied to the reconstruction and denoising of seismic data. Real data tests have proved that the proposed method performs better than single mathematical transformation and other combined dictionaries.

KEYWORDS

compressed sensing, dictionary learning, morphological component analysis, seismic data reconstruction, seismic data denoising

1 Introduction

The data required for seismic exploration are large and complex, and limited by terrain conditions and the economic cost, seismic data are often irregularly sampled, which may be caused by the complex terrain constraints, including buildings, lakes and forbidden areas for land acquisition (Sun et al., 2021). The inadequate excitation of the artificial source may also result in bad traces, thereby leading to irregular, incomplete and alias frequency of seismic data (Zhang et al., 2019). In order to meet the higher requirements for seismic data quality in subsequent processing, the most direct and effective method is to supplement acquisition and encryption of missing seismic traces (Huo et al., 2013; Zhang et al., 2017). However, restricted by economic cost, it is difficult to re-collect data. Therefore, it is necessary to use effective seismic data reconstruction methods to reconstruct the missing seismic traces in the later period. On the other hand, with the development of seismic exploration to more complex areas and deeper layers, the high elevation difference of topographic relief or the lateral change of near-surface velocity lead to shot excitation and poor receiver conditions, and the random noise interference in the collected single shot records is very strong. If these noises cannot be effectively removed, the migration imaging quality and reservoir prediction accuracy will be affected.

There are four main types of seismic data reconstruction methods. One is the filter-based method (Zhang and Tong, 2003), which using interpolation filtering. However, it usually treats the non-uniform grid sampling data as regular data, which leads to large errors. The second is wave field continuation method (Naghizadeh, 2010), which makes full use of underground information. However, the unknown prior information such as underground structure limits the application of this method. Third, the fast rank reduction method (Gao et al., 2013; Ma et al., 2013), which regards interpolation as an image filling problem, has fast calculation speed and simple parameter setting, but it still has limitations in the reconstruction of irregular missing channels under non-uniform grid sampling and its anti-aliasing ability. The fourth method is the Compressed Sensing (CS) method, which can reconstruct regular and irregular missing seismic traces without any prior information such as underground structures, and has high calculation speed and accuracy. Compressed sensing theory is considered a key method for dealing with the problem of data loss, and the three key factors of CS data recovery are data sparsity, random sampling and optimal reconstruction algorithm (Jiang et al., 2019). Since seismic data are not sparse, it is vital to find suitable sparse dictionaries, so that the coefficients of the signal in the dictionary can remain sparse. Indeed, there are many mathematical transformations used for CS, such as Radon transform (Xue et al., 2014; Tang et al., 2020), Fourier transform (Luo et al., 2015; Wen et al., 2018), Wavelet transform (Cui et al., 2003), Curvelet transform (Zhang et al., 2013; Han et al., 2018; Wang et al., 2018), Seislet transform (Liu et al.,. 2013), etc. (Wang et al., 2021). Seismic data are usually composed of different waves and cannot be fully and effectively represented by a single transformation (Wang et al., 2021). Li et al. proposed the application of Morphological Component Analysis (MCA) to seismic data reconstruction, which separates signals mainly by virtue of the difference between components of different signals (Li et al., 2012). MCA was first proposed to image denoising or restoration and achieved good results. At present, it is widely used in signal denoising, reconstruction, separation, repair and fusion and other fields. Zhou et al. quantitatively evaluated the data reconstruction effect of different sparse dictionary combinations under the framework of MCA, and found that the combination of discrete cosine transform (DCT) and curvelet dictionary is provided with the highest reconstruction accuracy (Zhou et al., 2015). Zhang et al. proposed the combination of the Shearlet and DCT dictionary that can represent seismic data more fully and guarantee more accurate reconstruction data (Zhang et al., 2019). In addition to MCA, many more advanced algorithms have been applied to seismic data processing. In 2014, the Variational Mode Decomposition (VMD) algorithm was first proposed and made a significant achievement in the field of signal decomposition. The VMD is an iterative search for the optimal solution of the variational model to determine what we know about the modes and their corresponding center frequencies and bandwidths. Each mode is a finite bandwidth with a central frequency, and the sum of all modes is the source signal (Dragomiretskiy and Zosso, 2014). Subsequently, many experts and scholars have applied the other decomposition methods to seismic data processing. In 2019, Liu et al. proposed an improved EWT (IEWT) to decompose a non-stationary seismic signal into several IMFs and describe its frequency features. Finally, an adaptive spectrum segmentation using detected boundaries based on the SSR is obtained (Liu et al., 2019). They also first identify the major components of the ground roll adopting the multichannel variational mode decomposition (MVMD), which shows significant improvements compared to the conventional single-channel VMD. Next, separating ground roll and reflections on the selected low-frequency IMFs through a curvelet based blockcoordinate relaxation method (Liu et al., 2021). However, the limitations of the mathematical dictionary remain unchanged. Dictionary learning (DL) is a new representative of interdisciplinary research field, which integrates the theoretical ideas of sparse representation, machine learning, image application and compressed sensing, and is mainly used to solve the problem of dictionary design of sparse representation model. Dictionary learning trains the dictionary according to the characteristics of the processed data, and can get the most adequate dictionary (Wang et al., 2021). K-means algorithm, also known as the clustering

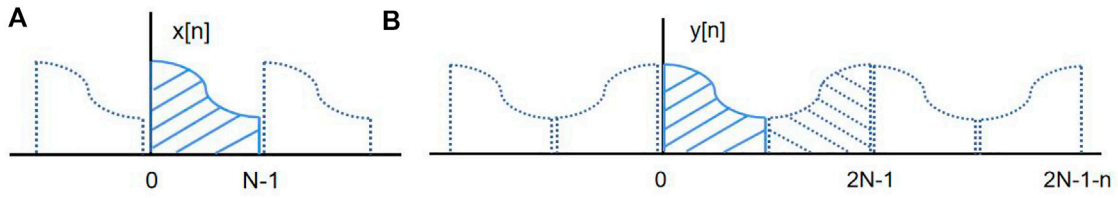


FIGURE 1
Schematic diagram of DFT and DCT. **(A)**DCT; **(B)** DCT.

algorithm, is considered the simplest dictionary learning method, and detects clusters in the sense of least square error by continuously classifying and updating center points. The K-SVD algorithm (SVD: Singluar Value Decomposition) is an extension of the K-means algorithm, which is also carried out by continuously updating the base and classification (Aharon, 2006). Compared with fixed basis functions, dictionary learning is self-adaptive and can achieve better reconstruction and denoising quality, but its application demands huge time cost and hardware requirements. Yu et al. has made important contributions to the development of dictionary learning, and has successively proposed learning by tight frame, Monte Carlo data driven tight frame, fast rank reduction algorithm, etc., which has improved the efficiency and effect of dictionary learning (Yu et al., 2015; Jia et al., 2016; Yu et al., 2016; Yu et al., 2018). In addition to compressed sensing, machine learning has also been well applied to seismic data reconstruction. SegNet network enables first-arrival picking at the same time as seismic data reconstruction, and forges a solid foundation for the development of data reconstruction and first-arrival picking (Yuan et al., 2022). Liu et al. propose propose a wavelet-based residual DL (WRDL) network to reconstruct the incomplete seismic data. It considers not only features in the time domain but also frequency features of seismic data, which obtains good reconstruction results in real data (Liu et al., 2022).

Compressive sensing usually transforms seismic data into sparse domain by some mathematical transformation method, and then designs a filter in the sparse domain for threshold processing, and then performs mathematical inverse transformation, and finally achieves the purpose of effectively removing noise in seismic data. In the case of the denoising method of seismic data, various theoretical methods have been academically proposed to filter out different types of noise. For regular noise, multiple waves can be removed by Radon transform (Shan et al., 2009), side waves by K-L filtering, and surface waves by least square filtering (Vaidyanathan, 1987). The theoretical basis of these methods is the difference between the effective signal and the regular noise in the characteristics such as frequency and propagation direction. Meanwhile, commonly used algorithms for random noise include frequency domain filtering based on Fourier transform, f-x domain prediction denoising (Spitz, 2012), wavelet transform (Jin et al., 2005), etc. Filtering based on Fourier transform can only filter out the random noise of the lowest and the highest frequency band at both ends; the noise will also be enhanced when the effective signal is enhanced in f-x domain prediction denoising (Spitz, 2012); the wavelet transform performs poorly in expressing the edge information of the curve, and is subject to certain limitations in expressing the hyperbolic features. Liu et al. proposed an EWT-based denoising method in 2020 and effectively suppressed noise. Synthetic data and 3D feld data examples also prove the validity and efectiveness of the TFPF-EWT for both attenuating random noise and preserving valid seismic amplitude (Liu et al., 2020). In this case, as with reconstruction, MCA and dictionary learning are also well applied to the field of denoising. Olshausenp et al. proposed the concept of learned dictionary in 1997, and applied overcomplete dictionary to image denoising. As an advanced and effective signal decomposition method, VMD is also well applied to seismic data denoising. Zhang et al. proposed a multi-channel scheme which is referred as the multi-channel variational mode decomposition (MVMD) based on multi-channel singular spectrum analysis (MSSA), to efficiently and effectively separate and attenuate seismic random noises. This method leverage the MSSA for each decomposed IMF to separate and attenuate random noises (Zhang et al., 2021). Lian et al. took the matching pursuit method as a continuation technique of sparse representation method, and obtained good progress (Lian et al., 2015). Subsequently, Chen proposed the basis tracing method to solve the sparse optimization problem (Chen et al.,. 2001), and Olshausen et al. proposed the self-adaptive learning complete dictionary (Olshausen et al., 2000). Tang et al. first applied the learning complete dictionary to the seismic data denoising. After many times of learning and training of the input signal, the dictionary was updated and the sparse representation coefficient was obtained, which achieved better denoising effect than the traditional method. However, the complexity of the seismic data led to long operation time (Tang et al., 2012). Xu proposed to

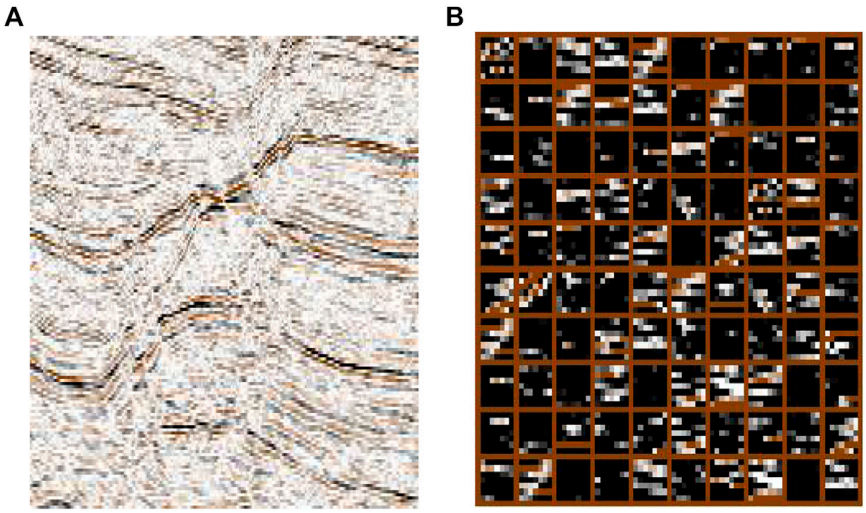


FIGURE 2
Training Set. **(A)** A seismic image; **(B)** Examples of non-flat patches.

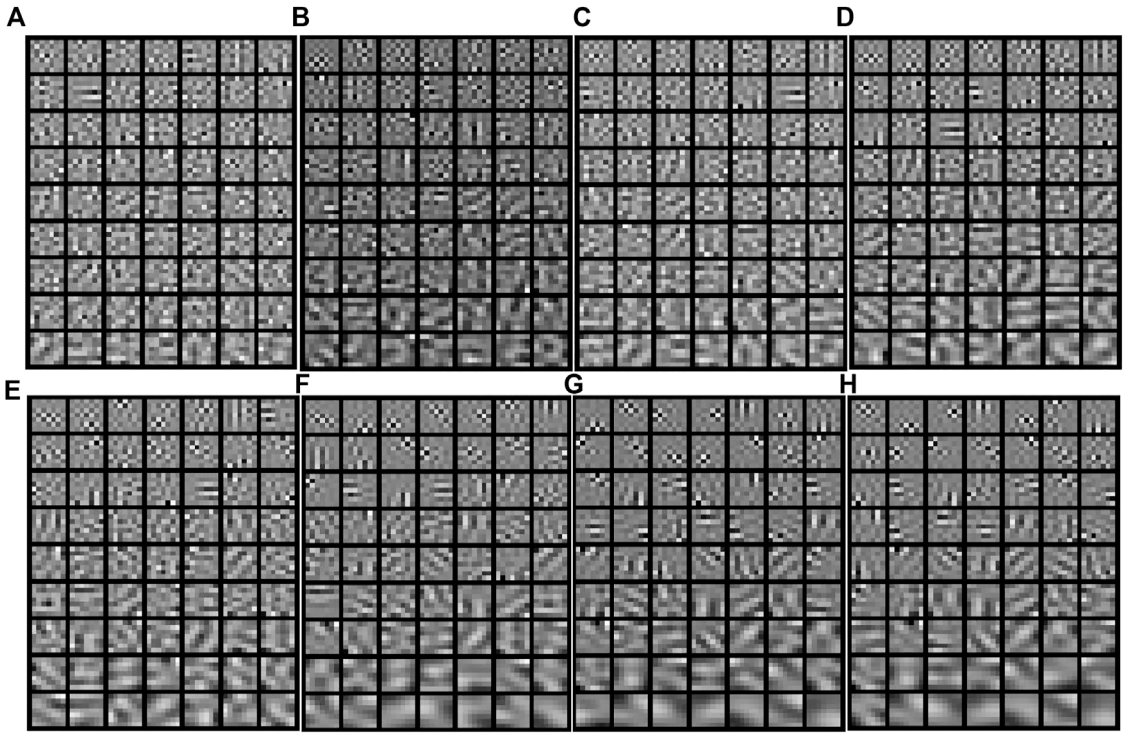


FIGURE 3
Partial dictionary images in different iterations. **(A)** Partial dictionary images in 5th iterations ($E = 17.57$); **(B)** Partial dictionary images in 10th iterations ($E = 15.85$); **(C)** Partial dictionary image in 15th iterations ($E = 14.16$); **(D)** Partial dictionary images in 20th iterations ($E = 10.23$); **(E)** Partial dictionary images in 25th iterations ($E = 6.89$); **(F)** Partial dictionary images in 30th iterations ($E = 3.82$); **(G)** Partial dictionary images in 35th iterations ($E = 1.22$); **(H)** Partial dictionary images in 40th iterations ($E = 0.53$).

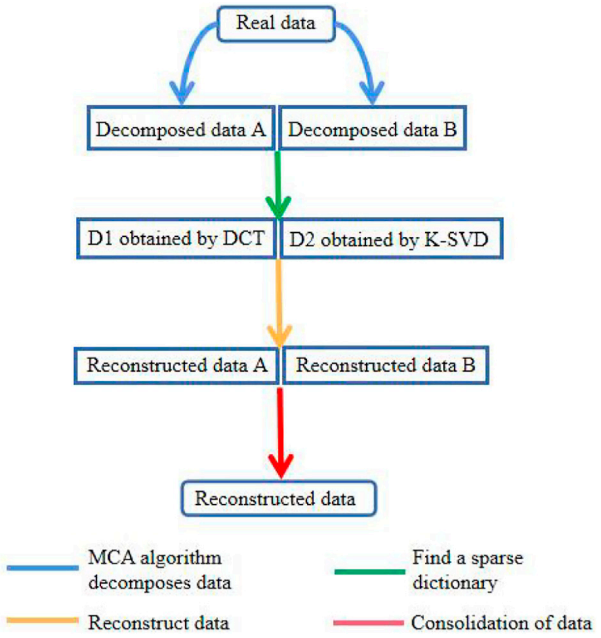


FIGURE 4
Workflow of the proposed method.

replace MOD algorithm in K-SVD algorithm with StOMP (Stagewise Orthogonal Matching Pursuit) algorithm, which not only overcomes the over-matching phenomenon caused by orthogonal matching pursuit (OMP) algorithm, but also significantly improves the convergence speed (Xu et al., 2016). With the development of learning methods, Artificial Intelligence (AI) has been gradually applied to seismic data denoising. Zhang et al. proposed a full convolution denoising network based on residual learning, which can remove various noises at the same time (Zhang et al., 2017). Mao et al. proposed a full convolution network, which uses convolution layer to encode to extract features, and deconvolution layer to decode to recover clean data. In recent years, DCNNs has also achieved good results in suppressing random noise (Sang, 2021).

Aiming at the limitations of dictionary combination and the inefficiency of dictionary learning, this paper comes up with a new dictionary library (K-SVD+DCT) and realizes the self-adaptive reconstruction and denoising of seismic data under the MCA framework. In addition, we can also simultaneously reconstruct and denoise to process missing noisy data. Tests of real data have proved the effectiveness and applicability of the proposed method.

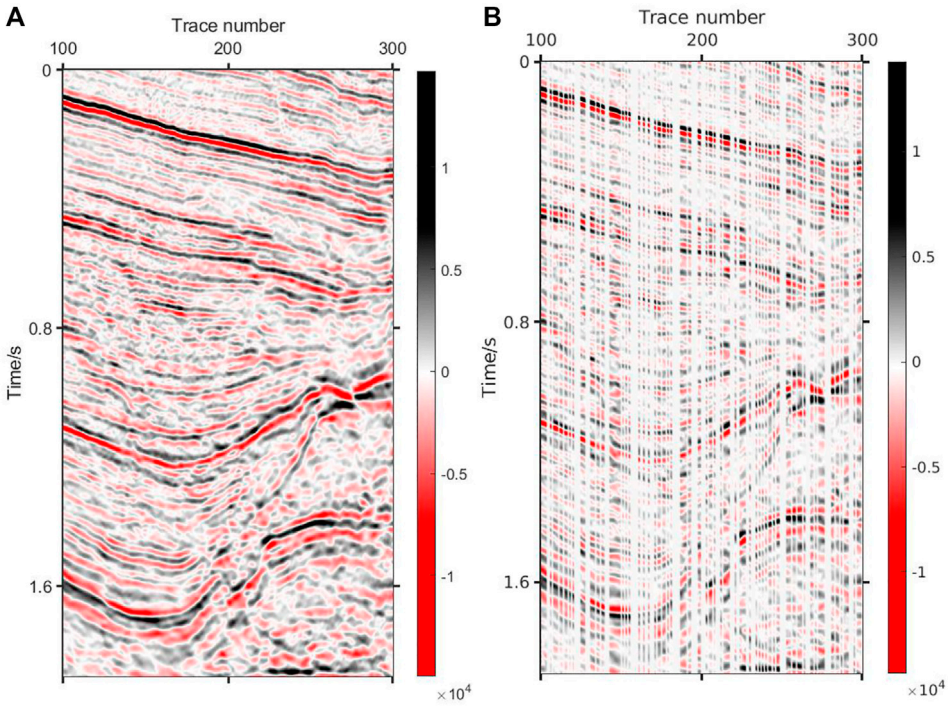


FIGURE 5
Original image and sampled image. **(A)** Original image; **(B)** 50% missing data (SNR=3.03 dB PSNR=19.38 dB).

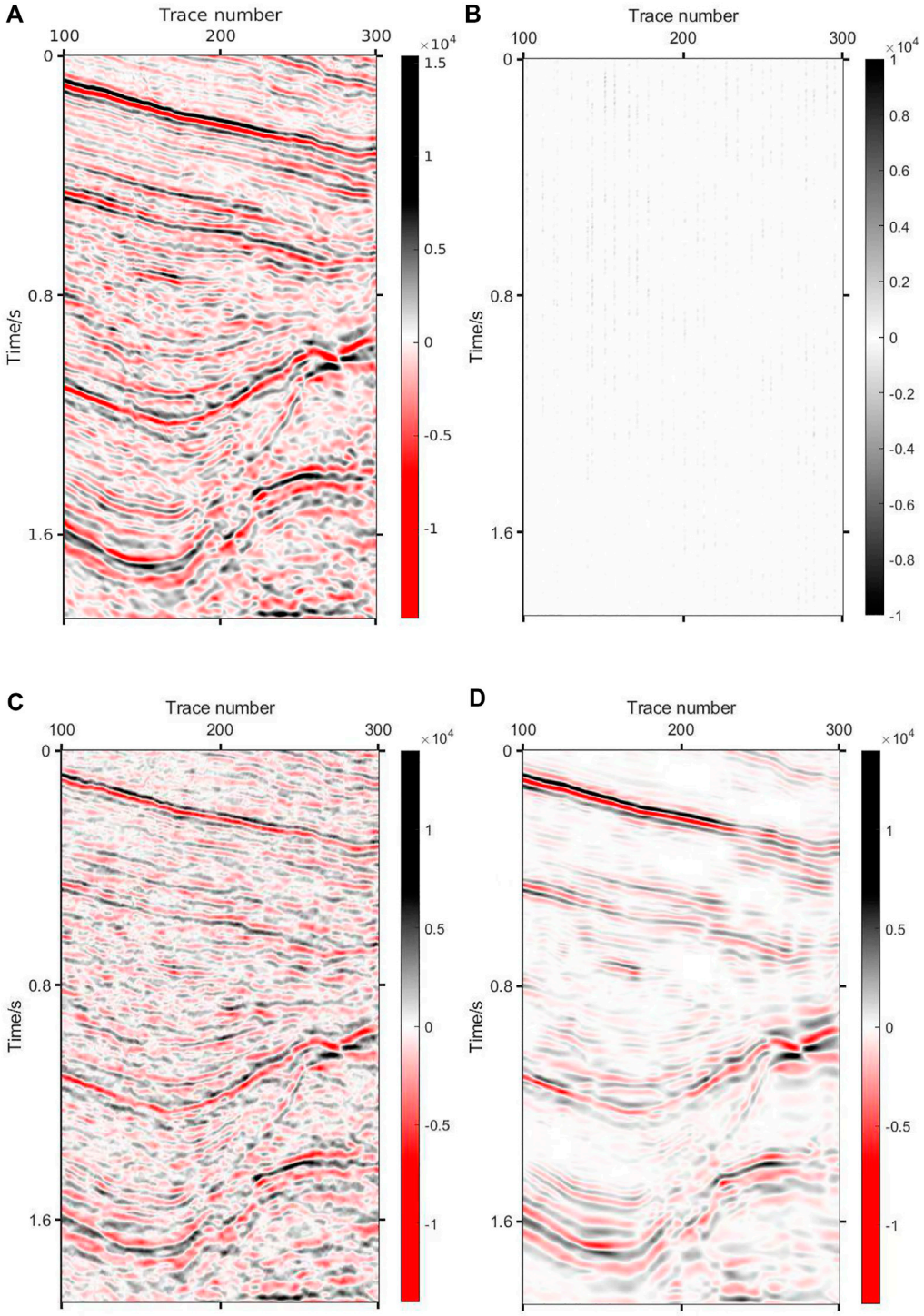


FIGURE 6
Result *via* DCT+K-SVD. **(A)** Result *via* DCT+K-SVD (SNR=12.1 dB PSNR=28.45 dB); **(B)** Absolute error *via* DCT+K-SVD; **(C)** Reconstruction component *via* K-SVD; **(D)** Reconstruction component *via* DCT.

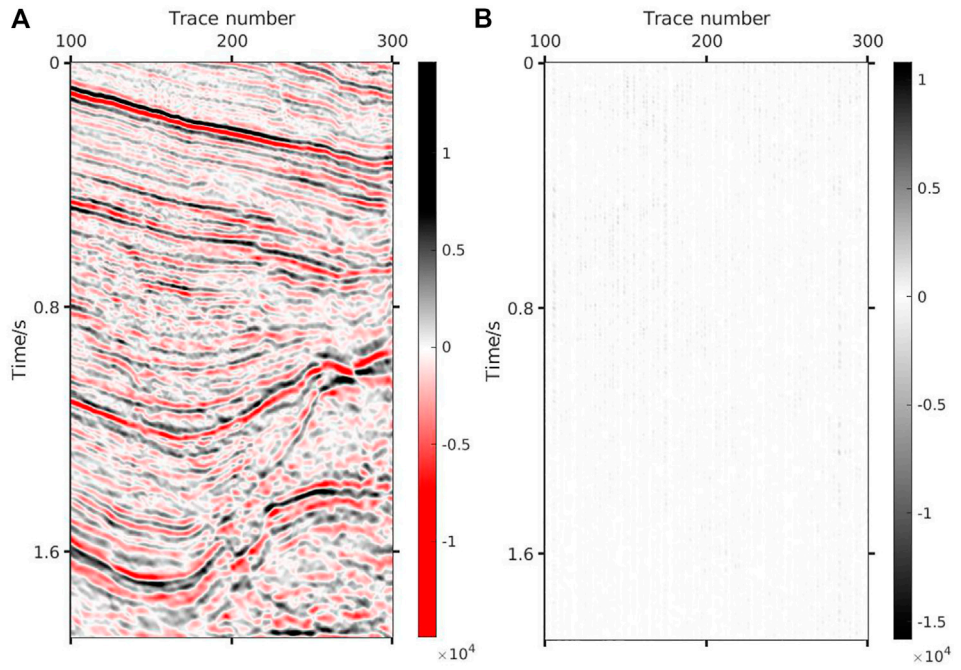


FIGURE 7
Result *via* K-SVD. **(A)** Result *via* K-SVD (SNR=15.88 dB PSNR=33.23 dB); **(B)** Absolute error *via* K-SVD.

2 Theory

2.1 Morphological component analysis (MCA)

Morphological Component Analysis (MCA) uses the individual matching of sparse dictionaries to signal features to achieve signal decomposition. A signal Y consists of K morphological components x_n:

$$Y = \sum_{k=1}^{K} X_n = \sum_{k=1}^{K} D_k \alpha_k \tag{1}$$

where, α_k is the sparse coefficient; D_k represents the sparse dictionary. Due to the incoherence between the various morphological components, the solution can be solved by Eq. 2:

$$\alpha_k = \arg\min_{\alpha} \|\alpha\|_0 \; s.t \; X_k = D_k \alpha \tag{2}$$

Feature selection on data using multiple dictionaries:

$$\{\alpha_1, ..., \alpha_k\} = \underset{\{\alpha_1, L, \alpha_k\}}{\operatorname{argmin}} \sum_{k=1}^{K} \|\alpha\|_1 \; s.t \; Y = \sum_{k=1}^{K} D_k \alpha_k. \tag{3}$$

To facilitate the solution, we convert Eq. 3 as Eq. 4:

$$\{\alpha_1, ..., \alpha_k\} = \underset{\{\alpha_1, L, \alpha_k\}}{\operatorname{argmin}} \sum_{k=1}^{K} \|\alpha_k\|_1 + \lambda \left\| Y - \sum_{k=1}^{K} D_k \alpha_k \right\|_2^2. \tag{4}$$

Considering the purpose of decomposing the signal, the vector $\{\alpha_1, ..., \alpha_k\}$ is transformed into $\{X_1, ..., X_k\}$, representing that the signal contains K morphological components, and each component of the signal is obtained by solving Eq. 5:

$$\{X_1, ..., X_k\} = \underset{\{X_1, L, X_k\}}{\operatorname{argmin}} \sum_{k=1}^{K} \left\| D_k^{-1} X_k \right\|_1 + \lambda \left\| Y - \sum_{k=1}^{K} X_k \right\|_2^2 \tag{5}$$

2.2 The theory of reconstruction based on MCA

MCA believes that a combined dictionary has the sum of the sparse representation capabilities of its combined components. For example, the combined dictionary of Fourier and Wavelet can well describe signals that contain both stationary and localized features. This is more conducive to the full expression of the data and the improvement of the reconstruction quality. The compressed sensing reconstruction process based on MCA is as follows:

A 2D signal X contains K components of different shapes, which is as Eq. 1. The reconstruction of seismic data can be expressed as follows:

$$Y = RX \tag{6}$$

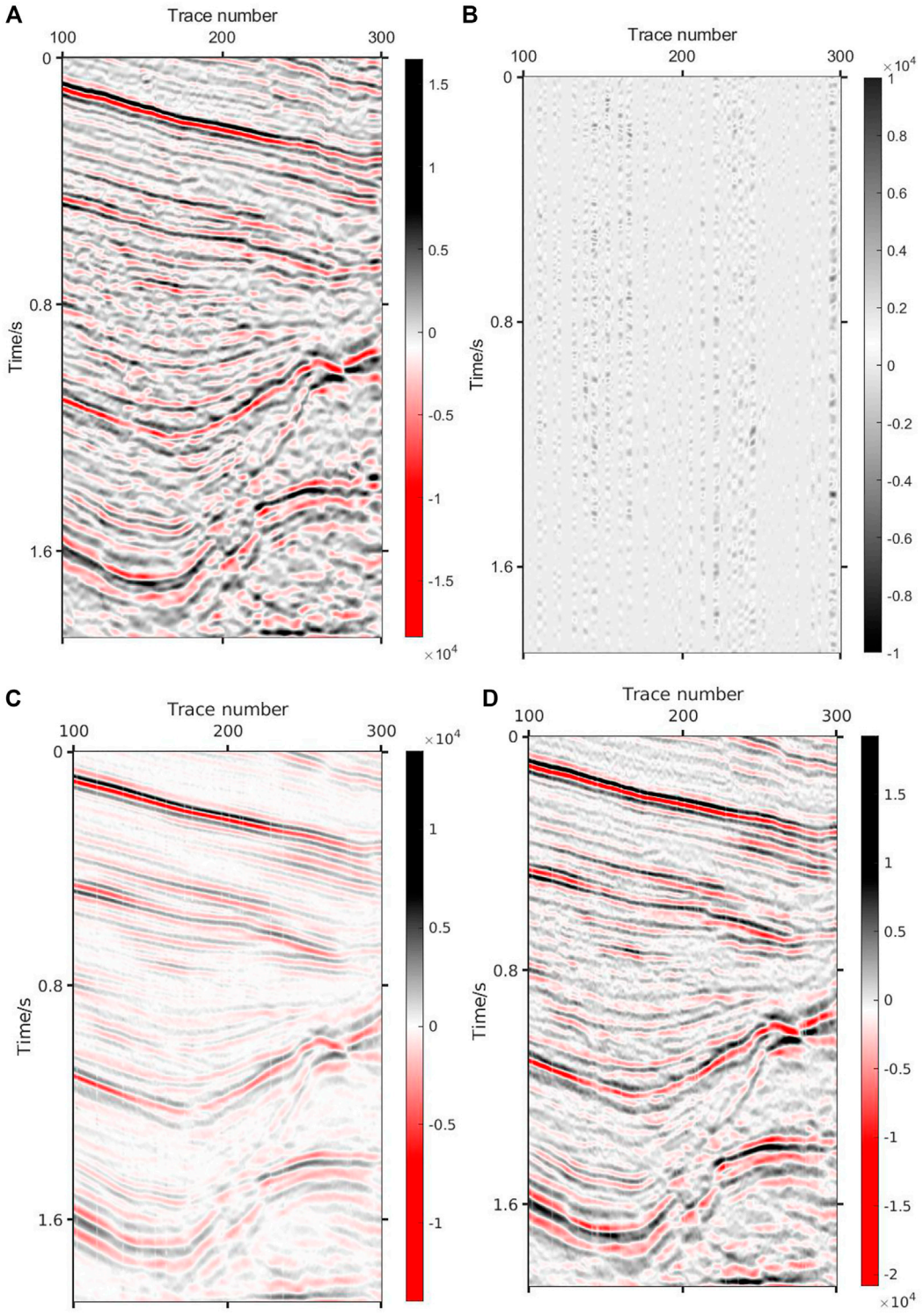


FIGURE 8
Result *via* Curvelet+DCT. **(A)** Result *via* Curvelet+DCT (SNR=11.45 dB PSNR=27.79 dB); **(B)** Absolute error *via* Curvelet+DCT; **(C)** Reconstruction component *via* Curvelet; **(D)** Reconstruction component *via* DCT.

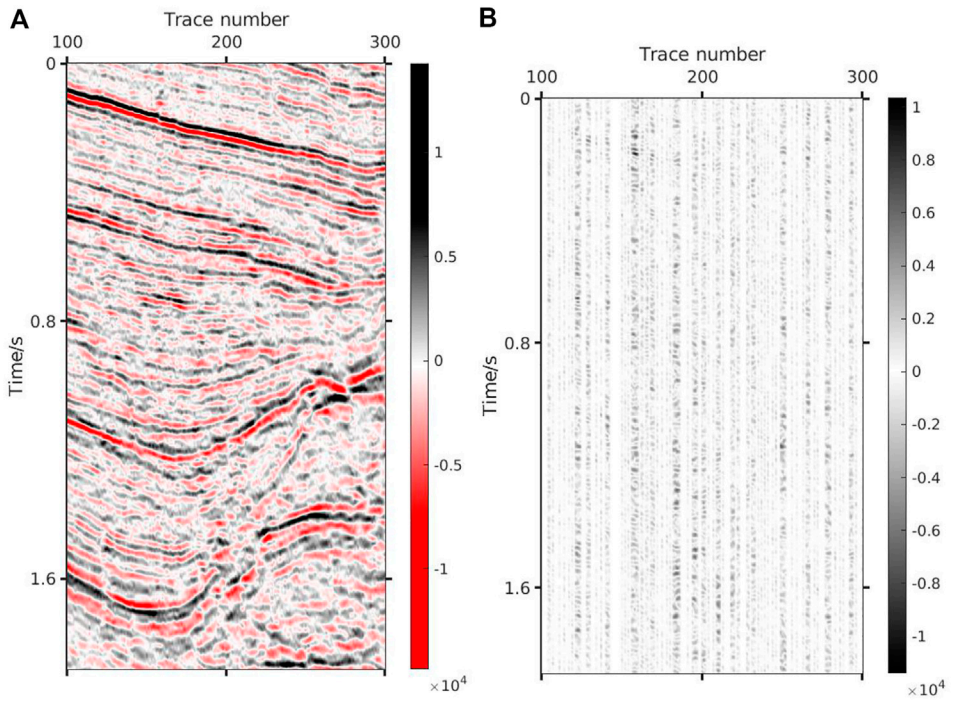


FIGURE 9
Result *via* Shearlet. **(A)** Result *via* Shearlet (SNR=7.59 dB PSNR=23.94 dB). **(B)** Absolute error *via* Shearlet.

where, X stands for ideal complete seismic data; Y represents the collected data with missing traces; R represents the sampling matrix. In the MCA framework, Eq. 1 can be expressed as:

$$Y = RX = R\sum_{k=1}^{K} X_k = R\sum_{k=1}^{K} D_k\alpha_k \tag{7}$$

Above is reformulated as an unconstrained optimization problem:

$$\alpha_k = \arg\ \min \|\alpha_k\|_1 + \lambda\left\|Y - R\sum_{k=1}^{K} D_k\alpha_k\right\|_2, \tag{8}$$

Where, λ is the Lagrange multiplier is used to measure the proportion of $\ell 1$ norm and $\ell 2$ norm.

The reconstruction algorithm's solution, combined with the BCR (Block Coordinate Relaxation, BCR) algorithm, offered the following solution method based on the morphological component. The solution process is:

Input: sample matrix R, the dictionary combination $D = [D_1, \cdots, D_k]$, missing seismic data Y, the total number of iterations N; Output: reconstructed seismic data X'; Initialize: each morphological component $X_i^{(0)} = 0,\ i = 1, \cdots, K$.

1) for: $n = 1$: N;
2) residual $r^{(n)} = Y - R \cdot sum(X_1, \ldots, X_k)$;
3) for: $k = 1$: K;
4) ${\alpha_k}^{(n)} = D_k({X_k}^{(n)}, \ldots, r^{(n)})$;
5) ${X_k}^{(n)} = {D_k}^{-1}T_\lambda({\alpha_k}^{(n)})$;
6) the threshold model is applied to reduce λ;
7) $X' = sum(X_1, \cdots, X_K)$.

where, D^{-1} represents the inverse transformation of the dictionary D; T_λ is the threshold function proposed, which formula is:

$$T(x,\lambda) = X \cdot \exp\left(-\left(\lambda/|x|\right)^{2-p}\right) \tag{9}$$

Besides the threshold λ, we have another independent parameter p, which can be flexibly chosen to achieve better performance. Based on the Taylor series, it is valuable to point out that the exponential shrinkage can be considered a smooth ℓ_0 constraint. For $|x| \geq \lambda$, it is a good approximation of the p-thresholding operator and does not suffer the bias when $p \neq 1$. It reduces to Stein thresholding operator for $p = 0$ and soft thresholding for $p = 1$. In solving the algorithm, it is necessary to adjust the Lagrange multiplier to get the optimal solution. The steps are as follows: First, a larger transform domain coefficient is selected as the threshold value to obtain the sparse approximate solution. Second, the value is continuously reduced to include more transform domain coefficients, and the optimal solution is approximated by continuous iteration. The threshold selection strategy is called

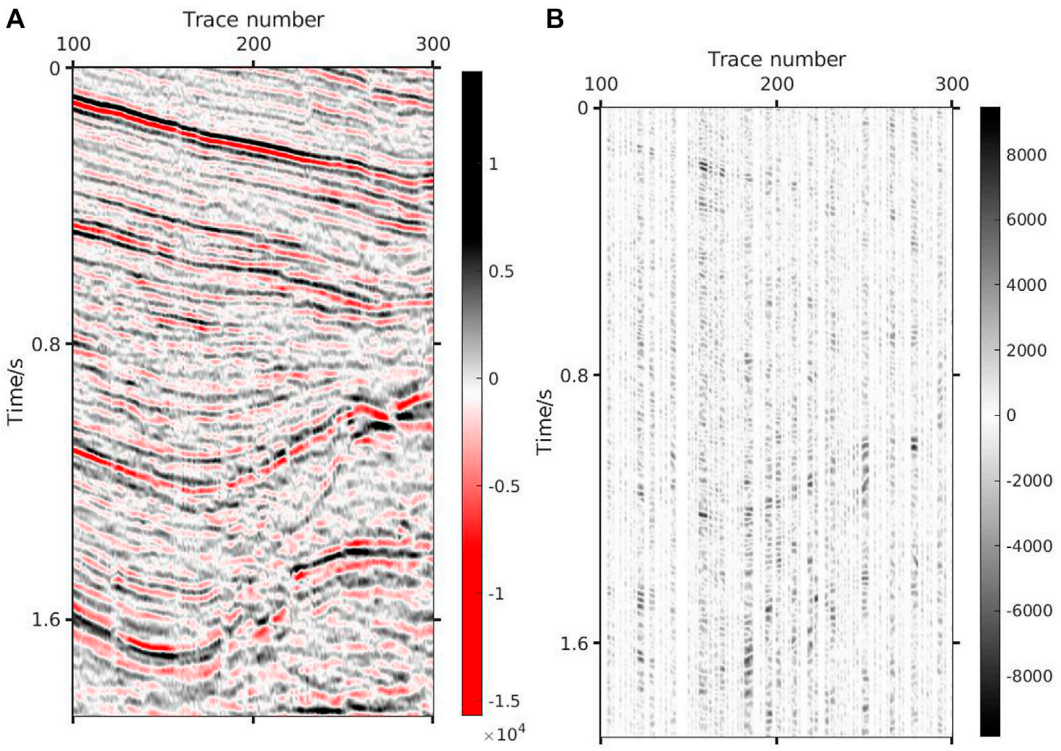


FIGURE 10
Result *via* DCT. **(A)** Result *via* DCT (SNR=3.03 dB PSNR=23.56 dB). **(B)** Absolute error *via* DCT.

a threshold model, which impacts the algorithm's speed and accuracy. Standard threshold models include linear models, exponential models, and data-driven models. In this paper, the exponential model is used, which has the form as:

$$\begin{aligned} \lambda_n &= \left(\lambda_{\min}/\lambda_{\max}\right)^{\frac{n-1}{N-1}} \cdot \lambda_{\max}, \quad n = 1, 2, \ldots, N \\ \lambda_{\max} &= q_{\max} \cdot \max_i \left\{|D^* Y|_i\right\} \\ \lambda_{\min} &= q_{\min} \cdot \max_i \left\{|D^* Y|_i\right\} \end{aligned} \tag{10}$$

where, $\lambda_{\max}$ and $\lambda_{\min}$ stand for the chosen maximum and minimum regularization parameters, respectively. $q_{\max}$ and $q_{\min}$ are user-defined percentages. N is the number of iterations.

2.3 The theory of denoising based on MCA

CS uses the structural differences between the useful signal and random noise in the sparse domain to denoising. A noisy seismic record y and its sparse representation can be expressed as:

$$y = x + \varepsilon = D\alpha + \varepsilon \tag{11}$$

where, ε represents random noise; D is a sparse dictionary; α is a sparse coefficient; x is the original signal. Equation 11 can be expressed in the form of MCA as:

$$y = \sum_{k=1}^{K} x_k + \varepsilon = \sum_{k=1}^{K} D_k \alpha_k + \varepsilon \tag{12}$$

To obtain sparse α, the following optimization problem is solved using the same method as in 2.2:

$$\min \|\alpha_k\|_1 \; s.\, t \left\| y - \sum_{k=1}^{K} D_k \alpha_k \right\| \le \sigma_\varepsilon \tag{13}$$

3 Dictionary selection

The selection of D is the core problem of the MCA Method. Different dictionaries have a significant influence on the effect of sparse representation. We mainly focus on the overall and local characteristics of seismic data. DFT (Direct Fourier Transformer) and DCT (Discrete Cosine Transform) can be used to transform the overall trend of

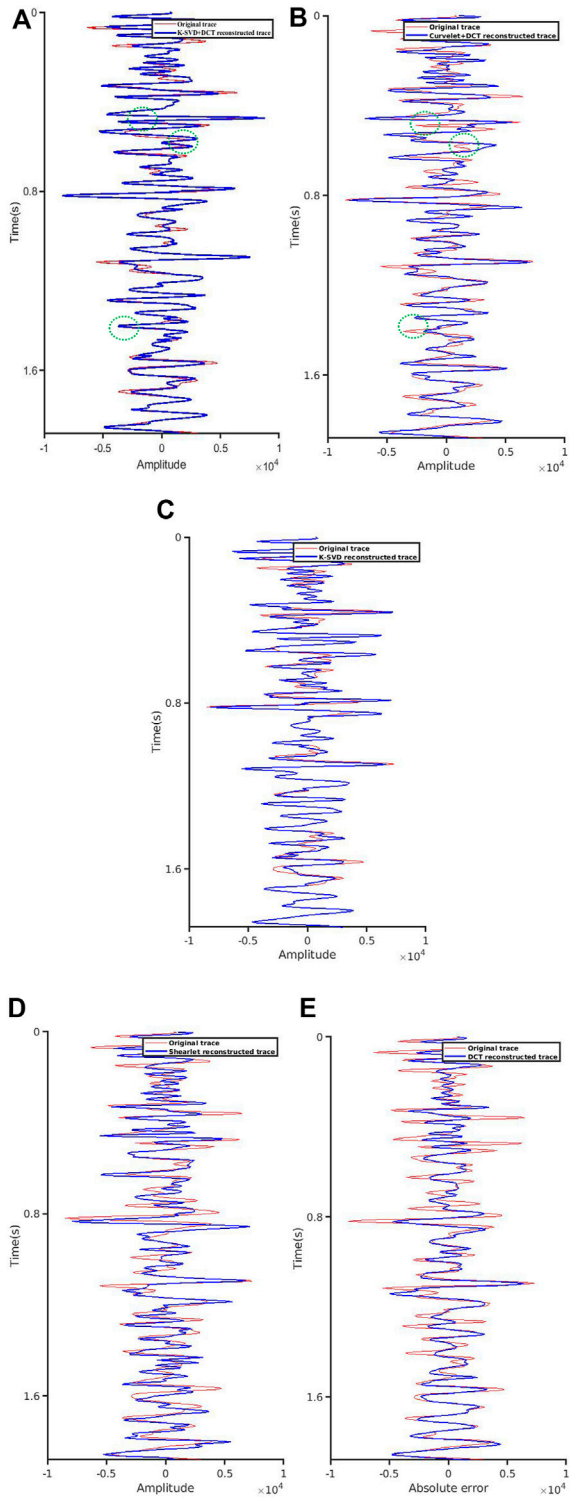


FIGURE 11
Reconstructed single trace result (Red: Original data 221th trace; Blue: Reconstructed data 221th trace). **(A)** Single trace result *via* DCT+K-SVD; **(B)** Single trace result *via* Curvelet+DCT; **(C)** Single trace result *via* K-SVD; **(D)** Single trace result *via* Shearlet; **(E)** Single trace result *via* DCT.

the data. We take x [n] and y [n] in Figure 1 as an example to illustrate the advantages of DCT. We start by considering the sequence y [n]=x [n]+x [2N-1-n] and this just consists of adding a mirrored version of x [n] to itself. When transform is DFT, we work with extension of x [n], when transform is DCT, we work with extension of y [n]. From Figure 1, it can be seen DFT case the extension introduces discontinuities but this does not happen for the DCT, due to the symmetry of y [n]. The elimination of this artificial discontinuity, which contains a lot of high frequencies, which is the reason why the DCT is much more efficient.

A single mathematical dictionary cannot adequately sparse the representation of seismic data, resulting in loss of information and bad reconstruction data, dictionary learning has been well applied to this problem. Therefor, for the local characteristics of the data, we consider the K-SVD dictionary learning algorithm. We divide the seismic data into two distinct components, so $K = 2$. Shearlet transform gets D_1 for $k = 1$, DCT transform gets D_2 for $k = 2$. The specific steps for training the dictionary are as follows:

Assuming that there is a training database $\{y_i\}_{i=1}^{M}$, which can determine the generated model dictionary A. Suppose the training error is ϵ, and the goal is to find A. Create the following training model:

$$\min_{A,\{x_i\}_{i=1}^{M}} \sum_{i=1}^{M} \|x_i\|_0 \; s.\; t \; \left\|y_i - Ax_i\right\|_2 \leq \epsilon \tag{14}$$

where, each signal y_i can be expressed as the sparsest representation A on the dictionary x_i to be obtained.

If a dictionary A is obtained whose coefficients are sparse when representing the data, a usable model R can be obtained. Predecessors have proved that when $\epsilon = 0$ and all elements in the training database can be represented by k_0 (k_0 <spark $(A_0)/2$) atoms, there is a unique sparse dictionary A_0. Therefore, sparsity can also be used as a constraint, and the goal is to obtain the best fit of the signal, as shown in Eq. 15:

$$\min_{A,\{x_i\}_{i=1}^{M}} \sum_{i=1}^{M} \left\|y_i - Ax_i\right\|_2^2 \; s.\; t \; \|x_i\|_0 \leq k_0 \tag{15}$$

The vectors in the database are combined into a matrix Y by column, and the corresponding sparse representation forms the matrix X. The problem of obtaining a dictionary is equivalent to the problem of decomposing matrix Y into AX, where the sizes of A and X are fixed and X has sparse columns. The inner layer of Eq. 15 is the sparse problem of representing vector x_i when A is known, and the outer layer is the minimization problem. In the kth iteration, each element y_i in the database is solved with the dictionary $A_{(k-1)}$ (the dictionary obtained by the k-1st iteration) to obtain the matrix $X_{(k)}$, and finally, the Least-Squares is used to solve $A_{(k-1)}$:

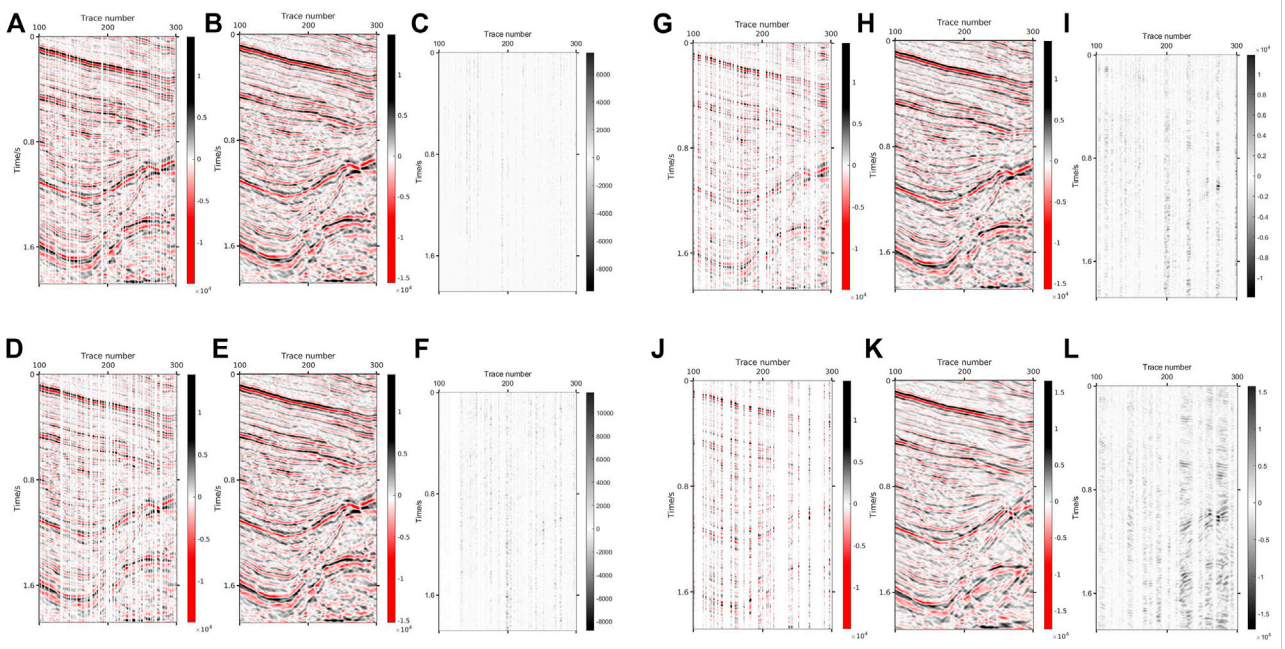


FIGURE 12
DCT+K-SVD results *via* missing in different proportions. **(A)** 20% missing image. **(B)** Reconstructed result *via* 20% missing (SNR=18.37 dB PSNR=34.72 dB). **(C)** Absolute error *via* 20% missing; **(D)** 40% missing image. **(E)** Reconstructed result *via* 40% missing (SNR=14.25 dB PSNR=30.59 dB). **(F)** Absolute error *via* 40% missing. **(G)** 60% missing image. **(H)** Reconstructed result *via* 60% missing (SNR=9.83 dB PSNR=26.17 dB). **(I)** Absolute error *via* 60% missing; **(J)** 80% missing image. **(K)** Reconstructed result *via* 80% missing (SNR=0.5 dB PSNR=21.29 dB). **(L)** Absolute error *via* 80% missing.

$$A_{(k)} = \underset{A}{\operatorname{argmin}} \left\| Y - AX_{(k)} \right\|_F^2 = YX_{(k)}^T \left(X_{(k)} X_{(k)}^T \right)^{-1} = YX_{(k)}^{-1} \tag{16}$$

$$\left\| Y - AX_{(k)} \right\|_F^2 = \left\| Y - \sum_{j-1}^{m} a_j x_j^T \right\|_F^2 = \left\| Y - \sum_{j \neq j_0}^{m} a_j x_j^T \right\|_F^2 \tag{17}$$

Where, x_j^T represents the jth row of X. The items in parentheses are used as the known error matrix, which is recorded as Eq. 18:

$$\mathrm{E}_{j_0} = \mathrm{Y} - \sum_{\mathrm{j} \neq \mathrm{j}_0} \mathrm{a_j x_j^T} \tag{18}$$

The optimal a_{j_0} and x_j^T are approximations of E_{j_0} and their rank must be 1. Solving with the SVD algorithm will generate a dense vector x_j^T, which increases the number of non-zero items in the X representation. To minimize the known error matrix while keeping all expressed potentials unchanged. We take a subset of the columns in E_{j_0} (the columns in this subset correspond to the signal using the j$_0^{th}$ atom in the sample set), and the items in the row x_j^T of these columns are non-zero. In this way, only the non-zero coefficient in x_j^T changes, and the potential remains unchanged. The detailed K-SVD dictionary learning algorithm is as follows:

Goal: Obtain the sparse dictionary A by solving the approximate solution of Eq. 15;

Initialize: $k = 0$ and:

1) Initialize the dictionary: form $A_{(0)} \in R^{n \times m}$ (random elements or m randomly selected samples).
2) Normalization: Normalize the columns of $A_{(0)}$

Main iteration: $k = k + 1$, and perform the following steps:

3) Approximate solution using tracking algorithm, get sparse representation $x_i^{'}(1 \leq i \leq M)$, they form the matrix $x_{(k)}$.

$$\mathrm{x_i^{'}} = \arg\min_{\mathrm{x}} \left\| \mathrm{y_i} - \mathrm{A_{(k-1)}x} \right\|_2^2 \text{ s.t. } \|\mathrm{x}\|_0 \leq \mathrm{k_0}$$

4) K-SVD dictionary update stage: use the following steps to update the columns of the dictionary and obtain $A_{(k)}$ (repeat $j_0 = 1, 2, \ldots, m$).
   1) Define the sample set $\Omega_{j_0} = \{i | 1 \leq i \leq M, X(K)[j_0, j] \neq 0\}$ that using the atom a_{j_0};
   2) Calculate the error $\mathrm{E_{j_0}} = \mathrm{Y} - \sum_{\mathrm{j} \neq \mathrm{j}_0} \mathrm{a_j x_j^T}$, where x_i is the j-th row of the matrix $x_{(k)}$;
   3) Limit $\mathrm{E_{j_0}}$ by selecting only the column corresponding to Ω_{j_0} to get $\mathrm{E_{j_0}^R}$;
   4) Apply SVD to decompose $\mathrm{E_{j_0}^R} = \mathrm{U \Delta V^T}$, update the dictionary atom $\mathrm{a_{j_0}} = \mathrm{u_1}$ and the representation $\mathrm{x_{j_0}^R} = \Delta[1,1] \cdot \mathrm{V_1}$.

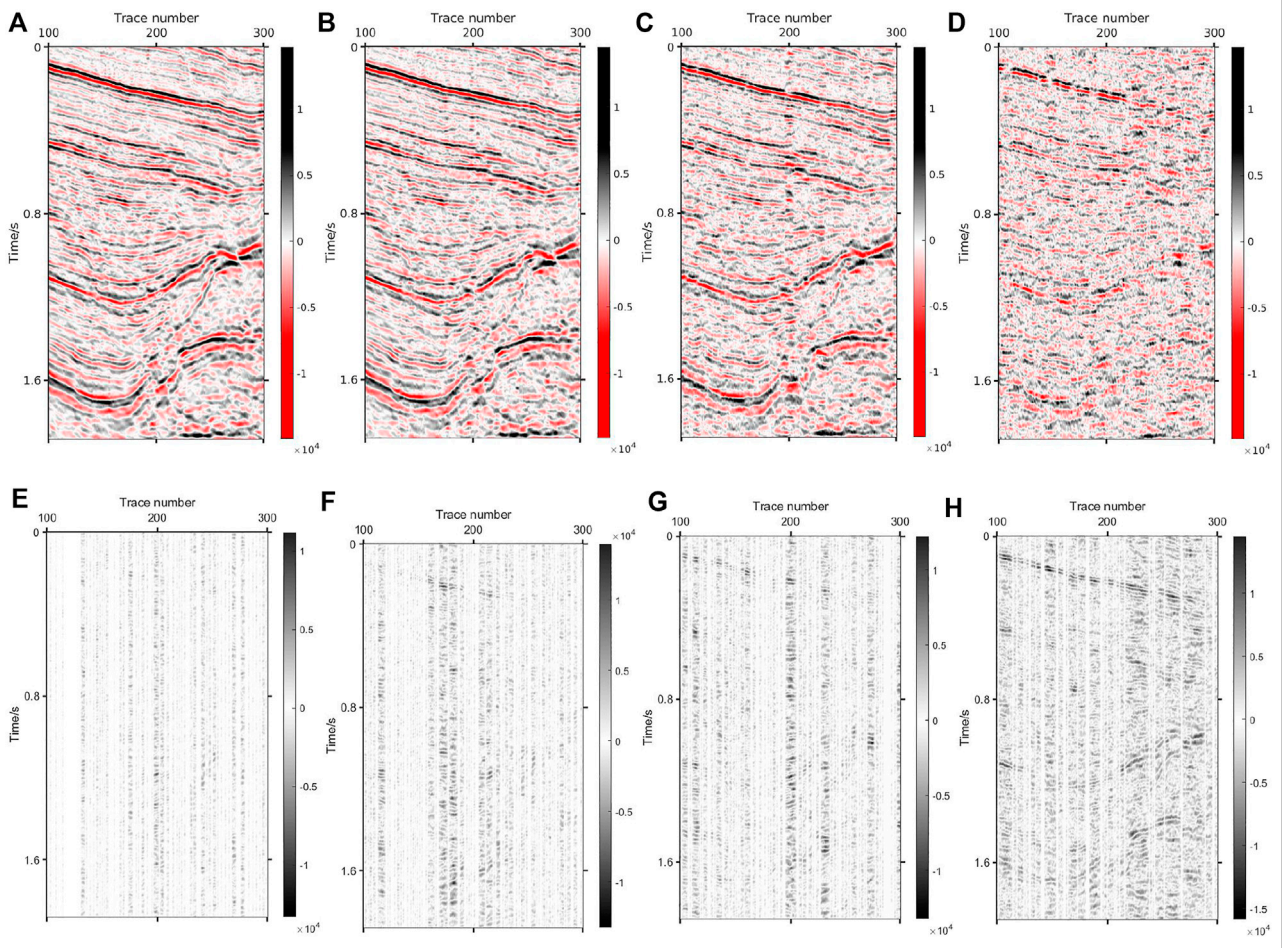


FIGURE 13
Curvelat+DCT results *via* missing in different proportions. **(A)** Reconstructed result *via* 20% missing (SNR=15.32dBdB PSNR=31.67 dB). **(B)** Reconstructed result *via* 40% missing (SNR=8.87 dB PSNR=25.21 dB). **(C)** Reconstructed result *via* 60% missing (SNR=4.52 dB PSNR=20.87 dB). **(D)** Reconstructed result *via* 80% missing (SNR=0.36 dB PSNR=16.71 dB). **(E)** Absolute error *via* 20% missing; **(F)** Absolute error *via* 40% missing. **(G)** Absolute error *via* 60% missing; **(H)** Absolute error *via* 80% missing.

5) Stop condition: If the change of $\|Y - AX_{(k)}\|_F^2$ is small enough, stop the iteration; otherwise, continue the iteration.

Output: get the result $A_{(K)}$.

Figures 2,3 show the update process of the K-SVD dictionary. We select a piece of seismic image for the test. We observe that out of a seismic data, less than half were 'active' (i.e., non-flat). We randomly choose 100 "active" patches for the dictionary training. The image and examples of 'active' patches extracted from it are shown in Figure 2. Figure 3 shows the change of the dictionary after every 5 iterations (from 5th to 40th). As the iterations proceed, the dictionary contains more and more basic features. At the 5^{th} iteration, the dictionary $A_{(5)}$ contains only some point features, which are insufficient to express the events. At the 40th iteration, the dictionary $A_{(40)}$ contains many linear features, and the linear basis functions of various forms are sufficient to achieve the best sparse representation of the seismic data. E is the error of each dictionary and the value is found by Eq. 18. Figure 4 shows the workflow of the method.

4 Test

4.1 Evaluation parameters

To quantitatively describe the reconstruction results, this paper introduces two evaluation parameters, which are defined as follows:

1) Signal Noise Ratio $R_{S/N}$

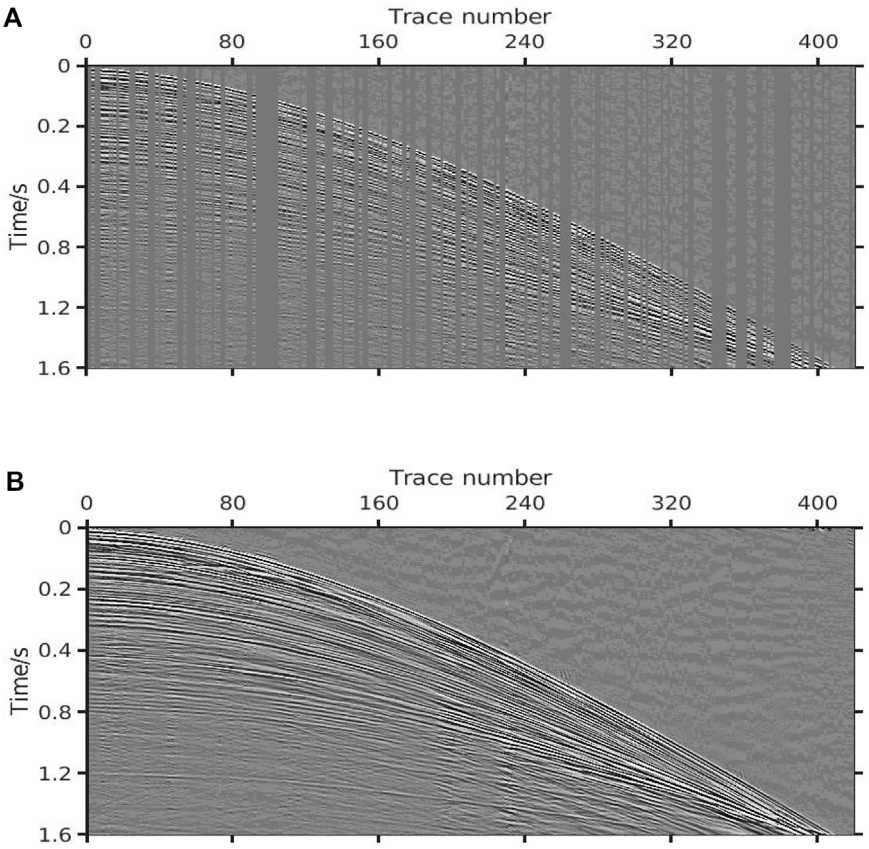


FIGURE 14
Regularized reconstruction of real data. **(A)** Original data (SNR=2.16 dB PSNR=11.2 dB). **(B)** Reconstructed data (SNR=23.89 dB PSNR=32.92 dB).

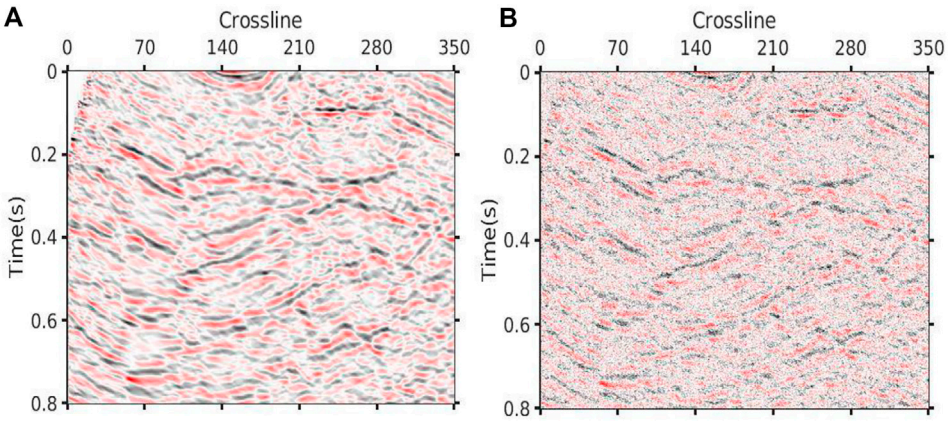


FIGURE 15
Original image and noisy image. **(A)** Original image. **(B)** Noisy image (SNR=−2.55 dB PSNR=13.79 dB).

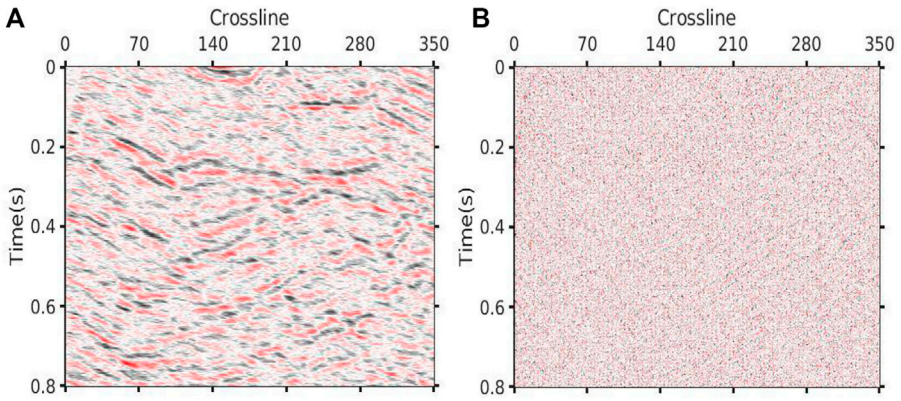


FIGURE 16
Denoising result *via* DCT+K-SVD. **(A)** Denoised data (SNR=3.67 dB PSNR=20.02 dB). **(B)** Removed noise.

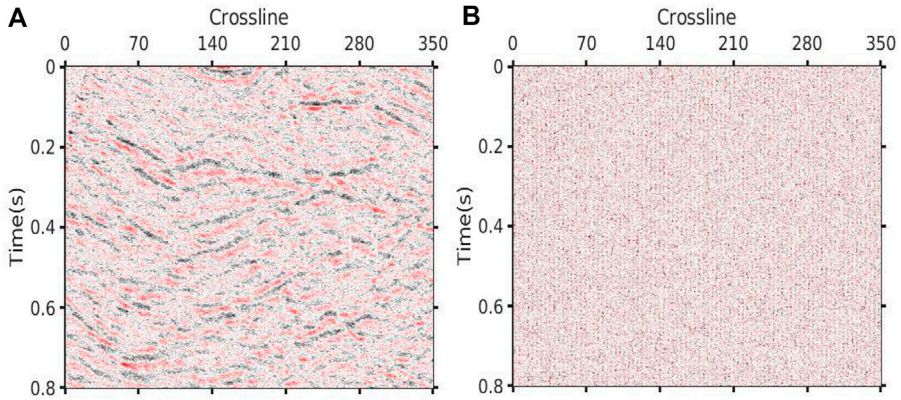


FIGURE 17
Denoising result *via* DCT. **(A)** Denoised data (SNR=−1.19 dB PSNR=15.15 dB). **(B)** Removed noise.

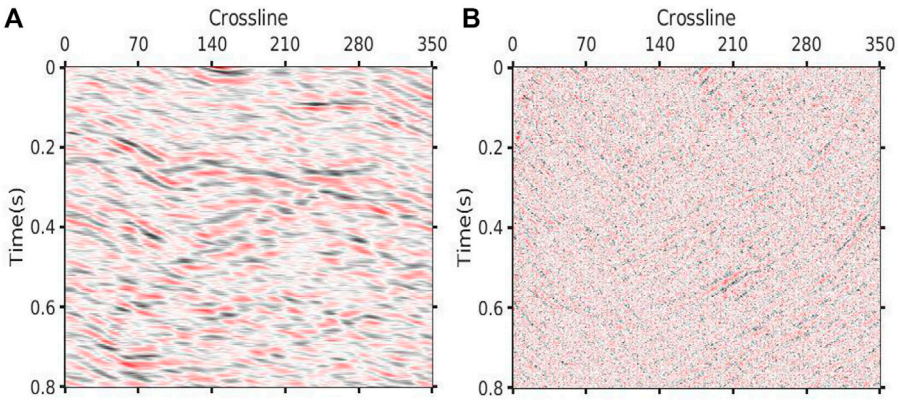


FIGURE 18
Denoising result *via* Shearlet. **(A)** Denoised data (SNR=3.93 dB PSNR=20.28 dB). **(B)** Removed noise.

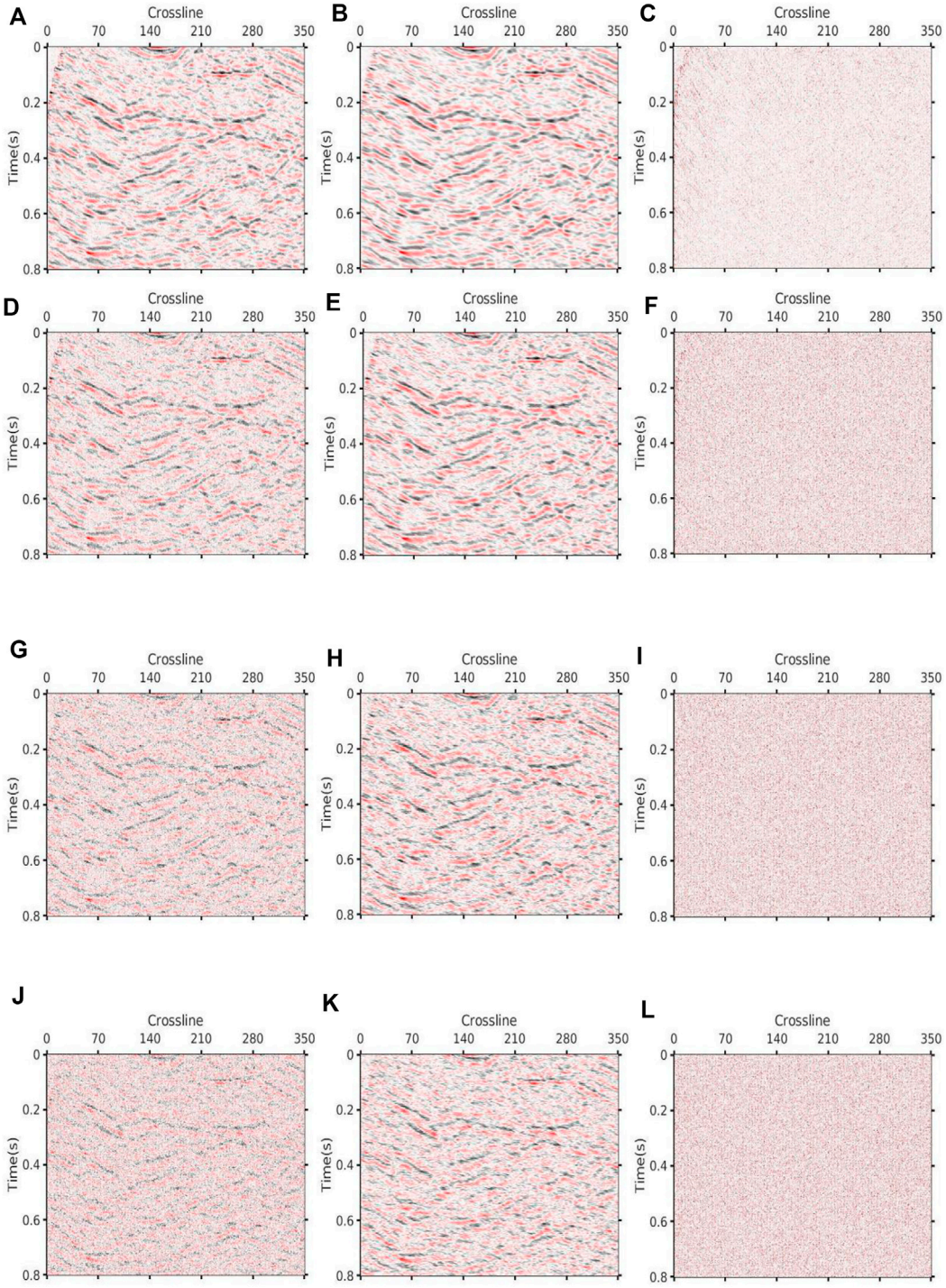


FIGURE 19
DCT+K-SVD denoising results *via* different levels of noise. **(A)** Image with 20 Hz noise (SNR=6.99 dB PSNR=23.34 dB). **(B)** Denoised data *via* 20 Hz noise (SNR=9.65 dB PSNR=25.99 dB). **(C)** Removed noise *via* 20 Hz noise. **(D)** Image with 35 Hz noise (SNR=0.96 dB PSNR=17.31 dB). **(E)** Denoised data *via* 35 Hz noise (SNR=5.9 dB PSNR=22.24 dB). **(F)** Removed noise *via* 35 Hz noise. **(G)** Image with 50 Hz noise (SNR=−3.88 dB PSNR=12.46 dB). **(H)** Denoised data *via* 50 Hz noise (SNR=1.94 dB PSNR=18.29 dB). **(I)**Removed noise *via* 50 Hz noise. **(J)** Image with 65 Hz noise (SNR=−6.98 dB PSNR=9.36 dB). **(K)** Denoised data *via* 65 Hz noise (SNR=−0.25 dB PSNR=16.09 dB). **(L)** Removed noise *via* 65 Hz noise.

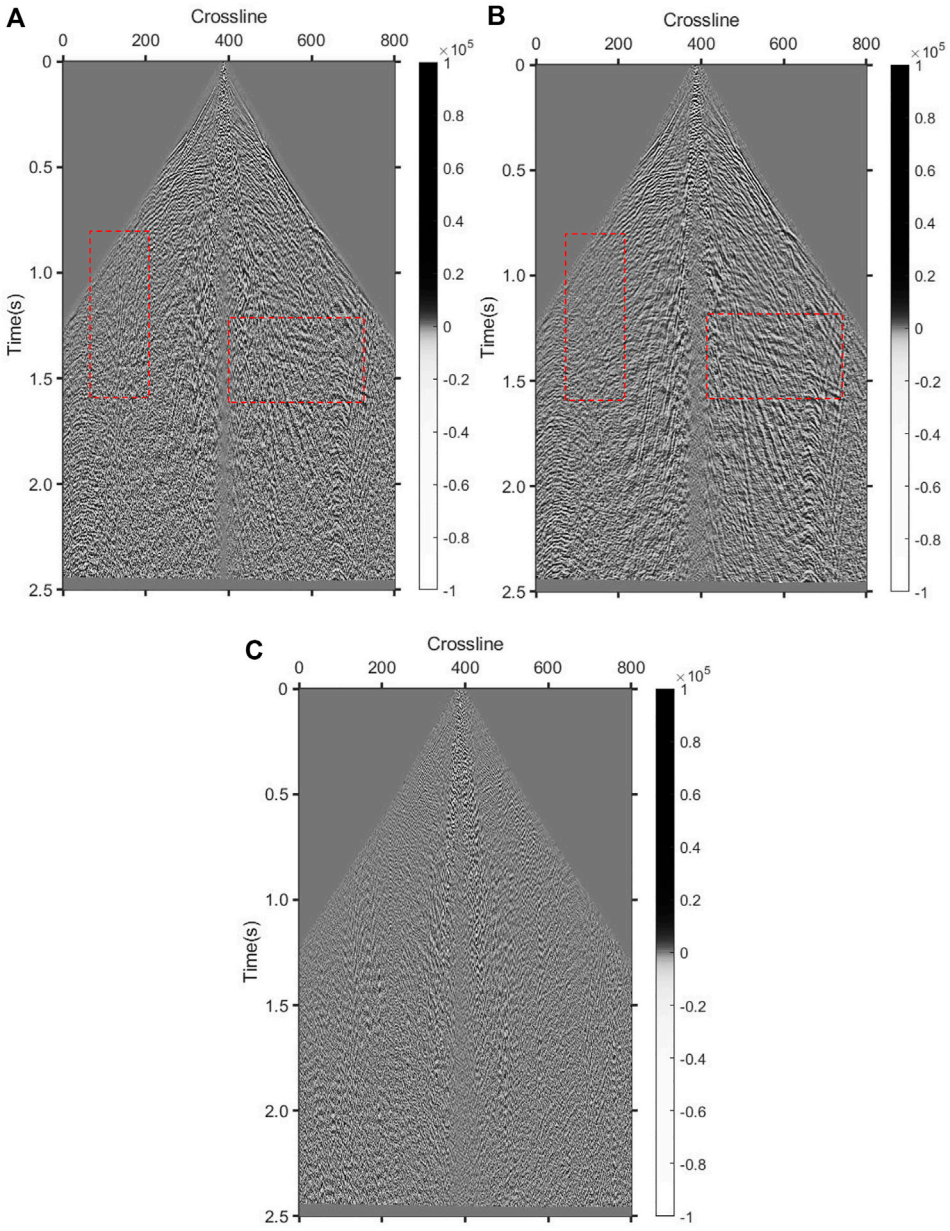


FIGURE 20
Real data denoising. **(A)** Original data (SNR=1.69 dB PSNR=26.79 dB). **(B)** Denoised data (SNR=5.93 dB PSNR=31.02 dB). **(C)** Removed noise.

$$R_{S/N} = \lg\left(\frac{\|X\|_2^2}{\|X' - X\|_2^2}\right)^{10} \tag{19}$$

Where, X represents complete seismic data; X' represents a seismic data reconstruction.

2) Peak Signal Noise Ratio $R_{S/N}^{P}$

$$R_{S/N}^{P} = \log_{10}\left(\frac{\max(X)^2}{M_{SE}}\right)^{10} \tag{20}$$

Among them, M_{SE} represents mean square error, and its calculation formula is as follows.

$$M_{SE} = \frac{1}{mn}\sum_{i=0}^{m}\sum_{j=0}^{n}\|x(i,j) - x'(i,j)\|_2^2 \tag{21}$$

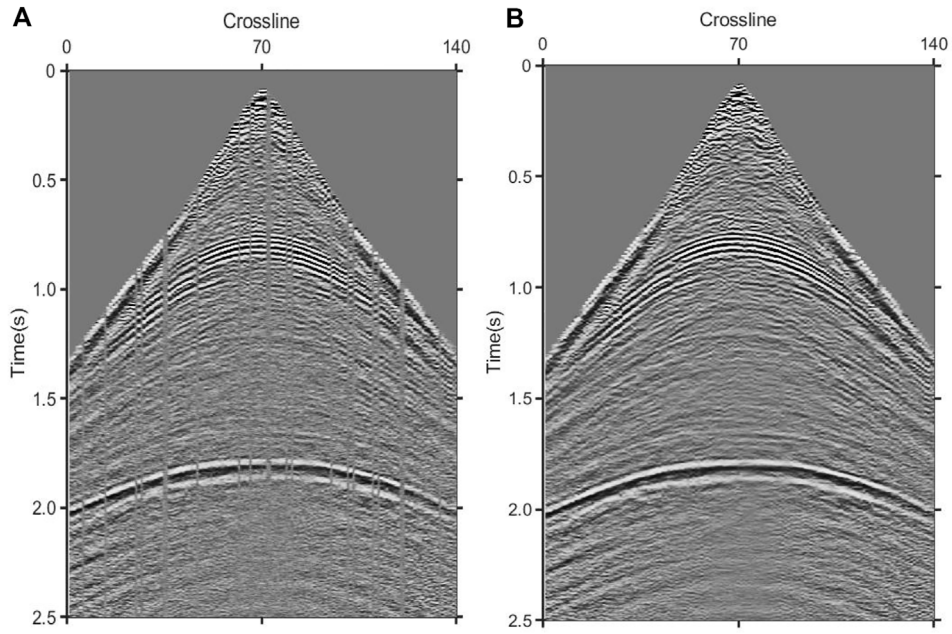


FIGURE 21
Real data processing. **(A)** Original data (SNR=13.61B PSNR=37.91 dB). **(B)** Reconstruct the denoised data (SNR=22.19 dB PSNR=46.49 dB).

TABLE 1 Evaluation values for different methods.

Before reconstruction	Methods	Evaluation values	After reconstruction	Values of change	Time (s)
SNR=3.03 dB PSNR=19.38 dB	K-SVD	SNR	15.88	+12.85	400
		PSNR	3.23		
	DCT+K-SVD	SNR	15.93 dB	+12.9 dB	280
		PSNR	32.32 dB		
	Curvelet+DCT	SNR	11.45 dB	+8.42 dB	156
		PSNR	27.79 dB		
	Shearlet	SNR	7.59 dB	+4.56 dB	128
		PSNR	23.94 dB		
	DCT	SNR	7.22 dB	+4.19 dB	2
		PSNR	23.56 dB		

4.2 Reconstruction of real data

This paper first verifies the effectiveness of the proposed method for the reconstruction of randomly sampled seismic data by comparing four sets of experimental data. Figure 5A shows a partial image of an offset profile, and Figure 5B depicts the image obtained by randomly sampling 50% of the traces.

The proposed method can effectively reconstruct the underground medium image in the case of a low sampling rate. It can be seen from Figure 6 that the proposed method can not only reconstruct the shallow reflection images, but the deep weak reflection images are also well reconstructed. Figure 6A presents the reconstruction result of DCT+K-SVD; Figure 6B describes the reconstructed error, which represents reconstructed quality without noise. Figures 6C,D show the reconstruction components of K-SVD and DCT. It can be seen that the reconstruction amount of K-SVD for small structures is rich, and most structures that cannot be accurately reconstructed by fixed dictionaries are well restored by K-SVD. The simple structures, such as smooth events and

TABLE 2 Evaluation values for different missing.

Before reconstruction	Proportion of missing (%)	Evaluation values	Method	After reconstruction (dB)	Values of change	Time (s)
SNR=3.03 dB PSNR=19.38 dB	20	SNR	DCT+K-SVD	18.37	+15.34 dB	250
		PSNR		34.72		
	40	SNR		14.25	+11.22 dB	
		PSNR		30.59		
	60	SNR		9.83	+6.8 dB	
		PSNR		26.17		
	80	SNR		5.56	+2.53 dB	
		PSNR		21.29		
	20	SNR	Curvelet+DCT	15.32	+12.29 dB	160
		PSNR		31.67		
	40	SNR		8.87	+5.84 dB	
		PSNR		25.21		
	60	SNR		4.52	+1.49 dB	
		PSNR		20.87		
	80	SNR		0.36	-2.67 dB	
		PSNR		16.71		

TABLE 3 Evaluation values for different methods.

Before denoising	Methods	Evaluation values	After denoising	Values of change	Time (s)
SNR=-2.55 dB PSNR=13.79 dB	DCT+K-SVD	SNR	3.67 dB	+6.22 dB	176
		PSNR	20.02 dB		
	DCT	SNR	−1.19 dB	+1.36 dB	16
		PSNR	15.15 dB		
	Shearlet	SNR	3.93 dB	+6.48 dB	98
		PSNR	23.56 dB		

strong low-frequency signal, can be easily reconstructed by DCT, which greatly reduces the calculation amount of dictionary learning and improves the efficiency. Figure 7 shows the reconstruction results of K-SVD. Due to the optimal expression of dictionary learning algorithm, K-SVD can reconstruct seismic data with very high accuracy. However, it will take a lot of time, which is particularly obvious when processing huge real data. Figure 8 shows the reconstruction result of Curvelet+DCT. Figure 8A reveals that although this method can recover all the missing information, the reconstructed data have insufficient energy, the details are clearly depicted, and the microstructure cannot be distinguished, while Figure 8B suggests the existence of errors at various locations in the reconstructed data. It can be seen from Figures 8C,D that the two dictionaries (Curvelet and DCT) have limited recovery ability for small structures. Curvelet transform is applicable to signals with curve characteristics, but DCT cannot perfectly reconstruct the remaining signals. In contrast, the Shearlet method is provided with a stronger reconstruction ability, which combines multi-scale geometric analysis through synthetic wavelet theory and affine system, and generates the basis function by stretching, translating, and rotating a base function. However, its lack of global sparse representation ability gives rise to reconstruction noise caused by undersampling in the reconstruction data, as shown in Figure 9A. Figure 10 indicates that the DCT method can

TABLE 4 Evaluation values for different noise.

Proportion of noise (Hz)	Evaluation values	Before denoising	After denoising	Values of change
20	SNR	6.99 dB	9.65 dB	+2.66 dB
	PSNR	23.34 dB	25.99 dB	
35	SNR	0.96 dB	5.9 dB	+4.94 dB
	PSNR	17.31 dB	22.24 dB	
50	SNR	−3.88 dB	1.94 dB	+5.82 dB
	PSNR	12.46 dB	18.29 dB	
65	SNR	−6.98 dB	-0.25 dB	+6.73 dB
	PSNR	9.36 dB	16.09 dB	

completely supplement the missing data, but its detail-describing ability is poor, and the image cannot be reconstructed with a high resolution. To further compare the reconstruction effects of the four methods, the 221th trace is taken for a separate comparison, and the results are shown in Figure 11. It can be seen from the curves that the results obtained by the proposed method have the best fit with the original signal. The Shearlet+DCT and Shearlet method is seriously inconsistent in some areas (For example, 0.6 s–0.7 s and 1.4 s–1.5 s). The curve of the signal obtained by the DCT and shearlet method poorly fits the original signal. Table 1 shows the quantitative evaluation parameters of the five methods, of which the proposed method has the highest improvement in SNR and takes less time than K-SVD.

The following conclusions can thus be drawn from numerical experiments: 1) DCT+K-SVD, Shearlet and DCT reconstruction methods based on CS can complement the missing data at lower sampling rates, while it is difficult for the Curvelet+DCT method to complete the reconstruction. This is because the Curvelet will recognize it as a boundary in the large vacancy position, and the Curvelet is equipped with good boundary protection properties, making it impossible to reconstruct the large missing location; 2) Sparse transformation is the key to the reconstruction algorithm, with sparser coefficients obtained after transformation indicating a better reconstruction effect. This group of experiments shows that Shearlet presents better reconstruction results than DCT because of the multi-directional and multi-scale characteristics of Shearlet transform, representing the signal more sparsely; 3. The hereby proposed reconstruction method based on the DCT+K-SVD sparse transformation can well reconstruct the underground medium. Overall, the difference from original images is the smallest, while that from SNR and PSNR are the biggest. The weak reflection image at the bottom is also endowed with a better reconstruction effect in terms of local details.

In order to explore the reconstruction ability of the proposed method, data are reconstructed with different degrees of random missing, including 20%, 40%, 60%, and 80%. The missing data with four missing levels, the reconstruction results and the errors are presented in Figure 12, where it can be noticed that the reconstruction effect gradually decreases with the increase of the number of deletions. However, it can be guaranteed that the proposed method can maintain satisfactory results in the case of less than 60% missing. When 80% of the data are missing, although some small features cannot be perfectly reconstructed, the information of the reconstructed data is complete, the event axis is continuous, and the SNR is improved. However, it can be seen from Figures 13D,H that other dictionary combination methods (Curvelet+DCT) are difficult to completely reconstruct a large number of missing data, and their reconstructed data have discontinuous events and low SNR ratio. Table 2 shows the parameters of K-SVD+DCT before and after reconstruction of data with different missing degrees.

The real seismic data of a certain work area are used to test the applicability of the method. The humid climate caused water accumulation in the ground, and multiple receivers on the survey line are damaged, resulting in incomplete acquisition data. Figure 14 is a partial display of the original data. Multiple breaks can be observed in the events, which will be further reconstructed to get complete data and smooth events. Figure 14B shows the reconstruction result, and all the missing data are found to have been recovered. In addition, the events of the reconstructed data are smooth and continuous, and the SNR of the data is also improved.

4.3 Denoising of real data

The proposed method is hereby applied to Compressed Sensing denoising that uses the differences between useful signal and random noise in sparse domain for denoising. In order to verify the effectiveness of the method, a part of the offset profile is selected for the experiment. The data have a total of 350 traces and 800 sampling points, as shown in Figure 15.

Figure 15B describes the data with noise, the main frequency of which is 40 Hz. DCT+K-SVD is hereby compared with DCT and Shearlet+DCT to prove its superiority. Figures 16–18 are the denoising result and the removed noise of the three methods. Figure 16 depicts the denoising result of the proposed method, and suggests that the method can efficiently extract useful signals of the deep and shallow layers. Additionally, this method does not cause any information loss and better retains the original data characteristics. Figure 17 shows that the denoising ability of the DCT method is weak and fails to effectively suppress random noise. In contrast, the Shearlet method has a stronger denoising ability and can basically suppress all the noise. However, due to the lack of self-adaptation, this method is still subject to the problem that the useful signals are suppressed, and the removed useful signals can be seen from Figure 18. Table 3 shows the noise reduction evaluation parameters of the three methods. The denoising ability of the proposed method is tested under different degrees of random noise, including 20Hz, 35Hz, 50Hz, and 65 Hz. Figure 19 shows the noisy data with four levels, the denoising results and the removed noise, and it can be observed from Figure 19J that even under the interference of strong noise, the effective signal is covered by a large area, making it still possible to extract useful signals and obtain satisfactory denoising results. Table 4 shows the relevant parameters.

To verify the applicability of this denoising method, the real data, a single shot record of a work area in western China, with a total of 800 traces and a sampling time of 2.5s, are further processed. Figure 19A reveals that the data contain a lot of noise. The existence of random noise and linear line noise makes the SNR of the shot set data low, and the continuity of the event is poor. Figure 20B shows the denoising result of DCT+K-SVD, indicating that most of the random noise is effectively suppressed, and that the event between 1.25 m and 1.5 m is clearer. The SNR of the data has been significantly improved, and forged a good foundation for subsequent processing. Figure 19C depicts the removed noise, and reveals that not only random noise is removed, but some linear noise and surface waves are also suppressed to a certain extent. The reason is that the dictionary learning algorithm self-adaptively learns the characteristics of the useful signal that can effectively distinguish the useful signal from other signals.

4.4 Process of real missing noisy data

This section presents the comprehensive application of the proposed method, based on which, the reconstruction and denoising of missing noisy data are implemented. Original data and processed data are shown in Figure 21. Figure 21A describes the original data, where partial missing and random noise can be observed. Figure 21B reveals that all the missing information is accurately recovered, and that the SNR is also significantly improved.

5 Conclusion

Problems such as missing tracks and bad tracks generally give rise to the incompletion of the seismic data. From the inversion perspective, incomplete image reconstruction is an ill-posed inverse problem, and seismic signals are inevitably affected by noise during the propagation process, which reduces the quality of seismic data and brings difficulties to subsequent interpretation work. However, compressed sensing reconstructs the data using the sparsity of the signal, and is well applied in the fields of regular reconstruction and denoising.

A new dictionary combination, i.e., K-SVD+DCT, is hereby proposed under the MCA framework, which overcomes the limitation of fixed base functions by training dictionaries fully suitable for processed data. DCT is a global type transformation used to reconstruct a smooth event. Therefore, the coefficients of the signal obtained using K-SVD+DCT are sparser, and have a good reconstruction and denoising effect on both pre-stack and post-stack data. Considerable experiments show that the hereby proposed method can reconstruct the image well, that the relative error of the reconstruction result is limited, and that the local details and the deep weak reflection signal can also be well reconstructed. Besides, even under the interference of strong noise, the effective signal is covered by a large area, and it is still possible to extract useful signals and obtain satisfactory denoising results. Indeed, this method retains both the fast operation of mathematical transformations and the high precision of dictionary learning. However, only the training time of dictionary learning is reduced by reducing the training data. To this end, the focus of future work will be placed on improving the dictionary learning time and developing efficient dictionary learning algorithms.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

The first author DW is the first owner of the research results, responsible for algorithm research and article writing, XX and HZ are in charge of auxiliary research.

Funding

This study was supported by the Research on Full-Frequency Processing Method of Thin Reservoir and Research on Target Fine Characterization Technology Research (2022KT1503) and Research and realization of seismic pre-stack imaging in the

western segment of Kedong structural belt in southwestern Tarim Basin in 2022 (2022KT0506).

Conflict of interest

The authors DW, XRX, HZ, JS and XX were employed by PetroChina. The author YZ was employed by Wuhua Energy Technology Co., Ltd.

Publisher's note

References

Aharon, M., Elad, M., and Bruckstein, A. K. S. V. D. (2006). $rm K$-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Process.* 54 (11), 4311–4322. doi:10.1109/TSP.2006.881199

Chen, S. S., Donoho, D. L., and Saunders, M. A. (2001). Atomic decomposition by basis pursuit. *SIAM Rev. Soc. Ind. Appl. Math.* 43 (1), 129–159. doi:10.1137/s003614450037906x

Cuifuqi, X. D., and Zhang, G. (2003). Seismic traces interpolation using wavelet transform[J]. *Oil Geophys. Prospect.* 38 (S1), 93–97.

Dragomiretskiy, K., and Zosso, D. (2014). Variational mode decomposition. *IEEE Trans. Signal Process.* 62 (3), 531–544. doi:10.1109/tsp.2013.2288675

Gao, J. J., Sacchi, M. D., and Chen, X. H. (2013). A fast re-duced-rank interpolation method for prestack seismic volumes that dependon four spatial dimensions. *Geophysics* 78 (1), V21–V30. doi:10.1190/geo2012-0038.1

Hanliang, L. (2018). *The study on seismic data reconstruction method based on curvelet transform.* : China University of Petroleum (EastChina), Qing-dao, Shandong

Huo, Z., Deng, X., and Zhang, J. (2013). The over view of seismic data reconstruction methods. *Prog. Geophys.* 28 (4), 1749–1756.

Jia, Y., Yu, S., Liu, L., and Ma, J. (2016). A fast rank-reduction algorithm for three-dimensional seismic data interpolation. *J. Appl. Geophys.* 132, 137–145. doi:10.1016/j.jappgeo.2016.06.010

Jiang, P., Zhang, K., Zhang, Y. K., and Tian, X. (2019). Noisy seismic data reconstruction method based on morphological component analysis framework. *Prog. Geophys.* 34 (2), 573–580.

Jin, Y., Angelini, E., and Laine, A. (2005). Wavelets in medical image processing: De-noising, segmentation, and registration[J]. *Behav. Process.* 45 (1-3), 33–57. doi:10.1007/0-306-48551-6_6

Li, H., Wu, G., and Yin, X. (2012). Morphological component analysis in seismic data reconstruction[J]. *Oil Geophys. Prospect.* 47 (2), 236–243.

Lian, Q. S., Shi, B. S., and Chen, S. Z. (2015). Research advances on dictionary learning models, algorithms and applications. *Acta Autom. Sin.* 41 (2), 240–260. (in Chinese).

Liu, N., Li, F., Wang, D., Gao, J., and Xu, Z. (2021). Ground-roll separation and attenuation using curvelet-based multichannel variational mode decomposition. *IEEE Trans. Geosci. Remote Sens.* 60, 1–14. doi:10.1109/tgrs.2021.3054749

Liu, C., Li, P., Liu, Y., Wang, D., and Liu, D.-M. (2013). Iterative data interpolation beyond aliasing using seislet transform. *Chin. J. Geophys.* 56 (5), 1619–1627.

Liu, N., Li, Z., Sun, F., Wang, Q., and Gao, J. (2019). The improved empirical wavelet transform and applications to seismic reflection data. *IEEE Geosci. Remote Sens. Lett.* 16 (12), 1939–1943. doi:10.1109/lgrs.2019.2911092

Liu, N., Wu, L., Wang, J., Wu, H., Gao, J., and Wang, D. (2022). Seismic data reconstruction via wavelet-based residual deep learning. *IEEE Trans. Geosci. Remote Sens.* 60, 1–13. doi:10.1109/tgrs.2022.3152984

Liu, N., Yang, Y., Li, Z., Gao, J., Jiang, X., and Pan, S. (2020). Seismic signal de-noising using time–frequency peak filtering based on empirical wavelet transform. *Acta Geophys.* 68, 425–434. doi:10.1007/s11600-020-00413-4

Luo, T. (2015). *Study on the reconstruction of seismic data based on compressive sensing theory[D].* Changchun, Jilin: Jilin University.

Ma, J. W. (2013). Three-dimensional irregular seismic data Reconstruction via low-rank matrix completion. *Geophysics* 78 (5), V181–V192. doi:10.1190/geo2012-0465.1

Naghizadeh, M., and Sacchim, D. (2010). Beyond alias hie-rarchical scale curvelet inter polation of regularly and Irregularly sampled seismic data. *Geophysics* 75 (6), WB189–WB202. doi:10.1190/1.3509468

Olshausen, B. A., and Millman, K. J. (2000). Learning sparse overcomplete image representations//wavelet applications in signal and image processing VIII. *Int. Soc. Opt. Photonics* 4119, 445–452.

Sang, W. J., Yuan, S. Y., Yong, X. S., Jiao, X. Q., and Wang, S. X. (2021). DCNNs-based denoising with a novel data generation for multidimensional geological structures learning. *IEEE Geosci. Remote Sens. Lett.* 18 (10), 1861–1865. doi:10.1109/LGRS.2020.3007819

Shan, H., Ma, J., and Yang, H. (2009). Comparisons of wavelets, contourlets and curvelets in seismic denoising. *J. Appl. Geophys.* 69 (2), 103–115. doi:10.1016/j.jappgeo.2009.08.002

Spitz, S. (2012). Seismic trace interpolation in the *F-X* domain. *Geophysics* 56 (6), 785–794. doi:10.1190/1.1443096

Sun, M., Li, Z., Liu, Y., Wang, J., and Su, Y. (2021). Low-frequency expansion approach for seismic data based on compressed sensing in low SNR. *Appl. Sci.* 11 (11), 5028. doi:10.3390/app11115028

Tang, G., Ma, J. W., and Yang, H. Z. (2012). Seismic data denoising based on learning-type overcomplete dictionaries. *Appl. Geophys.* 9 (1), 27–32. doi:10.1007/s11770-012-0310-z

Tang, H., Mao, W., and Zhang, Y. (2020). Recon-struction of 3D irregular seismic data with amplitude preserved by high-order parabolic Radon transform. *Chin. J. Geophys.* 63 (9), 3452–3464.

Vaidyanathan, P., and Nguyen, T. (1987). Eigenfilters: A new approach to least-squares fir filter design and applications including nyquist filters[J]. *Circuits Syst. IEEE Trans.* 34 (1), 11–23. doi:10.1109/TCS.1987.1086033

Wang, B., Lu, W., Chen, X., and Wang, Z., (2018). Efficient seismic data interpolation using three-dimen-sional Curvelet transform in the frequency domain[J]. *Geophys. Prospect. Petroleum* 57 (1), 65–71.

Wang, D., Zhang, K., Li, Z., Xu, X., Liu, Q., Zhang, Y., et al. (2021a). Application of greedy deep dictionary learning. Proceedings of the SEG 2021 Workshop: 4th International Workshop on Mathematical Geophysics: Traditional & Learning. virtual : 13–16. doi:10.1190/iwmg2021-04.1

Wang, D., Zhang, K., Li, Z., Xu, X., and Zhang, Y. Seismic data reconstruction using Shearlet and DCT dictionary combination[A]. First International Meeting for Applied Geoscience & Energy Expanded Abstracts SEG Technical Program Expanded Abstracts[C].Caofeidian area, Bohai Bay Basin, 2021b: 2615–2619. doi:10.1190/segam2021-3581920.1

Wen, R., Liu, G., and Yang, R. (2018). Three key factors in seismic data reconstruction based on com- pressive sensing[J]. *Oil Geophys. Prospect.* 53 (4), 682–693.

Xu, D. X. (2016). *Research on seismic denoising based on the sparse representation and dictionary learning.* Changchun: Jilin University. (in Chinese).

Xue, Y., Tang, H., and Chen, X., (2014). Seismic data Reconstruction based on high order high resolution Radon transform. *Oil Geophys. Prospect.* 49 (1), 95–100+131.

Yu, S., and Ma, J. (2018). Complex variational mode decomposition for slop-preserving denoising. *IEEE Trans. Geosci. Remote Sens.* 56 (1), 586–597. doi:10.1109/tgrs.2017.2751642

Yu, S., Ma, J., and Osher, S. (2016). Monte Carlo data-driven tight frame for seismic data recovery. *Geophysics* 81 (4), V327–V340. doi:10.1190/geo2015-0343.1

Yu, S., Ma, J., Zhang, X., and Sacchi, M. (2015). Interpolation and denoising of high-dimensional seismic data by learning a tight frame. *Geophysics* 80 (5), V119–V132. doi:10.1190/geo2014-0396.1

Yuan, S. Y., Zhao, Y., Xie, T., Qi, J., and Wang, S. X. (2022). SegNet-based first-break picking via seismic waveform classification directly from shot gathers with sparsely distributed traces. *Pet. Sci.* 19 (1), 162–179. doi:10.1016/j.petsci.2021.10.010

Zhang, H., and Chen, X. (2013). Seismic data reconstruction based on jittered sampling and curvelet transform. *Chin. J. Geophys.* 56 (5), 1637–1649.

Zhang, H., and Chen, X. (2017). *Theory and method of seismic data reconstruction*. Beijing: SciencePress.

Zhang, J., and Tong, Z., (2003). 3-D seismic trace interpolation in f-k domain[J]. *Oil Geophys. Prospect.* 38 (1), 27–30.

Zhang, K., Zuo, W., Chen, Y., Meng, D., and Zhang, L. (2017). Beyond a Gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Trans. Image Process.* 26 (7), 3142–3155. doi:10.1109/tip.2017.2662206

Zhang, K., Zhang, Y., Li, Z., Tian, X., Ouyang, Y., and Chen, J. (2019). Seismic data reconstruction method combined with discrete cosine transform and shearlet dictionary under morphological component analysis framework. *Cnkisunsydq* 54 (5), 1005

Zhang, Y., Zhang, H., Yang, Y., Liu, N., and Gao, J. (2021). Seismic random noise separation and attenuation based on MVMD and MSSA. *IEEE Trans. Geosci. Remote Sens.* 60, 1–16. doi:10.1109/tgrs.2021.3131655

Zhou, Y., Liu, Z., and Zhang, Z. (2015). Seismic signal reconstruction under the morphological component analysis framework combined with discrete cosine transform (DCT) and curvelet dictionary[J]. *Geophys. Prospect. Petroleum* 54 (5), 560–568. doi:10.3969/j.issn.1000-1441.2015.05.009

frontiers | Frontiers in Earth Science

TYPE Original Research
PUBLISHED 16 February 2023
DOI 10.3389/feart.2023.1090620

OPEN ACCESS

EDITED BY
Peng Zhenming,
University of Electronic Science and Technology of China, China

REVIEWED BY
Sanyi Yuan,
China University of Petroleum, Beijing, China
Shu Li,
Jishou University, China
Yingpin Chen,
Minnan Normal University, China

*CORRESPONDENCE
Hao Wu,
wuhao@cdut.edu.cn

SPECIALTY SECTION
This article was submitted to Environmental Informatics and Remote Sensing, a section of the journal Frontiers in Earth Science

RECEIVED 05 November 2022
ACCEPTED 06 February 2023
PUBLISHED 16 February 2023

CITATION
He L, Wu H and Wen X (2023), Prestack seismic random noise attenuation using the wavelet-inspired invertible network with atrous convolutions spatial pyramid. *Front. Earth Sci.* 11:1090620. doi: 10.3389/feart.2023.1090620

Prestack seismic random noise attenuation using the wavelet-inspired invertible network with atrous convolutions spatial pyramid

Liangsheng He[1,2], Hao Wu[1,2]* and Xiaotao Wen[1,2]

[1]Key Laboratory of Earth Exploration and Information Techniques, Chengdu University of Technology, Ministry of Education, Chengdu, China, [2]Chengdu University of Technology, Chengdu, China

Convolutional Neural Network (CNN) is widely used in seismic data denoising due to its simplicity and effectiveness. However, traditional seismic denoising methods based on CNN ignore multi-scale features of seismic data in the wavelet domain. The lack of these features will decrease the accuracy of denoising results. To address this barrier, a seismic denoise method based on the wavelet-inspired invertible network with atrous convolutions spatial pyramid (WINNet_ACSP) is proposed. WINNet_ACSP follows the principle of lifting wavelet transform. The proposed method utilizes the redundant orthogonal wavelet transform to obtain frequency multi-scale information from noisy seismic data. Then predict update network (PUNet) extracts spatial multi-scale features of approximate and detailed parts. The sparse driven network (SDN) learns the complex multi-scale information and obtains sparse features. These sparse features are processed to eliminate random noise. Compared to standard convolution, the atrous convolutions spatial pyramid (ACSP) can extract more features. The redundant features are the key to ensure the precision of multi-scale information. Therefore, the introduction of ACSP in PUNet can guarantee the denoising effect of the network. WINNet_ASCP combines the characteristics of wavelet transform and neural network and has a high generalization. Besides, transfer learning is used to overcome the difficulty caused by the training sample size of seismic data. The training process includes pre-training and post-training. The former is trained to obtain the initial denoising network by natural image samples. The latter is trained with a small sample of seismic data to enhance stratigraphic continuity. Finally, the proposed method is tested with synthetic and field data. The experimental results show that the proposed method can effectively remove random noise and reduce the loss of detailed information in prestack seismic data. In the future, we will make further improvements on this basis and conduct experiments on 3D prestack data.

KEYWORDS

random noise, seismic, wavelet transform, neural network, atrous convolution

1 Introduction

Random noise can affect the signal-to-noise ratio (SNR) of seismic data. Suppressing random noise and improving the SNR of seismic data is a critical step in seismic signal processing (Kong and Peng, 2015; Saad and Chen, 2020; Zhong et al., 2021). So far, researchers have come up with many methods to suppress random noise. Common methods include predictive filtering (Chen and Ma, 2014; Liu et al., 2020; Wang et al., 2021), mode decomposition (Han and van der Baan, 2015; Gómez and Velis, 2016; Long et al., 2021), low-rank constraints (Anvari et al., 2017; Chen et al., 2017; Huang, 2022), and transform domain (Kesharwani et al., 2022; Xie et al., 2022; Zhang et al., 2022).

Predictive filtering exploits the predictability of seismic data to suppress random noise. Canales and Lu (1993) first time proved the feasibility of predictive filtering technology in seismic data denoising field. Chen and Sacchi (2017) proposed a predictive filtering approach to simultaneously suppress mixed noises. This approach utilizes the hybrid L1/L2 norm to design a robust M-estimate of a special autoregressive moving-average model. The experimental results show that the model can effectively remove the mixed noise. Besides, Liu and Li (2018) proposed an adaptive predictive filtering method for non-stationary seismic signals. This method utilizes streaming characteristics to speed up the computation and uses signal-to-noise orthogonalization to enhance the denoising ability. Experiments on field data demonstrate the superiority of the method.

Mode decomposition-based denoising methods use correlation to separate seismic data into signal and noise components. Cai et al. (2011) utilized empirical mode decomposition to denoise seismic record. The denoising result showed that mode decomposition can suppress random noise. Zhang and Hong (2019) proposed a random noise suppression method based on the complete ensemble empirical mode decomposition. The results show that complete ensemble empirical mode decomposition has high feature recognition ability in complex random desert noise. Wu et al. (2022) uses multivariate variational modal decomposition on the segmented seismic data. This method significantly improves the lateral continuity and SNR of the seismic data.

Low-rank constrained denoising methods utilize seismic data's low-rank property to remove random noise. Wang et al. (2018) proposed a Hankel low-rank approximate denoising approach. Hankel structure can enhance the seismic low-rank

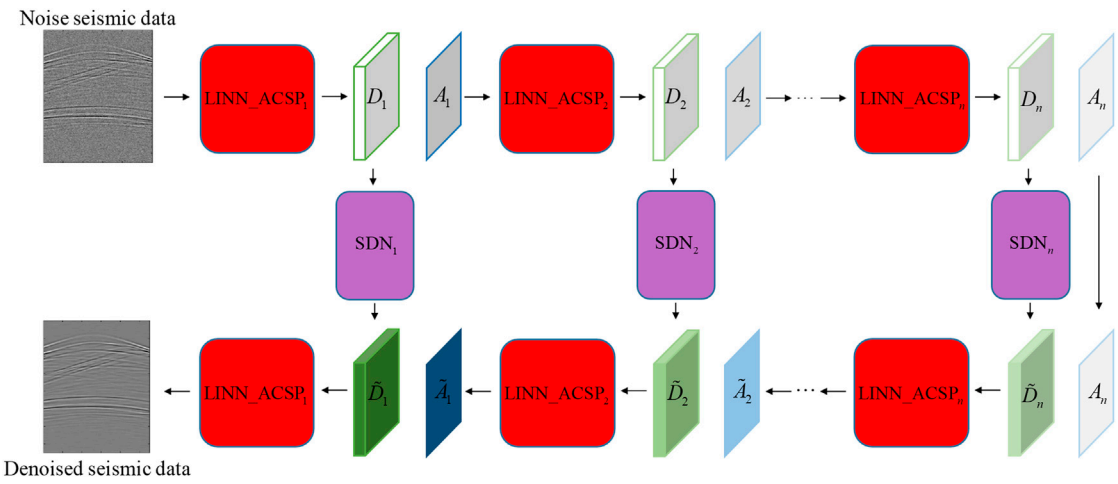


FIGURE 1
The network structure of WINNet_ACSP.

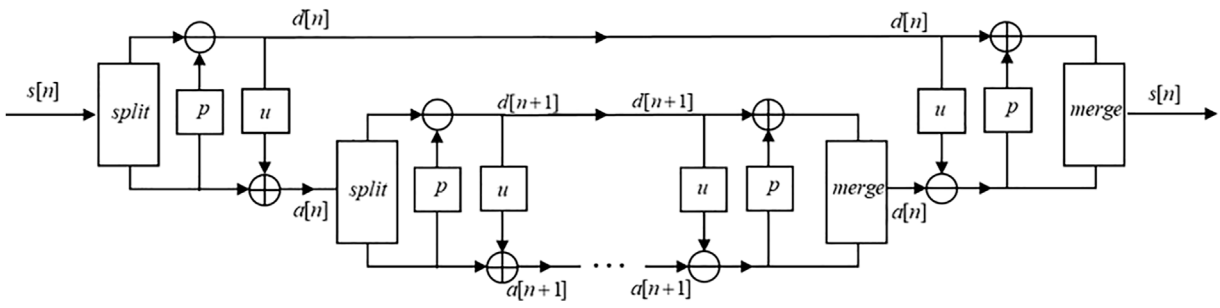


FIGURE 2
Lifting scheme wavelet transform.

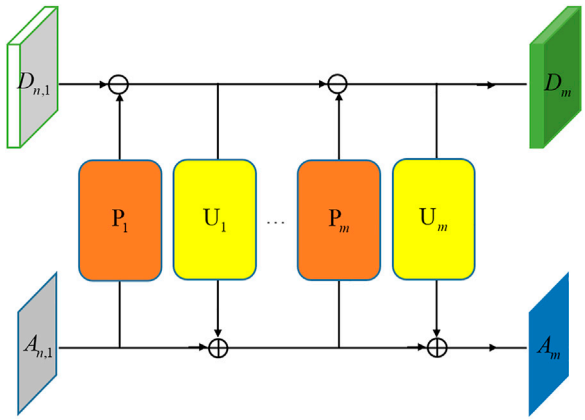


FIGURE 3
The process of predict and update.

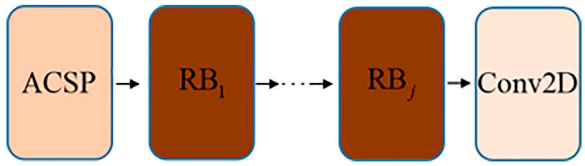


FIGURE 4
The structure of PUNet.

property. The enhanced low-rankness effectively removes random noise. Feng et al. (2021) proposed a denoising method using low-rank tensors. This method applies low-rank constraints to the seismic data tensor and improves the structural similarity of seismic data.

Transform domain-based denoising methods utilize the characteristics of seismic data in different transform domains to attenuate random noise. Zwartjes and Gisolf (2007) used Fourier transform to reconstruct seismic data. The high SNR reconstruction results demonstrate the feasibility of this method. Liang et al. (2018) proposed a denoising method based on the non-subsampled shearlet transform. The results show that the non-subsampled shearlet transform can suppress random noise and retain effective signals. Chen and Song (2018) used wavelet decomposition to decompose seismic data into multiple components. Then different threshold methods are applied to different seismic data components to achieve random noise suppression.

Predictive filtering, mode decomposition, low-rank constraints, and transform domain use the prior information of seismic data to construct suitable optimization strategies. Though these methods have good denoising ability and generalization abilities, denoising results are easily affected by human factors because of their large number of hyperparameters. To reduce the interference of human factors, researchers proposed the learning-based denoising method (Beckouche and Ma, 2014; Chen, 2017; Richardson and Feller, 2019; Yu et al., 2019). Dictionary learning and deep learning are commonly used strategies. Dictionary learning-based denoising methods train appropriate dictionary elements and linearly combine the elements to suppress random noise. Beckouche and Ma (2014) proposed a step-decomposable dictionary learning denoising method. The field data denoising result show that this method has a good denoising performance. Wang and Ma (2019) used the variation of noise variance in space to design a dictionary learning method with adaptive threshold parameters. The introduction of self-adaptation can realize blind denoising of seismic data and obtain signals with a high SNR. Kuruguntla et al. (2021) introduced a double sparse dictionary learning constraint to improve the denoising performance. This method combines the strength of the analytical transform and adaptive transform to suppress mixing noise. Chen et al. (2023) proposed a robust dictionary learning denoising method to reduce the loss of effective signal. This method retrieves leaked seismic signals by introducing a Huber-norm sparse coding model. Synthetic data and field data demonstrate the effectiveness of this method.

The denoising method based on deep learning distinguishes random noise from effective seismic signals by extracting the implicit features of seismic data through a neural network. Zhang et al. (2018) proposed a fast and flexible denoising convolutional neural network (FFDNet) to suppress noise. Numerous experimental results prove that FFDNet can flexibly and efficiently suppress random noise. Yu et al. (2019) attenuated the random and linear noise of complex

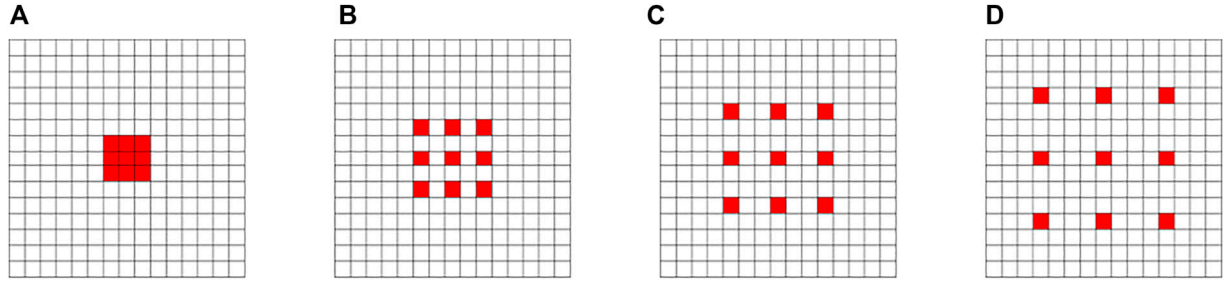


FIGURE 5
Atrous convolution with different dilation rates. **(A)** Dilation rate of 1; **(B)** Dilation rate of 2; **(C)** Dilation rate of 3; **(D)** Dilation rate of 4.

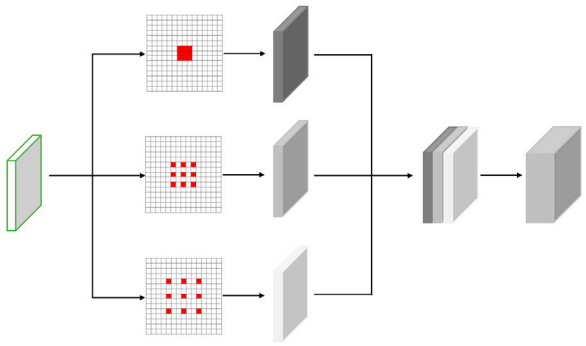


FIGURE 6
Atrous convolutions spatial pyramid.

seismic data using CNN. Experimental results prove the potential applications of CNN in suppression of random, linear, and multiples noise. Guo et al. (2019) proposed a convolutional blind denoising network (CBDNet) to eliminate random noise. The experimental results show that CBDNet can flexibly remove different levels of random noise by introducing a noise level estimation subnetwork. Sang et al. (2020) proposed a denoising method for multidimensional geological structure features based on the end-to-end deep denoising convolutional neural networks (DCNNs). DCNNs have a good denoising ability for complex geological structures, by extracting the characteristics of seismic data in different directions. Yang et al. (2021) proposed a denoising approach for 3-D seismic data by deep skip autoencoder. This approach uses the deep skip autoencoder to extract the waveform features of each seismic data patch. Feng and Li (2021) combined singular value decomposition (SVD) and neural networks to suppress noise interference in distributed acoustic sensing. The introduction of SVD improves the network' generalization and can accurately represent complex features in seismic data. Dong et al. (2022) utilized a spatial attention mechanism and convolutional neural network to distinguish weakly reflected seismic signals from strong random noise. The spatial attention further strengthens the denoising ability of the convolutional neural network.

Learning-based methods can extract various implicit features of seismic data. Through these implicit features, non-linear mapping of noiseless seismic data and noisy seismic data can be established. However, learning-based methods rarely take into account the advantages of other categories of methods approaches. For example, the wavelet transform threshold-based denoising method proved that the multi-scale features of seismic data can suppress random noise. But learning-based methods do not consider multi-scale information. The lack of multi-scale information results in a limited denoising effect of learning-based methods. To extract more abundant seismic information and improve the denoising effect, a wavelet-inspired invertible network with atrous convolutions spatial pyramid (WINNet_ACSP) is proposed for seismic denoising task. The proposed method consists of the lifting inspired invertible neural network with atrous convolutions spatial pyramid (LINN_ACSP) and sparse driven network (SDN). LINN_ACSP and SDN follow the principle of lifting wavelet transform and soft threshold operation, respectively. Therefore, LINN_ACSP inherits the multi-scale characteristic, sparsity, and perfect reconstruction characteristic of the lifting wavelet transform. Multi-scale features can ensure that the network effectively suppresses random noise. Sparsity can be exploited by soft-thresholding to distinguish random noise. Perfect reconstruction characteristic ensures that effective signals are not leaked. LINN_ACSP can obtain the frequency and spatial multi-scale information of seismic data through the splitting operator, prediction and update network (PUNet). The detail and approximate parts of the seismic data can be obtained by using this multi-scale information. Using the sparse detail part obtained by LINN_ACSP, the SDN learns to denoise the detail coefficients and obtains the denoised detail coefficients. Besides, to overcome the difficulty caused by the training sample size of seismic data, the proposed method utilizes transfer learning for training. Finally, the proposed method and other state-of-the-art methods are tested with synthetic and field seismic data. The experimental results show that the proposed method can effectively remove random noise and reduce the loss of detailed information in prestack seismic data.

2 Methods

Noisy seismic data can be expressed as follows

$$\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{N} \tag{1}$$

where $\boldsymbol{Y}$ represents noise-containing seismic data observed in the field, $\boldsymbol{X}$ denotes seismic data, and $\boldsymbol{N}$ indicates additive white Gaussian noise.

In this work, we propose a method for attenuating prestack seismic random noise using WINNet_ACSP. The LINN can obtain frequency multi-scale features in the wavelet domain (Huang and Dragotti, 2022). Embedding ACSP in LINN can extract spatial multi-scale features of the approximate or detail parts. LINN_

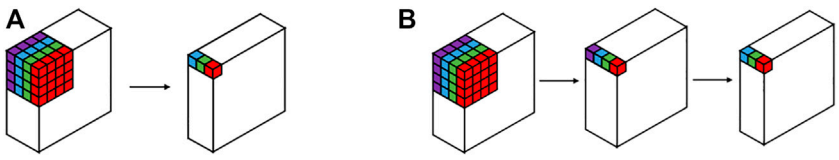


FIGURE 7
Convolution. **(A)** Standard convolution; **(B)** Depth-wise separable convolution.

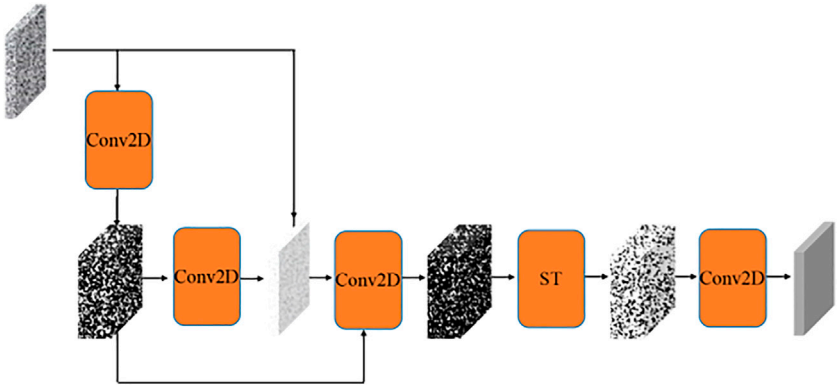


FIGURE 8
The structure of sparse driven network.

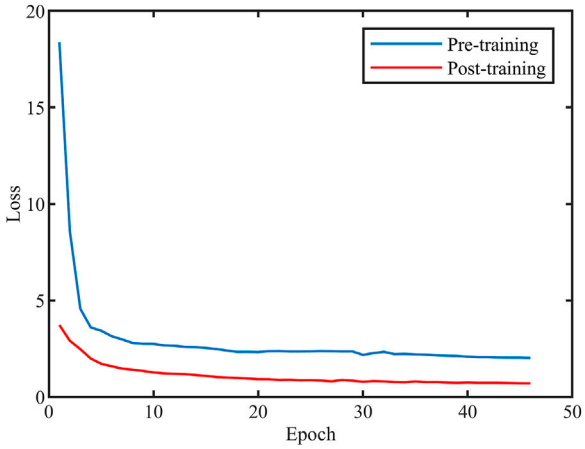


FIGURE 9
Training loss.

ACSP combines the characteristics of wavelet transform and neural networks. The entire network structure of WINNet_ACSP follows the wavelet threshold principle. At the same time, WINNet_ACSP as a neural network can realize non-linear mapping. The network structure of WINNet_ACSP is shown in Figure 1.

In Figure 1, LINN_ACSP represents lifting inspired invertible neural networks with atrous convolutions spatial pyramid, SDN denotes the sparse-driven network, $\boldsymbol{D}$ indicates the detail part, representing the boundary information, $\boldsymbol{A}$ is the approximate part, representing the smoothing information, n indicates the n-th scale, and the superscript ~ indicates the part after denoising.

WINNet_ACSP consists of LINN_ACSP and SDN. The forward pass of LINN_ACSP learns to perform a non-linear redundant transform on seismic data to obtain the multi-scale approximation part and detail part. The SDN learns to denoise the detail coefficients and obtains the denoised detail coefficients. Finally, using the backward pass of LINN_ACSP, the approximate part and the denoised detail part are reconstructed to obtain denoised seismic data.

2.1 LINN_ACSP

The denoising method based on wavelet transform can well remove the random noise in seismic data (Aghayan et al., 2016). The lifting scheme is known as the second-generation wavelet transform (Sweldens, 1998). The second-generation wavelet transform process can be divided into three steps: split, predict and update. Each step can be reconstructed by changing the direction and sign of the data flow. The splitting and merging process of the lifting scheme wavelet transform is shown in Figure 2.

In Figure 2, p represents the predict step, u denotes the update step, d [ ] indicates the detail part, a [ ] is the approximate part, s [ ] represents seismic data or approximate part, and n indicates the n-th scale.

However, for the split step, the lifting scheme wavelet transform uses a non-redundant transform. Affected by random noise, the non-redundant transformation will lose some important seismic information. For the prediction and update steps, the lifting scheme wavelet transform utilizes a simple linear formula and cannot accurately represent complex spatial features. For the above problems, some researchers use neural networks to complete the above requirements. LINN_ACSP is an invertible neural network with a structure inspired by the lifting scheme. LINN_ACSP inherits the sparsifying ability, perfect reconstruction characteristics, and multi-scale characteristics of the wavelet transform. Similar to the lifting scheme wavelet transform, LINN_ACSP consists of a splitting/merging operator, and a learnable predict and update network (PUNet).

2.1.1 Splitting/merging operator

LINN_ACSP uses redundant linear operators as a splitting operator, denoted as S. The split operator S is parameterized by a convolution kernel $K \in R^{c\times 1\times \rho\times \rho}$, where c denotes the number of channels and ρ denotes the spatial filter size.

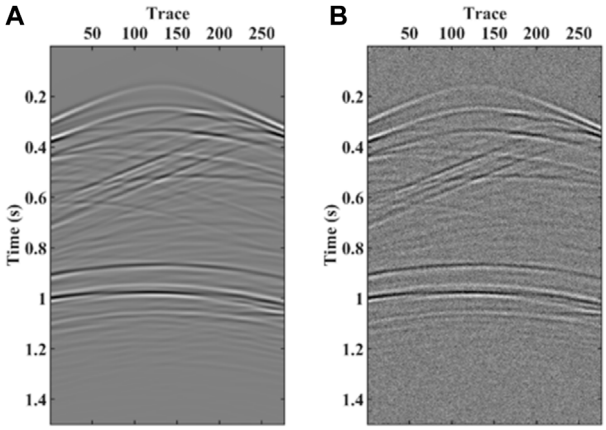


FIGURE 10
Synthetic Seismic Data. **(A)** Noise-free data; **(B)** Noisy data.

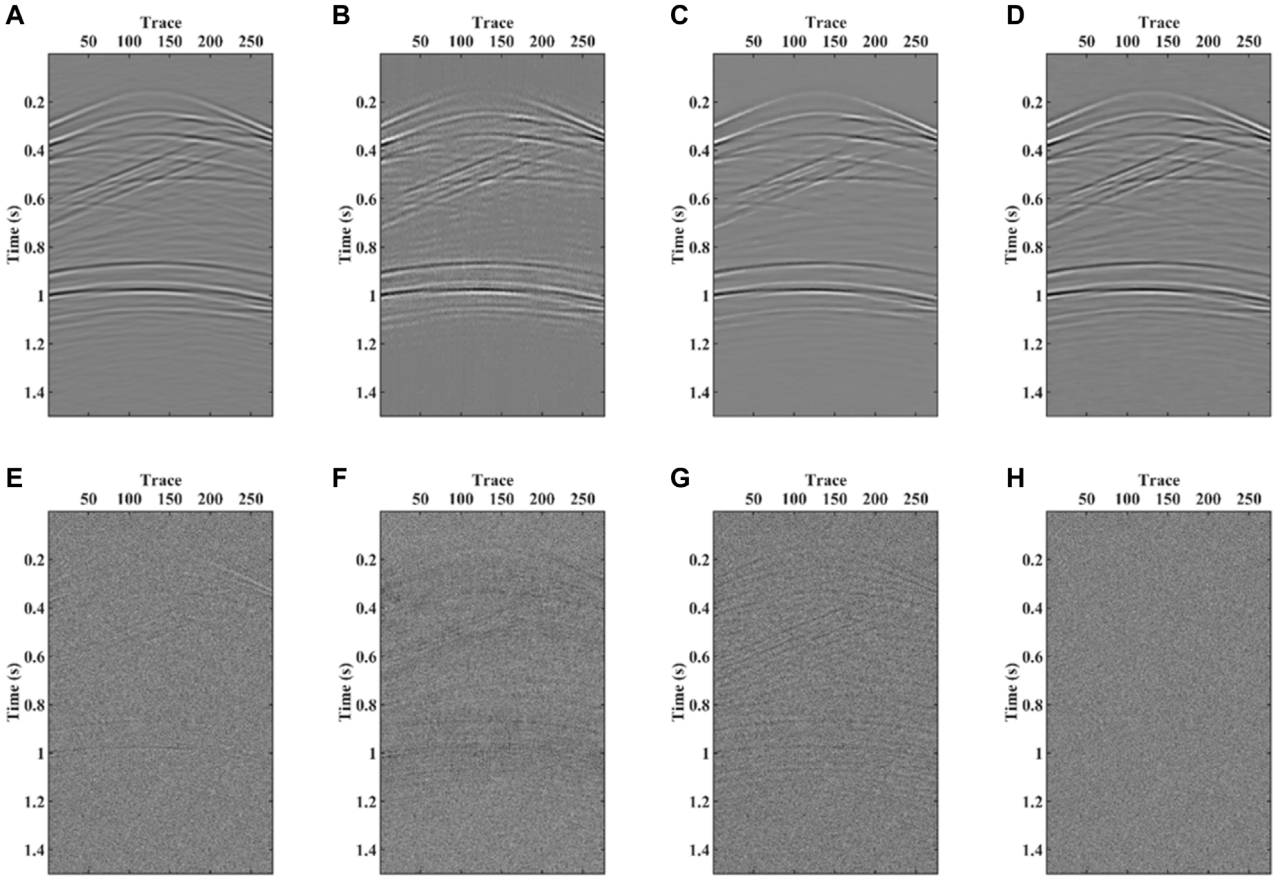


FIGURE 11
Synthetic seismic data denoising results section. **(A)** DMSSA; **(B)** SSWT-GoDec; **(C)** DnCNN; **(D)** The proposed method; **(E)** Removed noise by DMSSA; **(F)** Removed noise by SSWT-GoDec; **(G)** Removed noise by DnCNN; **(H)** Removed noise by the proposed method.

Using the redundant split operator to process the seismic data, the approximate part and the detail part with frequency multi-scale are obtained, as shown in the following formula

$$S(\boldsymbol{Y}) = (\boldsymbol{A}_1, \boldsymbol{D}_1) \tag{2}$$

where $\boldsymbol{A}_1$ represents the approximate part of the first scale, and $\boldsymbol{D}_1$ is the detail part of the first scale.

To ensure invertible, the merge operator M is parameterized by the transpose of the convolution kernel corresponding to the split operator. The merge operator reconstructs the

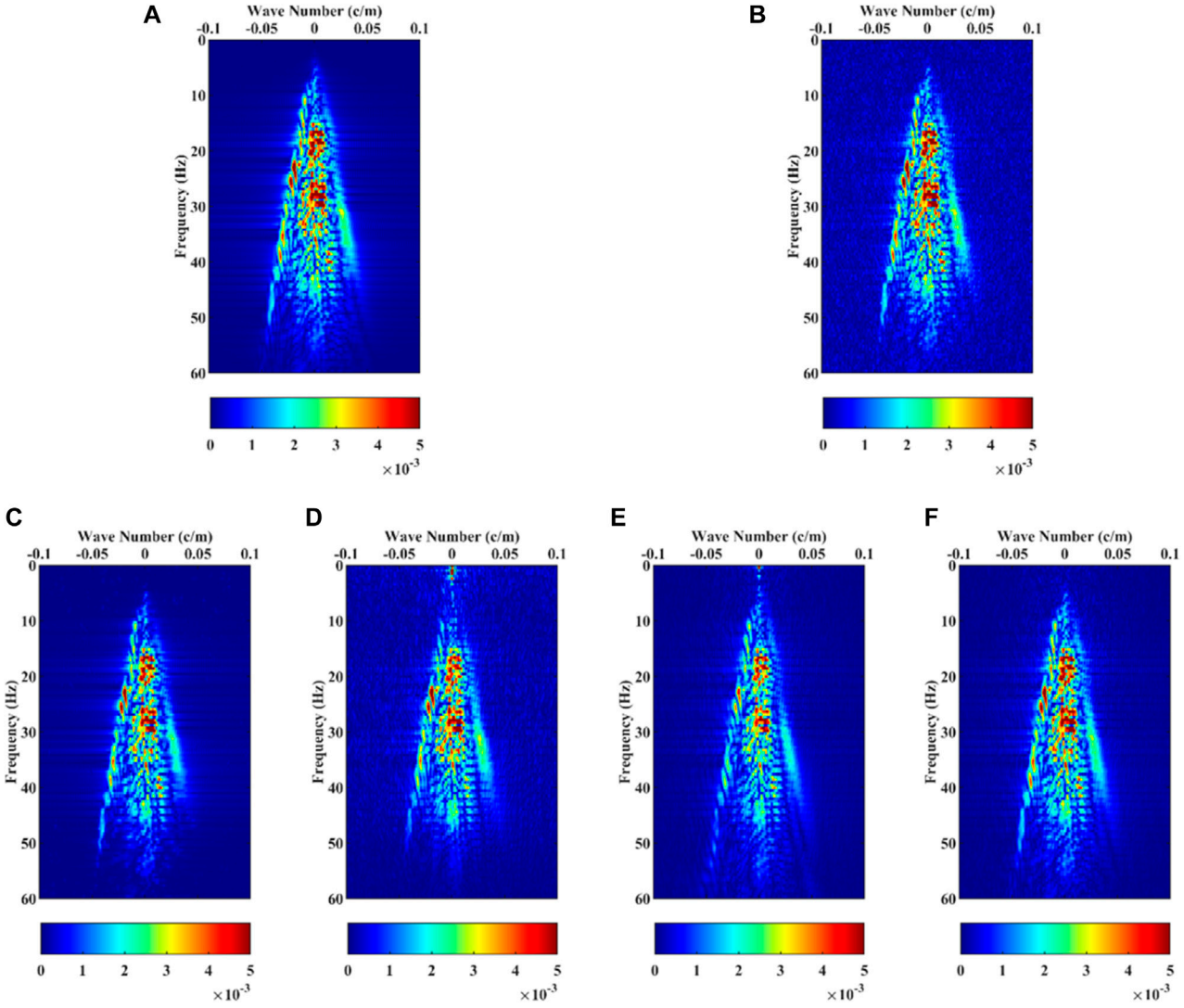


FIGURE 12
The F-K spectrum of the seismic data. **(A)** Clean data; **(B)** Noisy data; **(C)** DMSSA; **(D)** SSWT-GoDec; **(E)** DnCNN; **(F)** The proposed method.

approximate part and detailed part into seismic data. It can be defined as

$$M(\{\boldsymbol{A}_1, \boldsymbol{D}_1\}) = \boldsymbol{Y} \tag{3}$$

where { } is the concatenation operation.

Redundant representation can effectively reduce the leakage of seismic information and improve the stability of reconstruction results. Considering the waveform characteristics of seismic records, the sym2 wavelet is used to construct the convolution kernel K.

2.1.2 PUNet

LINN_ACSP uses a learnable convolutional neural network with ACSP to imitate the prediction and update operations in the lifting scheme wavelet transform. This convolutional neural network is named PUNet. PUNet can adaptively learn the corresponding non-linear features of the approximate part and the detailed part. These non-linear features are used to predict the detail part and update the approximate part. Completing one prediction and update process can be called one lifting step. Suppose there are m pairs of PUNet, in the n-th scale. The m times of lifting steps are shown in Figure 3.

In Figure 3, P represents the predict network, U denotes the update network, and the subscript m indicates the m-th lifting step. This paper sets $m = 4$.

In the forward transform, the approximate part and the detail part of the seismic data are non-linearly transformed by the neural network into a representation that is easier to denoising. For the approximate part $\boldsymbol{A}_n$ and detailed part $\boldsymbol{D}_n$ split in the n-th scale, the predict network uses the correlation between the approximate part and the detail part to perform prediction operation on the

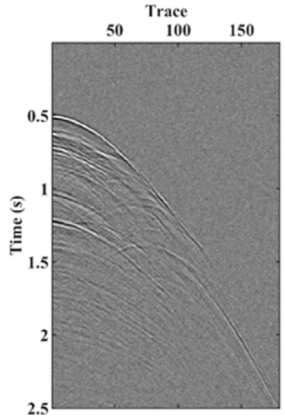


FIGURE 13
Field seismic data.

approximate part. The m-th pairs predict operation can be expressed as

$$\boldsymbol{D}_{n,m} = \boldsymbol{D}_{n,m-1} - \mathrm{P}_{n,m}\left(\boldsymbol{A}_{n,m-1}\right) \tag{4}$$

The purpose of the predict network is to make $\boldsymbol{D}_{n,m}$ sparser.

The update network act on the detail part to obtain the update result. Add the updated result and the approximate part to get the adjusted approximate part. The m-th pairs update operation can be expressed as

$$\boldsymbol{A}_{n,m} = \boldsymbol{A}_{n,m-1} + \mathrm{U}_{n,m}\left(\boldsymbol{D}_{n,m-1}\right) \tag{5}$$

The purpose of the update network is to make the approximate part $A_{n,m}$ smoother.

In the backward transform, the denoised detail part and approximate part are reconstructed back to the original domain by the same set of m pairs PUNet used in the forward transform. The formula is as follows

$$\boldsymbol{D}_{n,m-1} = \boldsymbol{D}_{n,m} + \mathrm{P}_{n,m}\left(\boldsymbol{A}_{n,m-1}\right) \tag{6}$$

$$\boldsymbol{C}_{n,m-1} = \boldsymbol{C}_{n,m} - \mathrm{U}_{n,m}\left(\boldsymbol{D}_{n,m-1}\right) \tag{7}$$

2.1.3 Structure of PUNet

To accurately predict and update the detail and approximate parts, PUNet needs to extract spatial multi-scale features of the detail and approximate parts. So PUNet is constructed by ACSP, residual blocks with depth-wise separable convolution, and the soft-thresholding operator approximated as the activation function. The network structure of PUNet is shown in Figure 4.

In Figure 4, ACSP represents atrous convolutions spatial pyramid, RB indicates residual block with depth-wise separable convolution, Conv2D is the 2D convolutional layer, and the subscript j represents j-th RB. This paper sets $j = 4$.

2.1.3.1 Atrous convolutions spatial pyramid

Atrous convolution is also called dilated convolution. Atrous convolution can change the receptive field by changing the dilation rate without increasing the number of convolution kernel parameters. The convolution kernel of atrous convolution is equivalent to inserting zeros between adjacent filter values in the horizontal or vertical direction of the convolution kernel of standard convolution. As shown in Figure 5, the larger the dilation rate, the larger the receptive field of the atrous convolution.

ACSP contains multiple parallel branches of the atrous convolutions with different dilation rates, shown in Figure 6. ACSP can extract spatial multi-scale features of approximate part and detail part (Ma et al., 2019). These spatial multi-scale features are fused by 1 × 1 convolution and input to the residual block.

2.1.3.2 Residual block with depth-wise separable convolution

The residual block directly stacks the input on the output through the skip connection to realize the feature fusion of the current module and the previous module. Feature fusion can solve the gradient vanishing problem during neural network training. Specifically, the residual block converts the original mappings that need to be learned into residual mappings, as shown in Eq. 8. And residual maps are easier to optimize for neural networks.

$$R(z) = O(z) - z \tag{8}$$

where z represents input features, $O(z)$ indicates original mapping, and $R(z)$ is residual mapping.

So residual learning can improve the stability of the network and allow more layers to be stacked to enhance the learning ability of the network.

Depth-wise separable convolution can reduce residual block parameters and ensure the accuracy of feature extraction by dividing the standard convolution operation into two parts (Chollet, 2017), as shown in Figure 7. The first part is the depth-wise convolution. The second part is the 1 × 1 convolution. Depth-wise convolution performs a separate convolution on each channel. The 1 × 1 convolution integrates all channel information. When the number of channels and the size of the convolution kernel are large, depth-separable convolution can effectively reduce memory and time costs during training.

2.1.3.3 Soft-thresholding

The soft-thresholding activation function expression is as Eq. 9

$$ST_{\lambda}(z) = \mathrm{sgn}(z)\max\left(|z| - \lambda, 0\right) \tag{9}$$

where, ST represents soft-thresholding operations, z is input features, and λ is a hyperparameters.

The soft-thresholding operator can be regarded as a two-sided ReLU function. Therefore, for seismic data with peaks and troughs, the soft-thresholding is more suitable as a non-linearization operator.

2.2 Sparse driven network (SDN)

The sparse driven network (SDN) consists of convolutional layers and soft-threshold sparse operators. For the detail parts

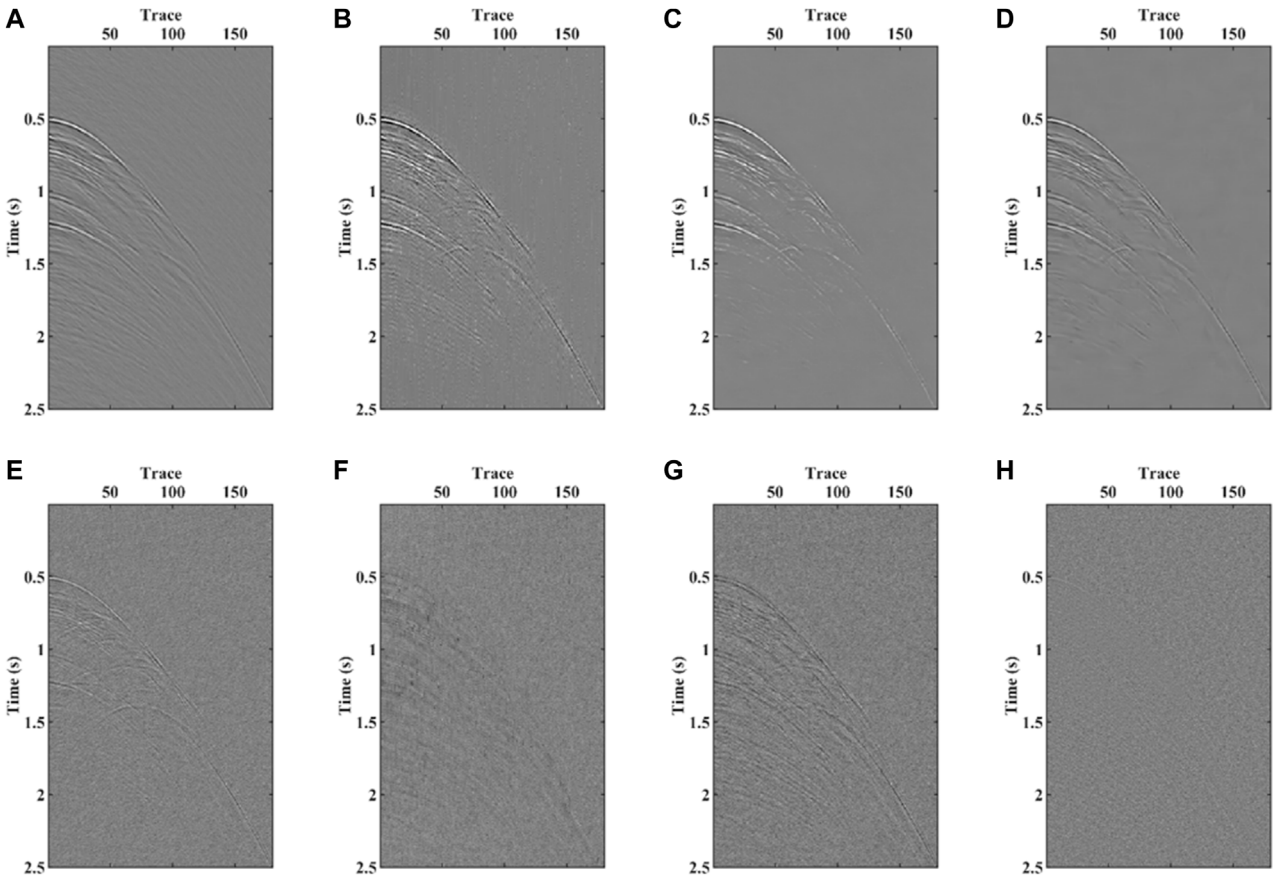


FIGURE 14
Field seismic data denoising results section. **(A)** DMSSA; **(B)** SSWT-GoDec; **(C)** DnCNN; **(D)** The proposed method; **(E)** Removed noise by DMSSA; **(F)** Removed noise by SSWT-GoDec; **(G)** Removed noise by DnCNN; **(H)** Removed noise by the proposed method.

at each scale, the denoising operation of the sparse drive network does not directly perform simple soft thresholding on the detail coefficients. The purpose of the sparse-driven network is to first utilize convolutional layers to transform the detail parts at each scale into a domain more suitable for denoising. In this domain, the eigencoefficient of the effective signal is made larger, and the eigencoefficient of random noise is made smaller. All feature coefficients are then processed using a learnable soft threshold operator. Finally, a convolutional layer is used to convert the feature coefficients back to the domain corresponding to the detail part. The network structure of SDN is shown in Figure 8.

2.3 Network training

To overcome the problem of the training sample size of seismic data, transfer learning (Pan and Yang, 2009) is used in this paper. The training process is divided into pre-training and post-training. In the pre-training step, a dataset of natural images is used to train the network. The pre-training can teach LINN_ACSP and SDN how to predict updates and denoising, respectively. In the post-training step, a small sample of seismic data is used for training to fine-tune the network. To reduce computer consumption, the dataset size is divided into 50×50 as the input of the neural network. The optimizer uses Adaptive Moment Estimation with a learning rate of 0.001 in the pre-training and 0.0001 in the post-training. Figure 9 is the training loss.

3 Examples and results

3.1 Evaluation of denoising performance

The SNR can directly reflect the quality of denoising results, it is defined as

$$\mathrm{SNR} = 10\log_{10}\frac{\|\boldsymbol{X}\|_{\mathrm{F}}^2}{\|\boldsymbol{X} - \boldsymbol{X}_{denoise}\|_F^2} \tag{10}$$

where $\boldsymbol{X}_{denoise}$ is the estimated or denoised seismic data.

SNR can evaluate the denoising effect of various methods as a whole. However, calculating SNR requires noise-free seismic data. So the SNR cannot be calculated in field

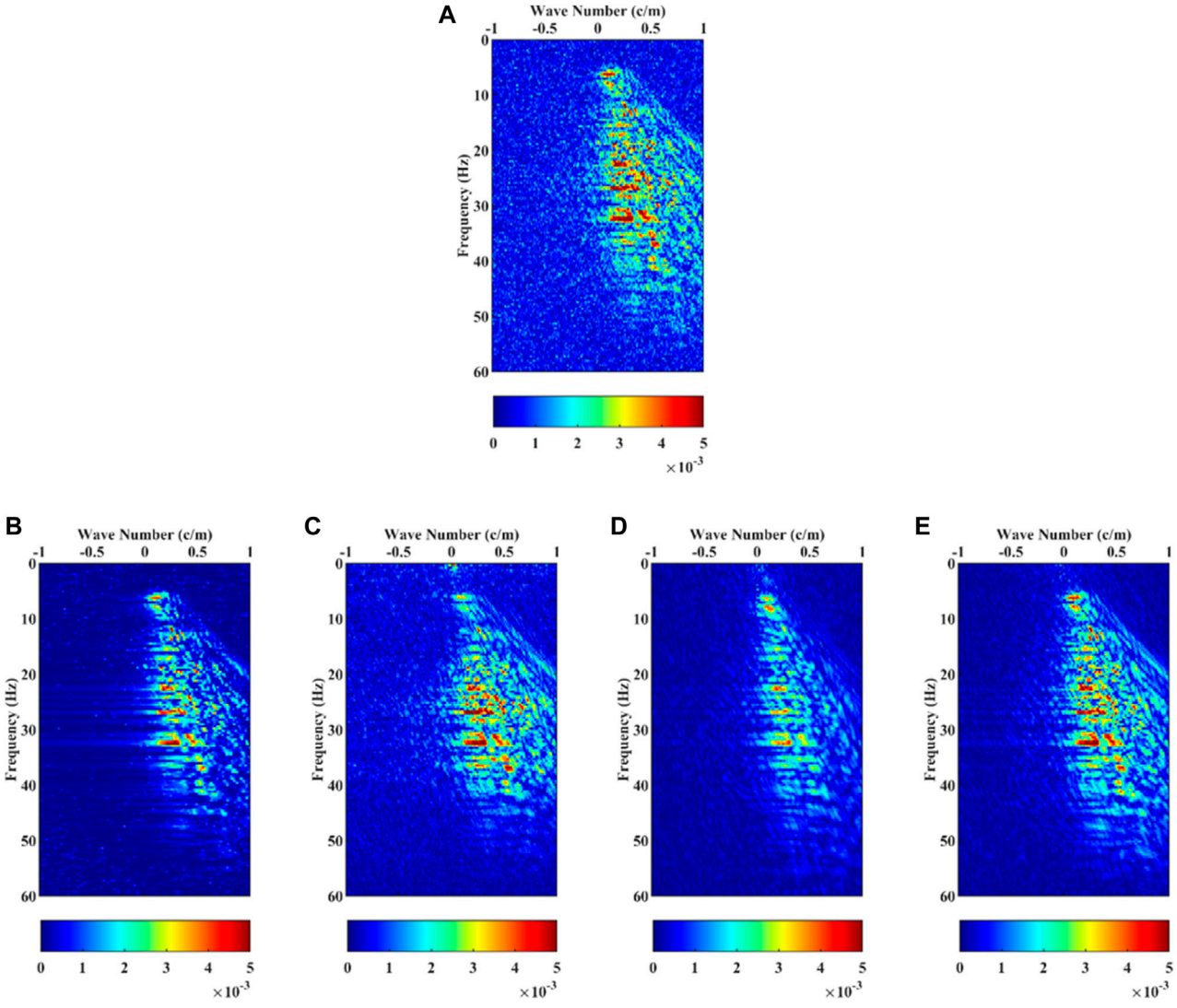


FIGURE 15
The F-K spectrum of the field seismic data. **(A)** Noisy data; **(B)** DMSSA; **(C)** SSWT-GoDec; **(D)** DnCNN; **(E)** The proposed method.

seismic data tests. To comprehensively evaluate the denoising results, the F-K spectrum is utilized to evaluate the denoising effect of various methods, too. F-K spectrum can analyze the advantages and disadvantages of various methods in terms of frequency.

3.2 Synthetic two-dimensional (2-D) seismic data

The marmousi2 P wave velocity model was used as the forward model. Combined with the first-order stress-velocity-sound wave equation, 31 shot synthetic data that conform to the law of field seismic data are obtained. The synthetic seismic data of each shot contains 277 traces, each trace has 3,000 sampling points, the sampling interval is 0.5 ms, and the domain frequency range is 20–30 Hz. The 30 shot synthetic seismic data were randomly selected as the post-training dataset. The selected seismic data of each shot is divided into datasets of size 50*50, as the input of the neural network. The remaining one-shot synthetic seismic data, shown in Figure 10, was used to test the denoising effect of the proposed method and other methods. Then AWGN was added to seismic data to generate noisy seismic data with SNR = −2 dB.

To evaluate the denoising effect of the proposed method, three state-of-the-art seismic denoising methods are used for comparison. Figures 11A–D are the denoising results of *f-x* damped multichannel singular spectrum analysis (DMSSA), SSWT-GoDec method, DnCNN and the proposed method, respectively. Figure 11A shows that the random noise in the sanction of the denoising results based on DMSSA is effectively suppressed. And this method does not cause the waveform

distortion of the seismic effective signal. The SNR of the DMSSA is 6.9 dB. Figure 11B contains a lot of random noise, and the waveform of the seismic signal is distorted. The SNR of the SSWT-GoDec is 3.3 dB. Figure 11C shows that DnCNN can suppress random noise, but also weaken the continuity of effective seismic signals. The SNR of the SSWT-GoDec is 4 dB. Figure 11D shows that the proposed method can effectively suppress random noise without causing distortion of the effective seismic signal waveform, nor weakening the continuity of the signal. The SNR of the proposed method is 7.3 dB. This result proves that the use of multi-scale features can improve the denoising effect of the neural network.

The removed noise section of the above method is shown in Figure 11E–H. Comparing Figure 11E–H the results show that there is obvious seismic reflection information in the whole removed noise section based on SSWT-GoDec and DnCNN. The denoising method based on DMSSA, when affected by random noise, will leak effective signals when recovering high-amplitude seismic signals. Finally, the seismic signal leakage cannot be observed in the removed noise section corresponding to the proposed method. This result demonstrates that the use of multi-scale features can prevent the leakage of valid seismic signals.

Figures 12A, B shows the F-K spectra of clean and noisy seismic data, respectively. Figure 12C–F are the F-K spectrum of the denoising results of the above methods. Figure 12C is the F-K spectrum obtained by DMSSA denoising. Comparing Figures 12A, C, when the frequency is higher than 30 Hz, the amplitude of the F-K spectrum shown in Figure 12C is smaller than the corresponding F-K spectrum of the noise-free seismic data. This result shows that the DMSSA-based denoising method loses high-frequency information, that is, leakage occurs when the seismic signal changes from low amplitude to high amplitude. Figure 12D is the F-K spectrum obtained by SSWT-GoDec denoising. Comparing Figures 12A, D, when the frequency is lower than 20 Hz, the F-K spectrum shown in Figure 12D is less consistent with Figure 12A. This result shows that the denoising method based on SSWT-GoDec will change the low-frequency information, that is, the waveform of the seismic signal is distorted. Figure 14E is the F-K spectrum obtained by DnCNN denoising. Comparing Figures 12A, E, the overall magnitude of the F-K spectrum shown in Figure 12E is lower than that in Figure 12A. The results show that the denoising method based on DnCNN will leak the effective seismic signal. Figure 12F is the F-K spectrum obtained by the proposed method for denoising. Comparing Figures 12A, F, the F-K spectrum shown in Figure 12F has the highest similarity with Figure 12A. The results show that the proposed method can effectively remove random noise and protect critical seismic signals.

3.3 Application on field seismic data

To verify the effectiveness of the proposed method in field seismic data, the single shot data shown in Figure 13 are used for testing. This single shot data contains 180 traces, each trace has 500 sampling points, and the sampling interval is 0.005 s.

Figure 14 shows the denoising results and removed noise section of the proposed method and other methods, respectively. Observing Figure 14, the results show that the DMSSA-based denoising method will seriously leak the effective signal. SSWT-GoDec-based denoising approach cause waveform distortion and lateral discontinuities. DnCNN-based denoising methods lose effective signals. The proposed method can effectively suppress random noise and retain valid signals.

Figure 15 shows the F-K spectrum of the field single shot seismic data and denoising results. The F-K spectrum amplitudes of DMSSA and DnCNN denoising results are low, again indicating that the effective signal will leak. The F-K spectrum of the SSWT-GoDec denoising results has a small amplitude in the low frequency part, which confirms the waveform distortion. The amplitude of the F-K spectrum of the denoising result of the proposed method is appropriate and focused. This result proves that the proposed method can effectively suppress random noise and retain important seismic signals.

4 Conclusion

This paper proposes a denoising method for prestack seismic data using WINNet_ACSP. This method can effectively suppress random noise and prevent the leakage of important seismic information. In the forward pass of WINNet_ACSP, the first step uses a redundant transformation to split the seismic data to obtain frequency multi-scale approximate and detail parts. The second step utilizes a learnable neural network with ACSP to extract spatial features for the approximate or detail parts. The third step uses the sparse drive network to process the coefficients of the detail part. Finally, the denoised seismic data is reconstructed using the backward pass of WINNet_ACSP. The whole denoising process follows the principle of wavelet transformation. The combination of redundant transformation and ACSP can obtain richer multi-scale information. These multi-scale features can effectively suppress random noise and retain important seismic information. Transfer learning divides the training process into pre-training and post-training. The former is trained using natural images. The latter is trained using a small amount of seismic data. Synthetic and field seismic data are utilized to test the proposed method and other methods. The results show that the proposed method can effectively suppress random noise, improve the SNR of seismic data, and prevent the leakage of effective signals. In the future, we will make further improvements on this basis and conduct experiments on 3D prestack data.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

LH proposes method and writes manuscript; HW proposes and participates in design research and reviews paper; XW participates in design research and reviews paper. All the authors approved the final version of the manuscript.

Funding

This work was supported by the Natural Science Foundation Project of Sichuan Province under Grant 22NSFSC3773.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Aghayan, A., Jaiswal, P., and Siahkoohi, H. R. (2016). Seismic denoising using the redundant lifting scheme. *Geophysics* 81, V249–V260. doi:10.1190/geo2015-0601.1

Anvari, R., Siahsar, M. a. N., Gholtashi, S., Kahoo, A. R., and Mohammadi, M. (2017). Seismic random noise attenuation using synchrosqueezed wavelet transform and low-rank signal matrix approximation. *IEEE Trans. Geosci. Remote Sens.* 55, 6574–6581. doi:10.1109/tgrs.2017.2730228

Beckouche, S., and Ma, J. (2014). Simultaneous dictionary learning and denoising for seismic data. *Geophysics* 79, A27–A31. doi:10.1190/geo2013-0382.1

Cai, H., He, Z., and Huang, D. (2011). Seismic data denoising based on mixed time-frequency methods. *Appl. Geophys.* 8, 319–327. doi:10.1007/s11770-011-0300-6

Canales, L. L., and Lu, L. (1993). "3-D trace interpolation in the fxp domain," in *63th Annual International Meeting* (Tulsa, OK, United States: Society of Exploration Geophysicists), 1174–1176.

Chen, K., and Sacchi, M. D. (2017). Robust f-x projection filtering for simultaneous random and erratic seismic noise attenuation. *Geophys. Prospect.* 65, 650–668. doi:10.1111/1365-2478.12429

Chen, W., and Song, H. (2018). Automatic noise attenuation based on clustering and empirical wavelet transform. *J. Appl. Geophys.* 159, 649–665. doi:10.1016/j.jappgeo.2018.09.025

Chen, Y., and Ma, J. (2014). Random noise attenuation by fx empirical-mode decomposition predictive filtering. *Geophysics* 79, V81–V91. doi:10.1190/geo2013-0080.1

Chen, Y. (2017). Fast dictionary learning for noise attenuation of multidimensional seismic data. *Geophys. J. Int.* 209, 21–31. doi:10.1093/gji/ggw492

Chen, Y., Zhou, Y., Chen, W., Zu, S., Huang, W., and Zhang, D. (2017). Empirical low-rank approximation for seismic noise attenuation. *IEEE Trans. Geosci. Remote Sens.* 55, 4696–4711. doi:10.1109/tgrs.2017.2698342

Chen, W., ObouÃ©, Y. a. S. I., and Chen, Y. (2023). Retrieving useful signals from highly corrupted erratic noise using robust residual dictionary learning. *Geophysics* 88, WA55–WA64. doi:10.1190/geo2022-0168.1

Chollet, F. O. (2017). "Xception: Deep learning with depthwise separable convolutions," in Proceedings of the IEEE conference on computer vision and pattern recognition (IEEE), 1251–1258.

Dong, X., Lin, J., Lu, S., Wang, H., and Li, Y. (2022). Multi-scale spatial attention network for seismic data denoising. *IEEE Trans. Geosci. Remote Sens.* 60, 212. doi:10.1109/TGRS.2022.3178212

Feng, J., Liu, X., Li, X., Xu, W., and Liu, B. (2021). Low-rank tensor minimization method for seismic denoising based on variational mode decomposition. *IEEE Geosci. Remote Sens. Lett.* 19, 1–5. doi:10.1109/lgrs.2021.3100262

Feng, Q., and Li, Y. (2021). Denoising deep learning network based on singular spectrum analysis—DAS seismic data denoising with multichannel SVDDCNN. *IEEE Trans. Geosci. Remote Sens.* 60, 1–11. doi:10.1109/tgrs.2021.3071189

Gómez, J. L., and Velis, D. R. (2016). A simple method inspired by empirical mode decomposition for denoising seismic data. *Geophysics* 81, V403–V413. doi:10.1190/geo2015-0566.1

Guo, S., Yan, Z., Zhang, K., Zuo, W., and Zhang, L. (2019). "Toward convolutional blind denoising of real photographs," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (IEEE), 1712–1722.

Han, J., and Van Der Baan, M. (2015). Microseismic and seismic denoising via ensemble empirical mode decomposition and adaptive thresholding. *Geophysics* 80, KS69–KS80. doi:10.1190/geo2014-0423.1

Huang, J., and Dragotti, P. L. (2022). WINNet: Wavelet-inspired invertible network for image denoising. *IEEE Trans. Image Process.* 31, 4377–4392. doi:10.1109/tip.2022.3184845

Huang, W. (2022). A two-step singular spectrum analysis method for robust low-rank approximation of seismic data. *Geophysics* 88, V9–V19. doi:10.1190/geo2021-0781.1

Kesharwani, A., Aggarwal, V., Singh, S., Br, R., and Kumar, A. (2022). Marine seismic signal denoising using VMD with Hausdorff distance and wavelet transform. *J. Def. Model. Simul.* 19, 739–746. doi:10.1177/15485129211036044

Kong, D., and Peng, Z. (2015). Seismic random noise attenuation using shearlet and total generalized variation. *J. Geophys. Eng.* 12, 1024–1035. doi:10.1088/1742-2132/12/6/1024

Kuruguntla, L., Dodda, V. C., and Elumalai, K. (2021). Study of parameters in dictionary learning method for seismic denoising. *IEEE Trans. Geosci. Remote Sens.* 60, 1–13. doi:10.1109/tgrs.2021.3107541

Liang, X., Li, Y., and Zhang, C. (2018). Noise suppression for microseismic data by non-subsampled shearlet transform based on singular value decomposition. *Geophys. Prospect.* 66, 894–903. doi:10.1111/1365-2478.12576

Liu, B., Fu, C., Ren, Y., Zhang, Q., Xu, X., and Chen, Y. (2020). Structural complexity-guided predictive filtering. *Geophys. Prospect.* 68, 1509–1522. doi:10.1111/1365-2478.12941

Liu, Y., and Li, B. (2018). Streaming orthogonal prediction filter in the t-x domain for random noise attenuation. *Geophysics* 83, F41–F48. doi:10.1190/geo2017-0322.1

Long, L., Wen, X., and Lin, Y. (2021). Denoising of seismic signals based on empirical mode decomposition-wavelet thresholding. *J. Vib. Control* 27, 311–322. doi:10.1177/1077546320926846

Ma, J., Dai, Y., and Tan, Y.-P. (2019). Atrous convolutions spatial pyramid network for crowd counting and density estimation. *Neurocomputing* 350, 91–101. doi:10.1016/j.neucom.2019.03.065

Pan, S. J., and Yang, Q. (2009). A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* 22, 1345–1359. doi:10.1109/tkde.2009.191

Richardson, A., and Feller, C. (2019). *Seismic data denoising and deblending using deep learning. arXiv preprint arXiv:1907.01497*.

Saad, O. M., and Chen, Y. (2020). Deep denoising autoencoder for seismic random noise attenuation. *Geophysics* 85, V367–V376. doi:10.1190/geo2019-0468.1

Sang, W., Yuan, S., Yong, X., Jiao, X., and Wang, S. (2020). DCNNs-based denoising with a novel data generation for multidimensional geological structures learning. *IEEE Geosci. Remote Sens. Lett.* 18, 1861–1865. doi:10.1109/lgrs.2020.3007819

Sweldens, W. (1998). The lifting scheme: A construction of second generation wavelets. *SIAM J. Math. Anal.* 29, 511–546. doi:10.1137/s0036141095289051

Wang, C., Zhu, Z., Gu, H., Wu, X., and Liu, S. (2018). Hankel low-rank approximation for seismic noise attenuation. *IEEE Trans. Geosci. Remote Sens.* 57, 561–573. doi:10.1109/tgrs.2018.2858545

Wang, H., Chen, W., Huang, W., Zu, S., Liu, X., Yang, L., et al. (2021). Nonstationary predictive filtering for seismic random noise suppression—a tutorial. *Geophysics* 86, W21–W30. doi:10.1190/geo2020-0368.1

Wang, X., and Ma, J. (2019). Adaptive dictionary learning for blind seismic data denoising. *IEEE Geosci. Remote Sens. Lett.* 17, 1273–1277. doi:10.1109/lgrs.2019.2941025

Wu, G., Liu, G., Wang, J., and Fan, P. (2022). Seismic random noise denoising using mini-batch multivariate variational mode decomposition. *Comput. Intell. Neurosci.* 2022, 2132732. doi:10.1155/2022/2132732

Xie, X., Ma, X., Long, H., Ming, Y., and Sun, S. (2022). Curvelet transform-based denoising of resonance interference induced by electrical poles in seismic exploration. *Geophys. Geochem. Explor.* 46, 474–481. doi:10.11720/wtyht.2022.2411

Yang, L., Wang, S., Chen, X., Saad, O. M., Chen, W., ObouÃ©, Y. a. S. I., et al. (2021). Unsupervised 3-D random noise attenuation using deep skip autoencoder. *IEEE Trans. Geosci. Remote Sens.* 60, 1–16. doi:10.1109/tgrs.2021.3100455

Yu, S., Ma, J., and Wang, W. (2019). Deep learning for denoising. *Geophysics* 84, V333–V350. doi:10.1190/geo2018-0668.1

Zhang, K., Zuo, W., and Zhang, L. (2018). FFDNet: Toward a fast and flexible solution for CNN-based image denoising. *IEEE Trans. Image Process.* 27, 4608–4622. doi:10.1109/tip.2018.2839891

Zhang, Q., Chen, Y., Zhang, F., and Chen, Y. (2022). Improving receiver function imaging with high-resolution Radon transform. *Geophys. J. Int.* 230, 1292–1304. doi:10.1093/gji/ggac116

Zhang, Z., and Hong, W.-C. (2019). Electric load forecasting by complete ensemble empirical mode decomposition adaptive noise and support vector regression with quantum-based dragonfly algorithm. *Nonlinear Dyn.* 98, 1107–1136. doi:10.1007/s11071-019-05252-7

Zhong, T., Cheng, M., Dong, X., and Wu, N. (2021). Seismic random noise attenuation by applying multiscale denoising convolutional neural network. *IEEE Trans. Geosci. Remote Sens.* 60, 1–13. doi:10.1109/tgrs.2021.3095922

Zwartjes, P., and Gisolf, A. (2007). Fourier reconstruction with sparse inversion. *Geophys. Prospect.* 55, 199–221. doi:10.1111/j.1365-2478.2006.00580.x

frontiers | Frontiers in Earth Science

TYPE Original Research
PUBLISHED 23 February 2023
DOI 10.3389/feart.2023.1101601

OPEN ACCESS

EDITED BY
Peng Zhenming,
University of Electronic Science and Technology of China, China

REVIEWED BY
Cai Hanpeng,
University of Electronic Science and Technology of China, China
Hao Wu,
Chengdu University of Technology, China

*CORRESPONDENCE
Yang Liu,
wliuyang@vip.sina.com

SPECIALTY SECTION
This article was submitted to Environmental Informatics and Remote Sensing, a section of the journal Frontiers in Earth Science

RECEIVED 18 November 2022
ACCEPTED 13 February 2023
PUBLISHED 23 February 2023

CITATION
Cova D and Liu Y (2023), Shear wave velocity prediction using bidirectional recurrent gated graph convolutional network with total information embeddings.
Front. Earth Sci. 11:1101601.
doi: 10.3389/feart.2023.1101601

Shear wave velocity prediction using bidirectional recurrent gated graph convolutional network with total information embeddings

David Cova[1,2] and Yang Liu[1,2,3]*

[1]State Key Laboratory of Petroleum Resources and Prospecting, China University of Petroleum (Beijing), Beijing, China, [2]CNPC Key Laboratory of Geophysical Exploration, China University of Petroleum (Beijing), Beijing, China, [3]China University of Petroleum (Beijing) Karamay Campus, Karamay, China

Shear wave velocity is an essential elastic rock parameter for reservoir characterization, fluid identification, and rock physics model building. However, S-wave velocity logging data are often missing due to economic reason. Machine learning approaches have been successfully adopted to overcome this limitation. However, they have shortcomings in extracting meaningful spatial and temporal relationships. We propose a supervised data-driven method to predict S-wave velocity using a graph convolutional network with a bidirectional gated recurrent unit (GCN-BiGRU). This method adopts the total information coefficient to capture non-linear dependencies among well-log data and uses graph embeddings to extract spatial dependencies. Additionally, the method employs a bidirectional gated mechanism to map depth relationships in both upward and backward directions. Furthermore, the prediction performance is increased by an unsupervised graph neural network to handle outliers and the generation of additional features by the complete ensemble empirical mode decomposition with additive noise method. Finally, the GCN-BiGRU network is compared with Castagna's empirical velocity formula, support vector regression, long-short-term memory (LSTM), GRU, and BiGRU methods over the North Sea open dataset. The results show that the proposed method performs better predicting S-wave velocity than the other ML and empirical methods.

KEYWORDS

shear wave velocity prediction, graph convolutional network, bidirectional gated recurrent unit, total information coefficient, graph neural network, outlier removal, ensemble empirical mode decomposition with additive noise

Introduction

In reservoir characterization, shear wave (S-wave) velocity is an essential elastic property for building accurate rock physics models and discriminating fluid content in geologic formations (Xu and White, 1995; Vernik and Kachanov, 2010; Refunjol et al., 2022). However, the availability of measured S-wave velocity logs in exploration projects is frequently scarce for an economic reason (Anemangely et al., 2019). Statistical and empirical methods address this problem using compressional wave velocity correlations (Castagna et al., 1985; Greenberg and Castagna, 1992). Nevertheless, statistical methods, such as linear regression (LR), often have low accuracy and fail to capture the complex

relationships among the data. Moreover, empirical methods require additional information, such as mineral composition, pore aspect ratio, fluid saturation, total organic carbon content, or formation pressure, for proper calibration and accurate results (Vernik et al., 2018; Omovie and Castagna, 2021). In contrast, machine learning (ML) methods discover intrinsic relationships, make accurate predictions, and overcome data scarcity efficiently (Ali et al., 2021). ML methods have been applied for predicting S-wave velocity using well-log data, such as support vector regression (SVR) (Ni et al., 2017), long-short-term memory (LSTM) (Zhang et al., 2020), gated recurrent units (GRUs) (Sun and Liu, 2020), and gradient boosting (Zhong et al., 2021).

The S-wave velocity prediction is frequently addressed as a multivariate time series problem by assuming independence among variables and calculating a single depth point without further considerations (Jiang et al., 2018). Alternatively, the S-wave velocity prediction can be reframed as a supervised data-driven learning problem with recursive neural networks (RNNs) by considering the trend variations in the rock properties with depth (Hopfield, 1982). GRU is an improved RNN, less complex, and easier to train than LSTM (Cho et al., 2014). GRU dynamically extracts patterns from previous depth points to forecast rock properties in the following depth points (Chen et al., 2020). Bidirectional gated recurrent units (BiGRUs) with attention consist of two synchronous GRU to increase the prediction performance. The input sequence starts from the top to the bottom for the first unit and from the bottom to the top for the second unit. At the same time, the attention mechanism selects the most important features contributing to the prediction (Zeng et al., 2020). However, GRU has limitations in extracting local spatial characteristics from data (Jiang et al., 2021). Therefore, recent models combine convolutional neural network (CNN) layers to extract local and global features (Wang and Cao, 2021).

Graph theory receives particular attention for representing complex models surpassing the limitations of Euclidean space (Zhou F. et al., 2019a). A graph is a collection of vertices and edges that shares a relationship, represented by a Laplacian matrix (Scarselli et al., 2009). A graph embedding translates the latent dependencies from the graph into a low-dimensional space while preserving the original features, structure, and information (Hamilton et al., 2017). In this context, graph neural networks (GNNs) are a learning algorithm that handles graphs and resembles RNNs (Gori et al., 2005; Di Massa et al., 2006; Xu et al., 2019). Graph convolutional networks (GCNs) are first-order approximations of local spectral filters on graphs that perform convolution operations with linear computational complexity (Defferrard et al., 2016; Kipf and Welling, 2017). Furthermore, GCN-GRU has been successfully used for time-series prediction by exploiting the advantages of both graph and recurrent network architectures (Zhao et al., 2020).

We propose a graph recurrent gated method to predict S-wave velocity and compare it with other ML methods. For added value, the proposed method includes unsupervised distance-based outlier elimination with GNN, empirical mode decomposition (EMD) as feature engineering, and non-linear graph embedding with the total information coefficient (TIC) for more meaningful results. The workflow contains four steps:

1) An unsupervised GNN is used to detect outliers by learning the information from the nearest neighbor samples (Goodge et al., 2022). The goal is to remove the extreme values in the well-logging data resulting from human, environmental, or instrumental errors that impact the final prediction.
2) The well-logging data are decomposed into intrinsic mode functions (IMFs) by the complete ensemble EMD with additive noise (CEEMDAN) algorithm. The IMFs represent features from the local oscillation frequency with a physical meaning similar to Fourier domain spectral decomposition (Huang et al., 1998; Gao et al., 2022). Furthermore, they are concatenated with the well-logging data to form sequences for the network input.
3) The well-logging data are converted into the graph domain by mapping their dependencies with the TIC. TIC is a noise-robust correlation coefficient representing intrinsic non-linear dependencies among variables (Reshef et al., 2018).
4) A modified GCN-GRU network with bi-recurrent units and an attention mechanism predicts the S-wave velocity (Zhao et al., 2020). The GCN captures the spatial dependencies among the well-logging data. At the same time, the bidirectional GCN-GRU maps the sequence of previous and subsequent depth points for the S-wave velocity prediction (Yu et al., 2018).

Finally, the GCN-BiGRU network is compared with other ML methods, including SVR, LSTM, GRU, BiGRU, Castagna's empirical formula, and LR. The root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and R^2 metrics are used to evaluate the performance of the models. The results show that the proposed method has a lower error in predicting the S-wave velocity than the other ML and empirical methods.

Methodology

Local outlier removal with graph neural networks

Identifying and eliminating potential outliers is an essential step in S-wave velocity prediction. Among different methods, local outliers are widely adopted to detect anomalies in multivariate data by measuring the distance between neighbor points (Breunig et al., 2000; Amarbayasgalan et al., 2018). However, they lack trainable parameters to adapt to particular datasets. In contrast, the message-passing abilities of GNNs can detect anomalies and outperform local outlier methods by learning the information from the nearest samples without supervision (Goodge et al., 2022).

The GNN uses the message-passing abilities of a direct graph for detecting outliers. In general, a graph $G(V, E)$ is defined by a set of N vertices or nodes, $V = \{v_1, v_2, \ldots, v_N\}$, with N nodes features, $X = \{x_1, x_2, \ldots, x_N\}$, a set of M edges, $E = \{e_1, e_2, \ldots, e_M\}$, with edge features defined as $e_j = (v_i, v_k)$, where $v_i, v_k \in V$ (Zhou F. et al., 2019a). The message-passing characteristic allows the graph to send and receive information through its connections with its neighbors in one direction. The message-passing workflow involves a message function, an aggregation function, and an update function. Then, the hidden representation of a node is calculated by

$$h_{N_i} = a_k\left(m\left(x_i, x_k, e_j\right)\right) \tag{1}$$

where x_i is the feature of the source node v_i, x_k is the feature of the adjacent v_k, m is the message function that sends the information to each neighbor node, a_k is the aggregation function that summarizes the incoming messages from k adjacent nodes of v_i, $k \in N_i$, N_i is the number of adjacent nodes to v_i, and h_{N_i} is the aggregation of the messages from its neighbors. Finally, the update function computes the following hidden representation by using the aggregated messages and the current state of the node by

$$h_i = u\left(x_i, h_{N_i}\right) \tag{2}$$

Then, the well-log data are represented as a graph for the outlier removal method using GNN, where each sample is equivalent to a node, and the value of each sample is the node feature. The edge connects the nearest neighbor samples to a given sample, and the network learns their distance as the anomaly score. Therefore, the edge feature e_j is defined by

$$e_j = \begin{cases} d\left(x_i, x_k\right), & k \in N_i \\ 0, & k \notin N_i \end{cases} \tag{3}$$

where d is the Euclidean distance between two point samples, x_i is the source sample, x_k is the adjacent sample, and k is the nearest neighbor sample set. The distance information is the message transmitted from the source sample to the adjacent sample, $m = e_j$. In addition, the aggregation function concatenates the distance of all neighbors of the source sample by

$$a_i = \left[e_1, \ldots, e_k\right] \in \mathbb{R}^k \tag{4}$$

Next, Eq. 1 can be rewritten as a neural network $\mathcal{F}$, where a_i represents the hidden representation h_{N_i} through the learnable weights Θ by

$$h_{N_i} = \mathcal{F}\left(a_i, \Theta\right) \tag{5}$$

Then, the update function in Eq. 2 is rewritten as the learned aggregated message h_{N_i} by

$$u = h_{N_i} \tag{6}$$

The GNN performance is compared with the isolation forest (IF) (Liu et al., 2008) and the local outlier factor (LOF) (Breunig et al., 2000). IF is an unsupervised ensemble method to separate anomalies from normal data. Based on the principle that a normal sample requires more partitions to be isolated, an anomaly sample requires fewer partitions. Then, the IF constructs a tree representing the number of divisions to isolate a sample. Normal samples have a path length that equals the distance from the root node to the terminating node. Anomalous samples have a shorten path length than normal samples. On the other hand, LOF is an unsupervised proximity algorithm for anomaly detection that calculates the local density deviation of a sample within its neighbors. The local density is calculated by comparing the distance between the neighboring samples. Normal samples have similar densities to their neighbors, while the samples with less density are considered outliers.

Feature engineering with empirical mode decomposition

The EMD is an adaptative and data-driven decomposition method suitable for non-stationary and non-linear data (Huang et *al.*, 1998). In contrast with the wavelet transformation, a wavelet definition is unnecessary for EMD (Zhou Y. et al., 2019b). EMD calculates IMFs with several frequency bands highlighting distinct stratigraphical and geological information that increases the network performance (Xue et al., 2016). IMFs are computed using the CEEMDAN algorithm, reducing model mixing and data loss (Colominas et al., 2014). This computation involves four steps. First, several types of Gaussian white noise w are added to the original data x as follows,

$$x^i = x + \varepsilon_0 w^i \tag{7}$$

where x^i is the data after adding white noise for an i-th time, and i denotes the number of modes (i.e., $i = 1, \ldots, I$), and ε_k is the fixed coefficient that regulates the signal-to-noise ratio at each k stage. Second, the adjoined noise data x^i are decomposed using the EMD. The fundamental EMD mode IMF_1 is averaged, and the first CEEMDAN mode $\overline{IMF}_1$ is calculated by

$$\overline{IMF}_1 = \frac{1}{I}\sum_{i=1}^{I} IMF_1^i \tag{8}$$

The first residual is calculated by subtracting $\overline{IMF}_1$ from x,

$$r_1 = x - \overline{IMF}_1 \tag{9}$$

Third, the second CEEMDAN mode $\overline{IMF}_2$ is calculated, where E_k is the k-th mode decomposed by the EMD algorithm,

$$\overline{IMF}_2 = \frac{1}{I}\sum_{i=1}^{I} E_1\left(r_1 + E_1\varepsilon_1 w^i\right) \tag{10}$$

Fourth, the process is repeated until the residual is unable to be further decomposed,

$$r_k = x - \sum_{k=1}^{K} \overline{IMF}_k \tag{11}$$

Then, the final residual is calculated by

$$\overline{IMF}_k = \frac{1}{I}\sum_{i=1}^{I} E_1\left(r_{k-1} + E_{k-1}\varepsilon_{k-1} w^i\right) \tag{12}$$

and the representation of the original data are defined by

$$x = \sum_{k=1}^{K} \overline{IMF}_k + r_k \tag{13}$$

The IMF approach is compared with the depth gradient and the spectral band methods. The gradient measures the rate of change of a well-log property in depth to map subtle changes in the subsurface. The spectral method decomposes the well-logging data into frequency bands to capture unseen relationships. In Figure 1, the IMF engineering features f_i are concatenated with the node features x_i of each node v_i at each depth point z_i. The result is an augmented feature matrix $\hat{\boldsymbol{X}}$, which serves as the input sequences for the GCN-BiGRU. Additionally, the logarithmic transformation is applied to the resistivity log to center the distribution. And the input sequences are normalized with the Minmax function for stable training by

$$x_{scaled} = \frac{x - min(x)}{max(x) - min(x)} \tag{14}$$

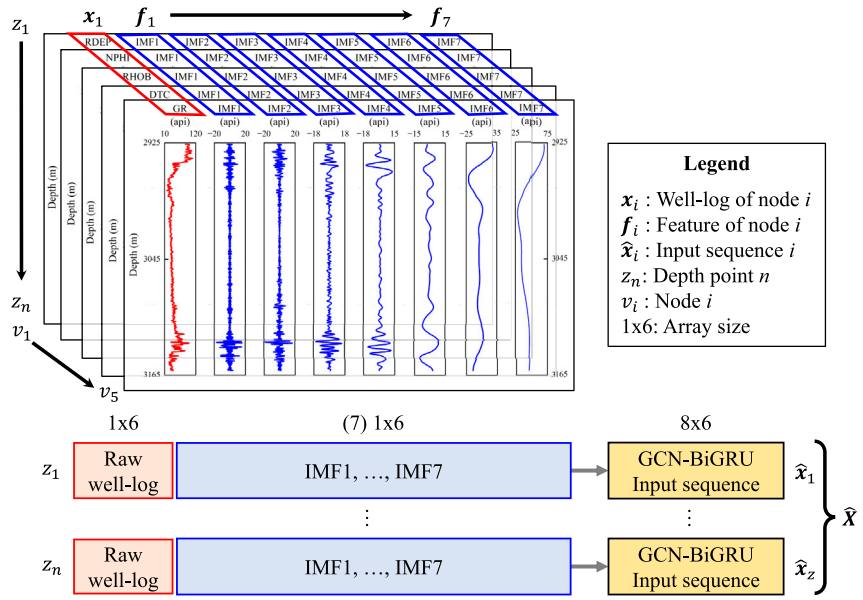


FIGURE 1
Sequence construction with the IMF features for the proposed model.

where x is the well-log data, $min(x)$ is the minimum value of the dataset, $max(x)$ is the maximum value of the dataset, and x_{scaled} is the normalized well-log data.

Graph construction

The S-wave velocity prediction is defined in the graph domain as follows. Given a certain number of training wells, expressed as an undirect graph $G(V, E)$, V are the well-logs, E are their complex dependencies, and $\boldsymbol{X}$ are the values of the well-log curves. The goal is to learn the intrinsic relationships and predict the S-wave velocity $\hat{y}_z$. The graph construction workflow calculates the edge weights and the graph convolution (Gconv) operator. Then, the node features are fed to the GCN-BiGRU network to output the S-wave velocity, as shown in Figure 2. Transforming well-log data into the graph domain is crucial since the Gconv operator requires reliable estimation of the graph interdependencies for an accurate prediction. Although there are no optimal methods to generate a graph from tabular data (Narayanan et al., 2017), we proposed a

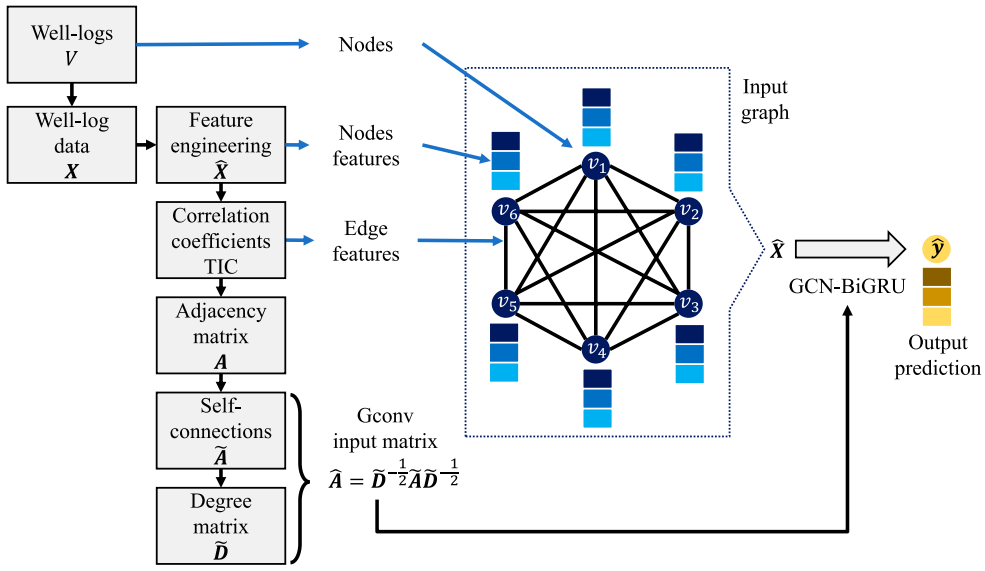


FIGURE 2
Graph construction workflow.

knowledge-based approach to aggregate information from the external domain. The edge features are calculated with the TIC to represent the complex intrinsic relationships between the physical rock properties measured by the well logs and highlight the most significant dependencies.

TIC is a variation of the maximal information coefficient (MIC) that integrates power, equitability, and performance to extract the potentially diverse relationships among well-log data (Reshef et al., 2018). MIC is a coefficient that detects non-linear dependencies by applying information theory and probability concepts and is robust to noise regardless of the relationship type (Reshef et al., 2011; Reshef et al., 2015). Mutual information (MI) is defined by the Kullback-Leibler divergence between two well logs joint and marginal distributions; the higher the variance, the higher the MI (Reshef et al., 2016). MIC is calculated by drawing a grid over a scatter plot to partition the data and embed the relationship. The well-log data distributions are discretized into bins, and the MI values are compared and divided by the theoretical maximum for a particular combination of bins. Then, MIC is defined as the highest normalized MI between two well-logs by

$$\mathrm{MIC}(x,y) = \frac{\max\left(I(x,y)\right)}{log_2\left(min\left(n_x,n_y\right)\right)} \tag{15}$$

where $I(x,y)$ is the MI between the well-logs x and y, n_x, n_y are the number of bins where x and y are partitioned. The MIC calculation becomes computationally expensive in large datasets. Therefore, the maximal grid size for simplification and optimization is defined by

$$n_x n_y < B(n) = n^6 \tag{16}$$

where n is the sample size. If $B(n)$ is significantly low, MIC searches only simple patterns, weakening the generality of MIC. If $B(n)$ is extremely high, MIC searches non-trivial coefficients for independent paired variables under finite samples. Therefore, MIC is redefined as

$$\mathrm{MIC}(x,y) = \max_{n_x n_y \le B(\mathrm{n})} \frac{\max\left(I(x,y)\right)}{log_2\left(min\left(n_x,n_y\right)\right)} \tag{17}$$

MIC measures equitability rather than the power to reject a null hypothesis of independence. Therefore, TIC is introduced to address this issue. Instead of choosing the maximal MI value, all entries are summed,

$$\mathrm{TIC}(\mathrm{x},\mathrm{y}) = \sum\nolimits_{n_x n_y \le B(\mathrm{n})} \frac{\max\left(I(\mathrm{x},\mathrm{y})\right)}{log_2\left(min\left(n_x,n_y\right)\right)} \tag{18}$$

Finally, the prediction performance of the graph embeddings using TIC and MIC are compared with other linear and non-linear correlation coefficients. The Pearson product-moment correlation coefficient (PC) (Szabo and Dobroka, 2017) quantifies linear relationships. And the Spearman rank correlation coefficient (SC) (Pilikos and Faul, 2019) and distance correlation coefficient (DC) (Skekely et al., 2007) measure non-linear relationships.

Network architecture

The GCN-BiGRU, GCN-GRU, and GCN network structures are shown in Figure 3. In Figure 3A, the input $\hat{\boldsymbol{X}} \in \mathbb{R}^{N\times Z}$ of the GCN-BiGRU is the feature matrix defined by well-logging data concatenated with the engineering features, where N is the number of well-log curves, and Z is the number of depth samples. The GCN-GRU predicts the spatial-temporal relationships in the forward and backward direction and transmits their final state $\boldsymbol{h}_z$ to the next GCN-GRU. The final output y_z is the predicted S-wave velocity at each depth point. In Figure 3B, the GCN-GRU consists of a reset gate $\boldsymbol{r}_z$, an update gate $\boldsymbol{u}_z$, a candidate memory $\boldsymbol{c}_z$, and a GCN to extract the most relevant information between depth points and output the state $\boldsymbol{h}_z$. In Figure 3C, the GCN concatenates the input node features with a hidden state, followed by the Gconv, an activation function, and a dropout layer. The Gconv captures the spatial relationships between the well-logs and hidden states within a first-order neighborhood radius. The GCN extracts spatial dependencies among nodes at each depth point, and the GCN-GRU extracts depth dependencies along depth points.

The Gconv uses the adjacency matrix $\boldsymbol{A}$, degree matrix $\boldsymbol{D}$, and feature matrix $\boldsymbol{X}$ to construct a normalized spectral filter in the Fourier domain (Kipf and Welling, 2017). The adjacency matrix $\boldsymbol{A} \in \mathbb{R}^{N\times N}$ describes the edge weights between N well-logs, defined by the calculated correlation coefficient. The diagonal matrix $\boldsymbol{D} \in \mathbb{R}^{N\times N}$ describes the number of edges at each node, computed from $\boldsymbol{A}$. Then, a single-layer Gconv operator is defined by

$$\mathrm{Gconv}\left(\hat{\mathbf{X}},\hat{\mathbf{A}}\right) = f_{drop}\left(\sigma\left(\hat{\mathbf{A}}\hat{\mathbf{X}}\mathbf{W}\right)\right) \tag{19}$$

where $\hat{\boldsymbol{A}}$ is the normalized self-connected adjacency matrix defined as $\hat{\boldsymbol{A}} = \tilde{\boldsymbol{D}}^{-\frac{1}{2}}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{D}}^{-\frac{1}{2}}$, $\tilde{\boldsymbol{A}}$ denotes the adjacency matrix with self-connections, defined as $\tilde{\boldsymbol{A}} = \boldsymbol{A} + \boldsymbol{I}$, where $\boldsymbol{I}$ is the identity matrix, $\tilde{\boldsymbol{D}}$ is the degree matrix of the adjacency matrix with self-connections $\tilde{\boldsymbol{A}}$, defined as $\tilde{\boldsymbol{D}} = \sum_j \tilde{\boldsymbol{A}}_{ij}$ where i is the number of nodes, j is the number of edges, $\boldsymbol{W}$ are the trainable weights, whose size is determined by the number of hidden units, $\sigma(\cdot)$ is the Mish activation function for non-linearity, and f_{drop} is a dropout layer with a given probability, activated during the training phase, to reduce overfitting. Mish is a novel self-regularized non-monotonic activation function that surpasses ReLU and Swish performances (Misra, 2019). Mish is defined by

$$\mathrm{Mish} = x\tanh\left(\ln\left(1+e^x\right)\right) \tag{20}$$

The GCN-BiGRU network comprises two GCN-GRUs for a forward and backward prediction. Each GCN-GRU has a two-gated mechanism to adaptively capture patterns from different depth points (Cho et al., 2014), as shown in Figure 3B. The activation gates are the reset gate $\boldsymbol{r}_z$ and the updated gate $\boldsymbol{u}_z$. The GCN-GRU requires two inputs, the feature matrix $\hat{\boldsymbol{x}}_z$ at depth z and the previous hidden cell state $\boldsymbol{h}_{z-1}$. First, the reset gate $\boldsymbol{r}_z$ controls the amount of information to preserve from the previous depth point to transmit to the memory candidate state $\boldsymbol{c}_z$. The reset gate $\boldsymbol{r}_z$ combines the previous memory information $\boldsymbol{h}_{z-1}$ with the current information $\hat{\boldsymbol{x}}_z$ by

$$\boldsymbol{r}_z = \sigma\left(\boldsymbol{W}_r\mathrm{Gconv}\left(\left[\hat{\boldsymbol{x}}_z,\boldsymbol{h}_{z-1}\right],\boldsymbol{A}\right) + \boldsymbol{b}_r\right) \tag{21}$$

where $\boldsymbol{W}_r$ and $\boldsymbol{b}_r$ are the trainable parameters of the reset gate, $\hat{\boldsymbol{x}}_z$ is the current state input, $\boldsymbol{h}_{z-1}$ is the hidden state from the previous depth point, $[\cdot]$ represents the concatenation, $\sigma(\cdot)$ is the logistic sigmoid function that forces the output range between [0,1], $\boldsymbol{r}_z$ output is a scalar, $\boldsymbol{r}_z \in [0,1]$, when $\boldsymbol{r}_z = 1$ the memory is preserved, when $\boldsymbol{r}_z = 0$, the memory is discarded. Second, the update gate $\boldsymbol{u}_z$

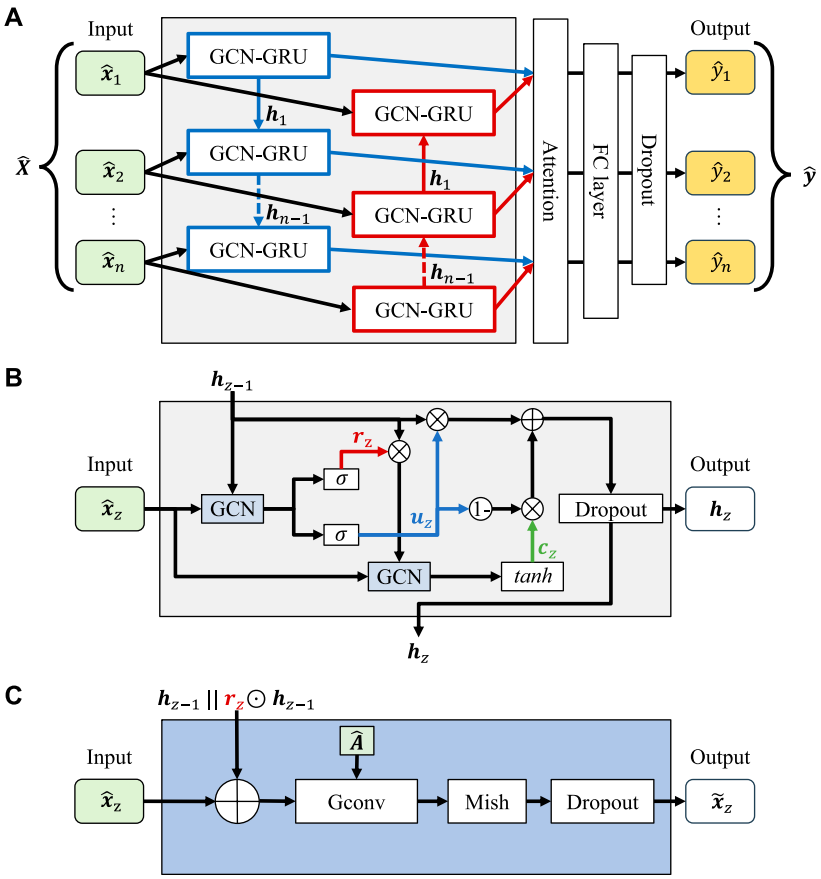


FIGURE 3
(A) GCN-BiGRU structure. **(B)** GCN-GRU structure. **(C)** GCN structure.

determines the amount of information to preserve from the previous depth point and the amount of current information to include from the current depth point, similar to the reset gate,

$$\boldsymbol{u}_z = \sigma(\boldsymbol{W}_u \mathrm{Gconv}([\hat{\boldsymbol{x}}_z, \boldsymbol{h}_{z-1}], \boldsymbol{A}) + \boldsymbol{b}_u) \tag{22}$$

where $\boldsymbol{W}_u$, and $\boldsymbol{b}_u$ are the trainable parameters of the update gate. Third, the memory candidate $\boldsymbol{c}_z$ is the present moment state at the depth point z and is defined by

$$\boldsymbol{c}_z = \tanh(\boldsymbol{W}_c \mathrm{Gconv}([\hat{\boldsymbol{x}}_z, (\boldsymbol{r}_z \odot \boldsymbol{h}_{z-1})], \boldsymbol{A}) + \boldsymbol{b}_c) \tag{23}$$

where $\boldsymbol{W}_c$, $\boldsymbol{b}_c$ are the trainable parameters of the candidate memory, $\odot$ is the Hadamard product (*i.e.*, element-wise multiplication), and tanh is the hyperbolic tangent function. Finally, the output state $\boldsymbol{h}_z$ at depth z is defined by

$$\boldsymbol{h}_z = (1 - \boldsymbol{u}_z) \odot \boldsymbol{c}_z + \boldsymbol{u}_z \odot \boldsymbol{h}_{z-1} \tag{24}$$

The update gate selectively stores or forgets memory information. The update gate acts as a forget gate when $\boldsymbol{u}_z \odot \boldsymbol{h}_{z-1}$ ignores unimportant information from the previous depth point. The update gate acts as a memory gate when $(1 - \boldsymbol{u}_z) \odot \boldsymbol{c}_z$ preserves relevant information in memory for the next depth point. Additionally, a dropout layer is added at the end of each GCN-GRU to increase the network generalization ability and reduce overfitting. Next, the output state $\boldsymbol{h}_z$ is fed into a hierarchical attention mechanism to highlight the essential features and attenuate the less significative information contributing to the S-wave velocity prediction (Bahdanau et al., 2014; Yang et al., 2016). The output state $\boldsymbol{h}_z$ is fed into a fully connected (FC) layer, followed by an activation function to obtain a hidden representation $\boldsymbol{u}_z$ defined by

$$\boldsymbol{u}_z = \tanh(\boldsymbol{W}_a \boldsymbol{h}_z + \boldsymbol{b}_a) \tag{25}$$

where $\boldsymbol{W}_a$ and $\boldsymbol{b}_a$ are the trainable parameters of the FC layer. Next, the feature importance is measured between the hidden representation $\boldsymbol{u}_z$ and a trainable context vector $\boldsymbol{u}_w$ that is a high-level representation of a static query of the features (Sukhbaatar et al., 2015; Kumar et al., 2016). Then, importance weights are normalized through a softmax function $\boldsymbol{\alpha}_z$ by

$$\boldsymbol{\alpha}_z = \frac{\exp(\boldsymbol{u}_z \boldsymbol{u}_w)}{\sum_z \exp(\boldsymbol{u}_z \boldsymbol{u}_w)} \tag{26}$$

Then, the weighted sum of the hidden states is the new high-level presentation of the output state $\hat{\boldsymbol{h}}_z$ and is defined by

$$\hat{\boldsymbol{h}}_z = \sum_z \boldsymbol{\alpha}_z \boldsymbol{h}_z \tag{27}$$

and the new output state $\hat{\boldsymbol{h}}_z$ is fed to a fully connected (FC) layer with a Mish activation function and a dropout layer to predict the S-wave velocity by

$$\hat{y}_z = f_{drop}\left(\sigma\left(\boldsymbol{W}_f \hat{\boldsymbol{h}}_z + \boldsymbol{b}_f\right)\right) \tag{28}$$

where $\boldsymbol{W}_f$ and $\boldsymbol{b}_f$ are the trainable parameters of the final FC layer. Lastly, the training dataset is rearranged into sequences in the supervised training process and matched with the labels. The Huber loss function is implemented to minimize the difference between the predicted S-wave velocity $\hat{y}_z$ and the actual S-wave velocity y_z at depth z. The Huber loss is less sensitive to outliers and noise by combining L1 and L2 norms (Yu et al., 2016) and is defined by

$$\mathcal{L} = \begin{cases} \frac{1}{2}\left(y_z - \hat{y}_z\right)^2, & \left|y_z - \hat{y}_z\right| \le \delta \\ \delta\left(\left|y_z - \hat{y}_z\right| - \frac{1}{2}\delta\right), & \left|y_z - \hat{y}_z\right| > \delta \end{cases} \tag{29}$$

where $\delta = 0.1$ is the threshold parameter for the L1/L2 norm. The adaptive movement estimation algorithm (Adam) is used for network optimization with a learning rate of 0.001 (Kingma and Ba, 2015). The grid search strategy is applied to optimize the network parameters selection and a 10-fold cross-validation method to evaluate the accuracy and generalization ability of the model while reducing randomness impact (Hampson et al., 2001). Finally, the metrics to evaluate the difference between the actual S-wave velocity and the predicted S-wave velocity are RMSE, MAE, MAPE, and R^2. RMSE measures the average weighted performance of the model. MAE estimates the average error of the model for the prediction. MAPE measures the percentage of the average difference between the actual and the predicted value. The coefficient of determination measures the performance of the model over a regressor that outputs the mean value of the label that is used in training. The error metrics are defined as follows,

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{z=1}^{N}\left(y_z - \hat{y}_z\right)^2} \tag{30}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{z=1}^{N}\left|y_z - \hat{y}_z\right| \tag{31}$$

$$\mathrm{MAPE} = \frac{1}{N}\sum_{z=1}^{N}\frac{\left|y_z - \hat{y}_z\right|}{\left|y_z\right|} \times 100 \tag{32}$$

$$\mathrm{R}^2 = 1 - \frac{\sum_{z=1}^{N}\left(y_z - \hat{y}_z\right)^2}{\sum_{z=1}^{N}\left(y_z - \bar{y}\right)^2} \tag{33}$$

where $\bar{y}$ is the mean of the actual S-wave velocity, and N is the number of samples.

Field data example

The study comprises a selection of 30 wells from the North Sea area (Bormann et al., 2020). The training dataset consists of 21 wells, the validation dataset includes 5 wells, and the testing dataset consists of 4 blind wells. Each well has 6 well-log curves: Gamma-ray (GR), compressional wave transit-time (DTC), shear wave transit-time (DTS), bulk density (RHOB), neutron porosity (NPHI), and deep resistivity (RDEP). The original sampling interval is 0.152 m, and the range of the training dataset is constrained for stability purposes, as shown in Table 1. The GCN-BiGRU employs a prediction window of 1 sample, a sequence length of 8 samples, and 8 hidden units for the dimension of the hidden state. The training time is 20 min for 50 epochs and a batch size of 128 samples in an Nvidia GeForce GTX 960M.

TABLE 1 Well-log data range for the training dataset.

Well-logs	Data range
GR	0–200 api
DTC	50–200 us/ft
DTS	80–500 us/ft
RHOB	1.8–3.0 g/cc
NPHI	0–1 dec
RDEP	0–20 Ωm

Additionally, the robustness of the proposed method is evaluated by measuring the impact of the proportion of the training dataset and the sensitivity to Gaussian noise. First, the training dataset is divided into ten groups based on the number of wells (i.e., 4, 7, 9, 10, 12, 15, 16, 17, 20, and 21), corresponding to a ratio of 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1 of the training dataset, respectively. Second, the noise resistance is analyzed by adding Gaussian noise with mean zero and standard deviation of the training dataset (i.e., σ_{GR} = 40.96 api, σ_{DTC} = 20.09 us/ft, σ_{DTS} = 79.20 us/ft, σ_{RHOB} = 0.17 g/cc, σ_{NPHI} = 0.10 dec, σ_{RDEP} = 21.07 Ωm) to each sample. Then, the performance is evaluated by examining ten fraction levels of the defined noise (i.e., 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1). Finally, the RMSE on the validation and testing datasets is calculated for both analyses.

The effects of the GNN, LOF, and IF methods are shown in Figure 4. GNN uses 13 samples as neighbors, LOF 50 nodes as neighbors, and IF 100 estimators. Additionally, a contamination value of 0.1 is employed for the three methods. The GNN handles the spikes located on the RHOB log below the 2,200 m better than the alternative outlier removal methods and other abrupt values on the rest of the well-logs below the 2,400 m while preserving the essential well-log information, as shown in Figure 4. In the prediction performance, the GNN surpasses LOF and IF methods, with lower RMSE error for the predicted DTS log, as shown in Figure 5. The RMSE for the training, validation and testing dataset with the GNN are 21.0482 us/ft, 22.7562 us/ft, and 23.5854 us/ft, respectively. Compared with the LOF and IF methods, the main drawback of GNN is the higher computation time.

The cross-plot between the DTS and the well-log curves is shown in Figure 6. The color represents the distribution density of the samples. The higher the density, the higher the color intensity. And the line represents the minimum squares regression line. The RHOB and NPHI show a good linear trend, the DTC behaves linearly for low values, and the relationship changes for higher values. The DEPT and RDEP trend is logarithmic, while the GR is unclear due to the bimodal distribution between sand and shales. The cross-plot shows that a linear correlation coefficient is insufficient to capture the intrinsic relationships of the rock properties and to build a meaningful graph structure. Therefore, a non-linear correlation coefficient is more suitable for this task.

The relationship strength between the DTS log and the other well logs curves with the six correlation coefficients is shown in

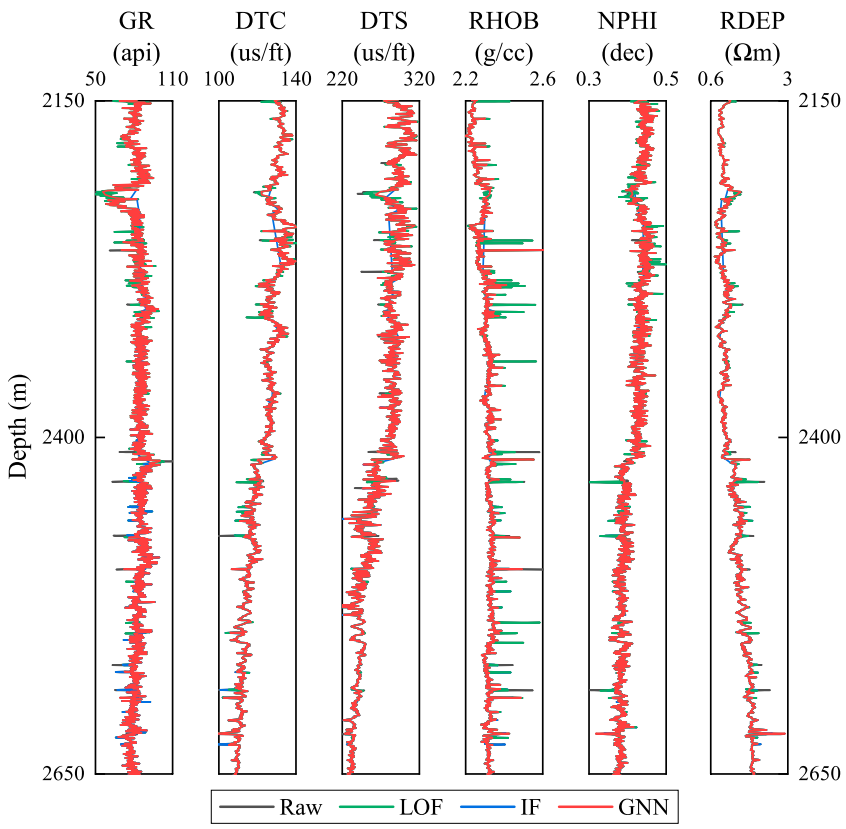


FIGURE 4
Comparison of the three methods for outlier removal at well W3.

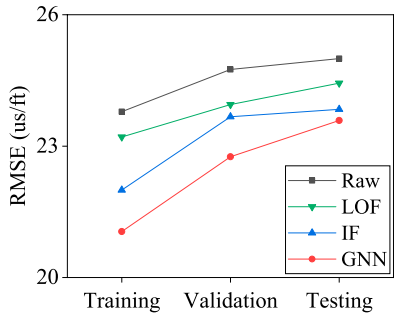


FIGURE 5
RMSE of predicted DTS for the three outlier removal methods.

Figure 7. The hexagons range from 0 to 1, with an increment of 0.2. The closer to the center, the lower the correlation; the closer to the edges, the higher the correlation. On average, the DTC, NPHI, and RHOB logs show a high correlation, consistent with the definition of S-wave velocity. The DTC correlation is higher because it shares the shear modulus and density parameters. The density is a very sensitive parameter for rock velocity, and the porosity directly impacts the rigidity of the rock and reduces its value. The DEPT shows a moderate correlation due to the dependency on changes in pressure and temperature that affect the rock properties. RDEP has an average correlation linked to the lithology characteristics of the layer. In contrast, the low correlation in GR is probably due to averaging effect between sand and shale lithologies. These results constitute the building block to constructing a graph with meaningful physical rock relationships, proven by external knowledge.

The evaluation of the prediction results for the six correlation coefficients is shown in Figure 8. TIC accuracy is higher than other approaches, with an RMSE value of 22.1603 us/ft, 23.3468 us/ft, and 24.2019 us/ft for the training, validation, and testing datasets, respectively. TIC is more reliable for embedding the non-linear physical correlation between the rock properties and the well-logs into the graph edges. However, MI approaches have a high computational cost for extensive datasets than other correlation coefficients. The TIC matrix used as the adjacency matrix to represent the edge features in the proposed method is shown in Figure 9. The DTC and DTS are the only pair that achieves a high correlation, with a value of 0.57, which is consistent with the theoretical and empirical results for S-wave velocity prediction.

The evaluation of the three feature engineering methods for the DTS log prediction is shown in Figure 10. The gradient method adds the first derivative as a component. The frequency band method uses three components. The low-frequency band (i.e., 20 Hz) isolates the significant geological trend changes. The middle-frequency band (i.e., 40 Hz) is related to third-order sequence events, while the high-

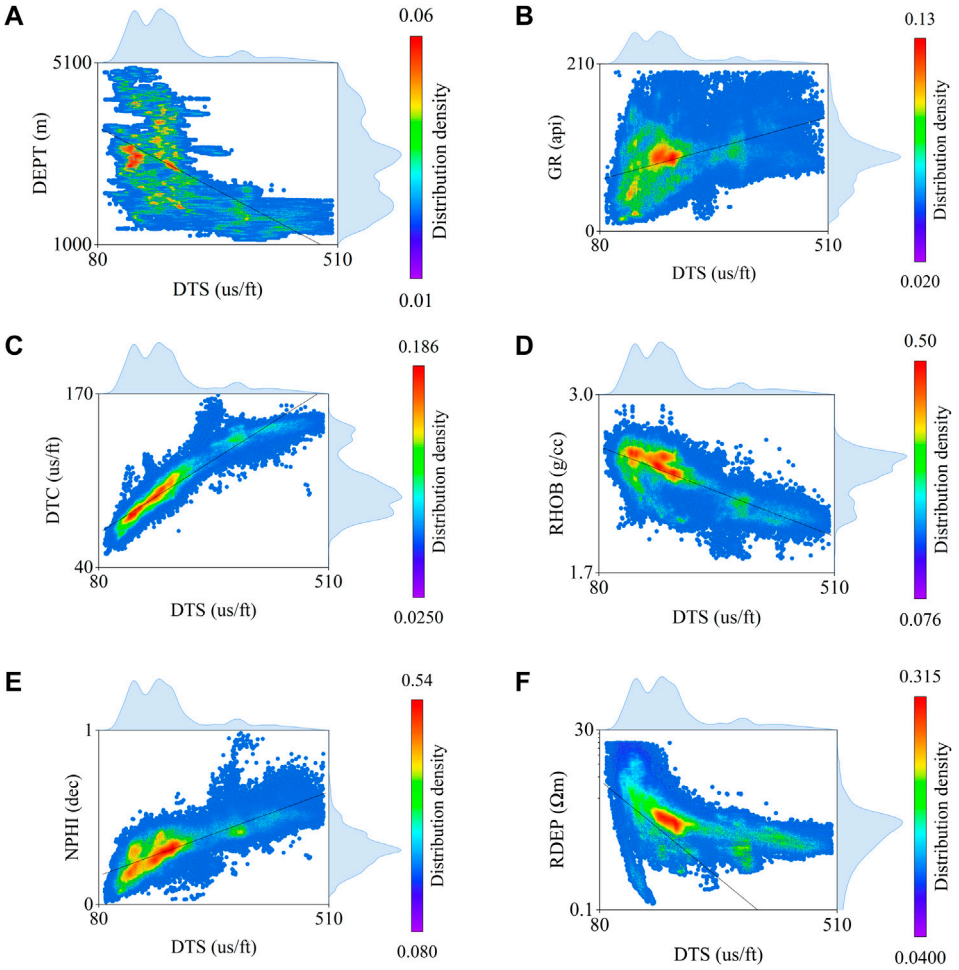


FIGURE 6
Cross plots between DTS and the well-logs. **(A)** DEPT. **(B)** GR. **(C)** DTC. **(D)** RHOB. **(E)** NPHI. **(F)**. RDEP.

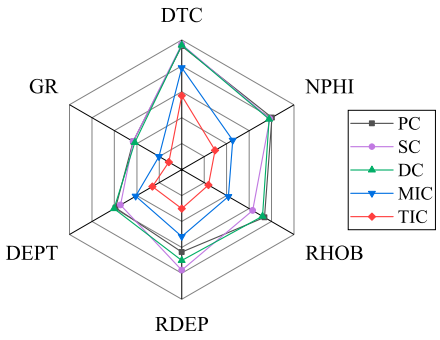


FIGURE 7
The correlation coefficients between DTS and the others well-log curves.

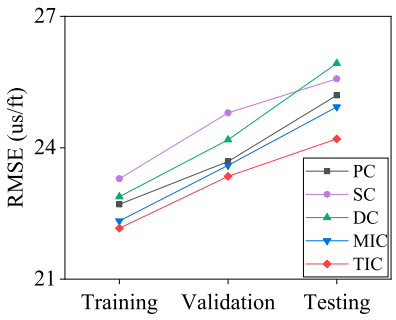


FIGURE 8
RMSE of predicted DTS of the six correlation coefficients.

frequency band (i.e., 200 Hz) focuses on the local changes inside the geological formations. After experimentation with the training dataset, the CEEMDAN method decomposes the data into 7 IMFs. This number preserves the uniformity size in all the IMFs for the sequence aggregation step and reduces overfitting by avoiding high-order IMFs without a reliable meaning. Results show that the feature engineering method can improve the prediction accuracy of the GCN-BiGRU network. The gradient

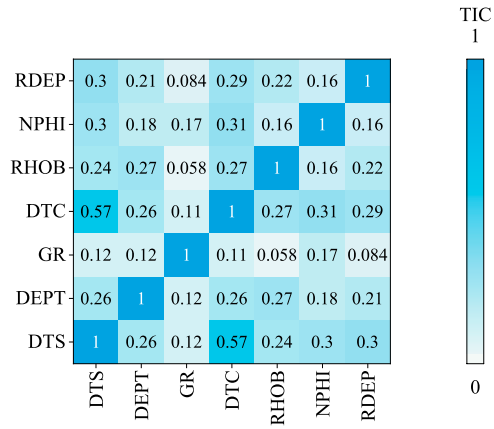


FIGURE 9
TIC for the training dataset.

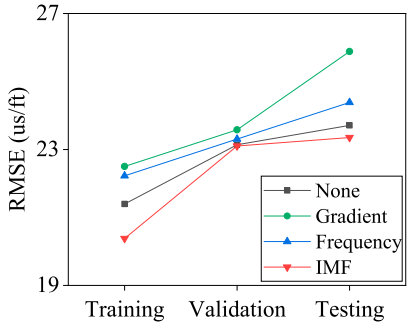


FIGURE 10
RMSE of predicted DTS for the three feature engineering methods.

shows the lowest performance because the contribution of its high frequency is less significant for the prediction. Although the computational time for the IMFs is longer than other methods, the RMSE is lower, with 20.3805 us/ft, 23.1001 us/ft, and 23.3531 us/ft, for the training, validation, and testing datasets, respectively.

The results for the DTS log prediction during the network optimization are shown in Figure 11. The RMSE of the GCN-BiGRU network is 19.6581 us/ft, 23.5363 us/ft, and 24.3045 us/ft for the training, validation, and testing datasets, respectively, improving the performance compared with the original GCN-GRU network (Zhao et al., 2020), as shown in Figure 11A. The Mish activation function shows superior regularization and overfitting reduction abilities than other state-of-the-art activation functions, such as Leaky ReLU, GELU, SELU, and Swish. The RMSE with the Mish activation function is 21.3972 us/ft, 23.1146 us/ft, and 23.6318 us/ft for the testing, validation, and testing datasets, respectively, as shown in Figure 11B.

The prediction performance by the number of well-logs is shown in Figure 11C. The node configurations are tested based on their coefficient ranking. Thus, the GR log is excluded. The node configurations are defined as follows: The 3 nodes include the DTC, NPHI, and RHOB logs. The 4 nodes include the DTC, NPHI, RHOB, and RDEP logs. The 5 nodes include the DTC, NPHI, RHOB, RDEP, and DEPT logs. The 6 nodes include all logs. Although the RMSE error decreases with 5 nodes for the training and validation datasets, the overall performance of the GCN-BiGRU decreases for the testing dataset. The RMSE for 6 nodes is 20.6117 us/ft, 22.8539 us/ft, and 22.9764 us/ft for the training, validation, and testing datasets, respectively. The 6 nodes are used since the GCN extracts meaningful embeddings based on the number of adjacent nodes for aggregation. When the number of nodes is reduced, the GCN embeddings are shallower, and the ability to map complex physical relationships among the input data is also reduced. Then, the prediction is compared with two attention mechanisms. The hierarchical attention shows a lower RMSE than soft attention, with a value of 19.7153 us/ft, 22.9858 us/ft, and 23.1156 us/ft, for the training, validation, and testing datasets, respectively, as shown in Figure 11D. However, the attention mechanism occasionally creates spike artifacts.

The impact of the proportion of the training dataset ratio is shown in Figure 11E. The RMSE is higher for a ratio of 0.1, with a value of 35.8539 us/ft for the validation dataset and 36.8711 us/ft for the testing dataset. The RMSE reduces between a ratio of 0.2–0.5, reaching a value of 32.4315 us/ft for the validation dataset and 32.7639 us/ft for the testing dataset at a ratio of 0.5. The RMSE shows a stability plateau between a ratio of 0.7–1, achieving a value of 22.2465 us/ft for the validation dataset and 22.9672 us/ft for the testing dataset at a ratio of 1.

The prediction performance in the presence of Gaussian noise is shown in Figure 11F. The RMSE is high when the added noise is equal to the standard deviation of the training dataset (i.e., a noise fraction of 1) with a value of 28.6670 us/ft for the validation dataset and 42.8739 us/ft for the testing dataset. The RMSE gradually decreases until a noise fraction of 0.5 with a value of 24.9382 us/ft for the validation dataset and 28.5342 us/ft for the testing dataset. The RMSE is stable when the noise fraction is lower than 0.2, with a value of 22.9373 us/ft for the validation dataset and 23.5910 us/ft for a noise fraction of 0.1.

Finally, the DTS log prediction results for all the models are shown in Figure 12. The GCN-BiGRU shows lower error in the training, validation, and testing dataset with an RMSE of 19.3260 us/ft, 22.4905 us/ft, and 22.7120 us/ft, respectively. The evaluation for the testing dataset is shown in Table 2. The GCN-BiGRU shows an MAE of 17.2842 us/ft, MAPE of 6.7880%, and R^2 of 0.9470. The GCN-BiGRU performs better than other ML baseline models and empirical equations without adding mineral components, fluid properties, pore aspect ratio, or thermal maturity information. Some discrepancies in the predicted DTS log and the actual DTS log value are due to the presence of fluids, unbalanced lithologies samples, or the inherent covariance shift problem.

The results for the testing well B9 are shown in Figure 13. The predicted DTS log is consistent with the actual DTS log, as shown in Figure 13A. The model performs satisfactorily when constant or missing values are present, such as the depths 2,900 m, 3,100 m, and 3,400 m. The distribution of the predicted DTS and the true DTS are consistent, as shown in Figure 13B. The range of the predicted DTS for higher values is reduced due to the constraints established during the training phase. The R^2 coefficient between the true and predicted DTS is 0.9593, as shown in Figure 163. The high coefficient indicates that the proposed model can explain a significant variation in the actual DTS log. Moreover, the homoscedasticity analysis shows that the variance of the residuals is homogeneous, thus increasing the robustness and feasibility of the method, as shown in Figure 13D.

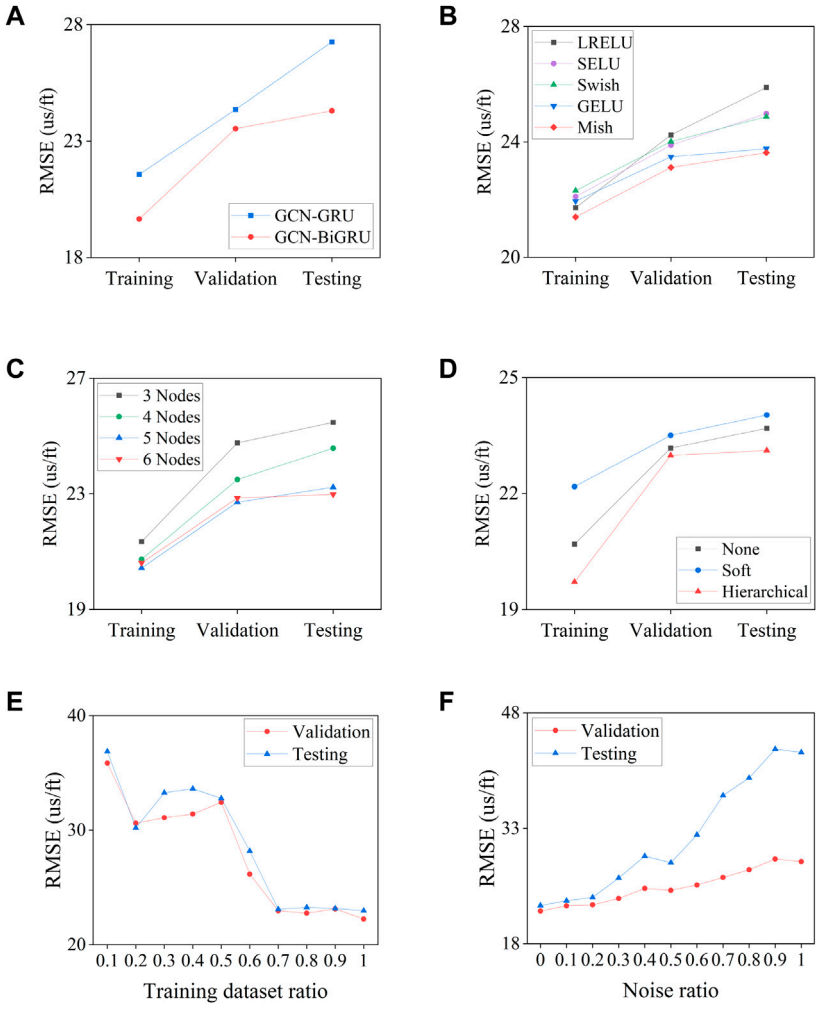


FIGURE 11
Proposed method optimization. **(A)**. Network structure. **(B)** Activation Function, **(C)** Nodes number, **(D)** Attention mechanism, **(E)** Training ratio test, **(F)** Noise sensitivity test.

Discussion

The proposed GCN-BiGRU method predicts the S-wave velocity by extracting the spatial and depth relationships among well-log data. The model combines GCN into GRU to create a GCN-GRU network, which is implemented to predict the S-wave velocity in two directions, forming the GCN-BiGRU network. The performance of the method is evaluated with a training dataset ratio test and a noise sensitivity test. The GCN-BiGRU has a lower error than Castagna's equation, LR, SVR, LSTM, GRU, and BiGRU baseline methods using the well-logs from the North Sea area. The approach is feasible and could be further extended for reservoir properties prediction using inverted seismic data as input and output maps and volumes of rock properties.

The GCN embeds the topological information, the intrinsic relationships, and the measured physical properties of the geological formations by an external-knowledge approach. The Gconv aggregates nearby information from the nodes, resembling a spectral Fourier filter. The number of nodes in the graph impacts the quality of the embeddings. Fewer nodes create shallow embeddings that reduce the representation ability.

Although 1-layer GCN is adopted due to the current graph topology, the GCN can extract deeper patterns from the well-log data with multiple GCN layers (Magner et al., 2022). Further research could reframe the graph creation process and add more

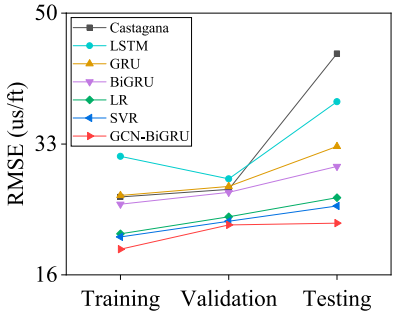


FIGURE 12
RMSE of the predicted DTS for all the compared methods.

TABLE 2 DTS prediction results for the testing dataset.

Model	RMSE (us/ft)	MAE (us/ft)	MAPE (%)	R^2
Castagna	44.7420	32.8800	12.4500	0.9374
LR	26.0257	21.1107	10.4064	0.9315
SVR	24.9529	19.9018	8.9553	0.9403
LSTM	38.4902	32.6003	14.7848	0.9079
GRU	32.6786	26.6913	11.2101	0.9201
BiGRU	30.0930	24.8923	11.0732	0.9313
GCN-BiGRU	22.7120	17.2842	6.7880	0.9470

hierarchy nodes (*i.e.*, nodes connected below the first-level nodes) for a meaningful aggregation during the graph embeddings.

The GCN-GRU extracts patterns over previous data windows to map the changes in rock properties with depth. The number of hidden units inside the GCN-GRU impacts the ability to memorize the most important information for the S-wave velocity prediction. Moreover, the dimension of the hidden states balances the generalization and overfitting of the GCN-GRU.

GNNs are a versatile approach to solving problems by the intrinsic message-passing characteristic. As an unsupervised outlier removal method, GNN shows promising results in handling anomalous values based on the sample distance between neighbor samples. GNN adapts to particular datasets by fine-tuning the number of nearest neighbor samples, which is essential for the detection performance. GNN for local outlier removal increases the accuracy of the model at the expense of a higher computational cost than IF and LOF.

The feature engineering process improves the prediction ability of the GCN-BiGRU. The prediction error is reduced with the IMFs. However, the network complexity and the training time increase with a higher number of features. The frequency bands are an alternative trade-off between accuracy and efficiency.

The construction of the graph is an essential step for the success of graph embedding. The proposed approach constructs the adjacency matrix from the correlation coefficients among well-log data. This supervised external-knowledge approach links the relationships between the measured rock properties and the wave propagation parameters into the network. Linear coefficients have limitations for capturing intrinsic rock dependencies and are more sensitive to their variation with depth. Non-linear coefficients extract suitable representations of the complex relationships between rocks and measured physical properties and are more

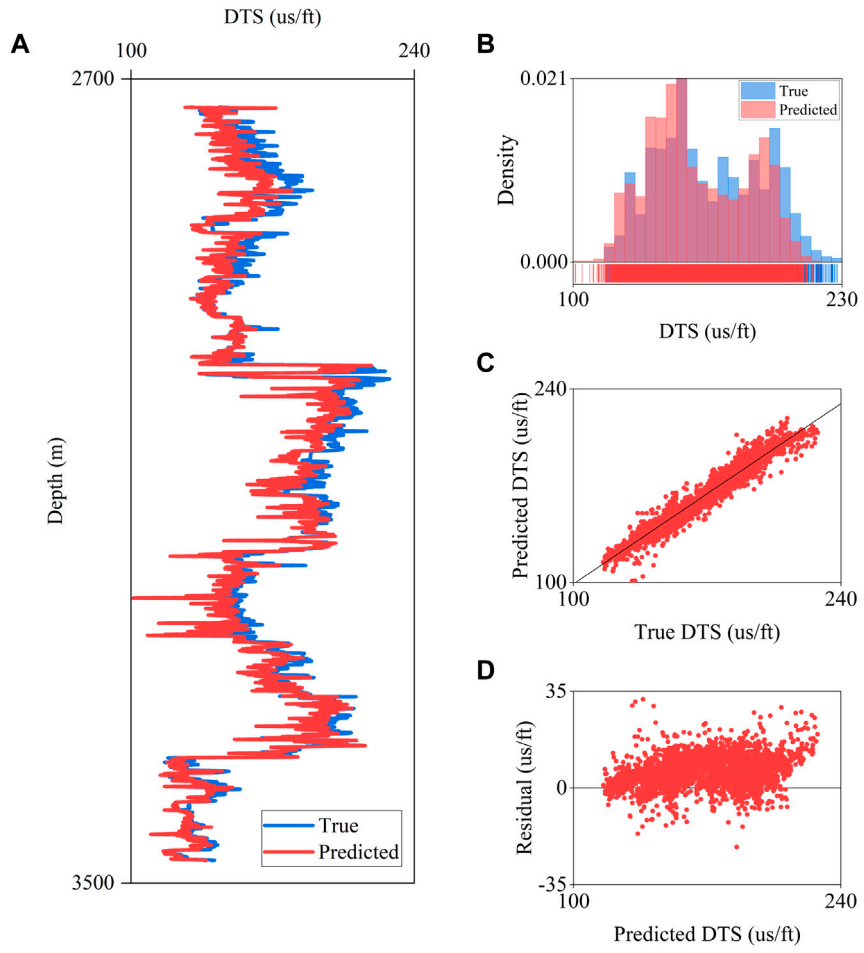


FIGURE 13
DTS prediction results on testing well B9. **(A)** DTS log curve, **(B)** True and predicted DTS distribution, **(C)** R^2 analysis, **(D)** Homoscedasticity analysis.

robust to well-log data variance, preserving the intrinsic dependencies that govern depth.

Depth changes are affected by temperature, pressure, fluid, and lithology, among other factors. Difficulties arise with a fixed adjacency matrix in complex geological scenarios by approximating the global properties variation with depth. Specifically, the GCN has limitations for predicting local minima and maxima due to the smooth moving average filter in the Fourier domain. Therefore, further research towards a dynamic graph representation to recreate more realistic models and map depth-dependent representations is encouraged.

The GCN-BiGRU uses point-wise activation functions as a non-linear operator. Nevertheless, further research is required to adapt non-linearities directly into the graph domain and increase the generalization of the model. The contribution of conventional attention mechanisms for the S-wave velocity prediction should be further explored. Graph attention networks or graph transformers have the potential to improve the ability of the network in abrupt lithology changes.

Conclusion

This study introduces a novel method for predicting S-wave velocity with a GCN-BiGRU network. GCN captures the spatial dependencies from the well-log data, while bidirectional GCN-GRU maps the changes in the rock properties with depth in both upward and backward directions. The well-log data are transformed into the graph domain by integrating external knowledge into the model. The well-logs are the graph nodes, well-logging data are the node features, and their intrinsic non-linear relationships are the edges features defined by TIC. Moreover, an unsupervised GNN is implemented for outlier removal to increase the network performance. And IMFs are aggregated to the node features, improving the prediction accuracy. The proposed method performs better than LR, SVR, LSTM, GRU, BiGRU methods, and Castagna's empirical equation. Finally, this method shows promising applications for predicting reservoir properties using inverted seismic data.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author. The North Sea dataset can be downloaded through the dGB Earth Sciences Open Seismic Repository https://terranubis.com/datainfo/FORCE-ML-Competition-2020-Synthetic-Models-and-Wells.

Author contributions

DC and YL contributed to the idea and methodology. DC was responsible for the field data training and testing and the writing of the manuscript, and YL checked and polished the manuscript. All authors have read the manuscript and agreed to publish it.

Funding

National Natural Science Foundation of China (NSFC) under contract no. 42274147.

Acknowledgments

We want to thank the Editor, Zhenming Peng, and the two external peer reviewers for their insightful comments and suggestions, which helped improve the quality of our manuscript. We also thank Gui Cheng for his helpful discussion and valuable recommendations. We are particularly grateful to China's National Natural Science Foundation for its founding and support.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Ali, M., Jiang, R., Ma, H., Pan, H., Abbas, K., Ashraf, U., et al. (2021). Machine learning – a novel approach of well logs similarity based on synchronization measures to predict shear sonic logs. *J. Pet. Sci. Eng.* 203, 108602. doi:10.1016/j.petrol.2021.108602

Amarbayasgalan, T., Jargalsaikhan, B., and Ryu, K. H. (2018). Unsupervised novelty detection using deep autoencoders with density based clustering. *Appl. Sci.* 8 (9), 1468. doi:10.3390/app8091468

Anemangely, M., Ramezanzadeh, A., Amiri, H., and Hoseinpour, S. (2019). Machine learning technique for the prediction of shear wave velocity using petrophysical logs. *J. Pet. Sci. Eng.* 174, 306–327. doi:10.1016/j.petrol.2018.11.032

Bahdanau, D., Cho, K., and Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. ArXiv abs/1409.0473.

Bormann, P., Aursand, P., Dilib, F., Dischington, P., and Manral, S. (2020). FORCE Mach. Learn contest. Avaliable At: https://github.com/bolgebrygg/Force-2020-Machine-Learning-competition.

Breunig, M. M., Kriegel, H. P., Ng, R. T., and Sander, J. (2000). Lof: Identifying density-based local outliers. *ACM SIGMOD Rec.* 29 (2), 93–104. doi:10.1145/342009.335388

Castagna, J. P., Batzle, M. L., and Eastwood, R. L. (1985). Relationships between compressional-wave and shear-wave velocities in clastic silicate rocks. *Geophysics* 50 (4), 571–581. doi:10.1190/1.1441933

Chen, W., Yang, L., Zha, B., Zhang, M., and Chen, Y. (2020). Deep learning reservoir porosity prediction based on multilayer long short-term memory network. *Geophysics* 85 (4), WA213–WA225. doi:10.1190/geo2019-0261.1

Cho, K., Merrienboer, B., Bahdanau, D., and Bengio, Y. (2014). "On the properties of neural machine translation: Encoder-decoder approaches," in *Proc. Syntax semant. Struct. Stat. Transl* (Doha, Qatar: Association for Computational Linguistics), 103–111. doi:10.3115/v1/W14-4012

Colominas, M. A., Schlotthauer, G., and Torres, M. E. (2014). Improved complete ensemble EMD: A suitable tool for biomedical signal processing. *Biomed. Signal Process. Control.* 14 (1), 19–29. doi:10.1016/j.bspc.2014.06.009

Defferrard, M., Bresson, X., and Vandergheynst, P. (2016). "Convolutional neural networks on graphs with fast localized spectral filtering," in Proc. Int. Conf. Neural inf. Process. Syst, Barcelona, Spain, December 5–10, 2016 (Neural Information Processing Systems (NIPS)), 3844–3852. doi:10.5555/3157382.3157527

Di Massa, V., Monfardini, G., Sarti, L., Scarselli, F., Maggini, M., and Gori, M. (2006). "A comparison between recursive neural networks and graph neural networks," in Proc. Int. Jt. Conf. Neural Netw, Vancouver, BC, Canada, 16-21 July 2006 (IEEE), 778–785. doi:10.1109/IJCNN.2006.246763

Gao, H., Jia, H., and Yang, L. (2022). An improved CEEMDAN-FE-TCN model for highway traffic flow prediction. *J. Adv. Transp.* 2022, 1–20. doi:10.1155/2022/2265000

Goodge, A., Hooi, B., Ng, S. K., and Ng, W. S. (2022). Lunar: Unifying local outlier detection methods via graph neural networks. *Proc. AAAI Conf. Artif. Intell.* 36 (6), 6737–6745. doi:10.1609/aaai.v36i6.20629

Gori, M., Monfardini, G., and Scarselli, F. (2005). A new model for learning in graph domains. *Proc. Int. Jt. Conf. Neural. Netw.* 2, 729–734. doi:10.1109/IJCNN.2005.1555942

Greenberg, M. L., and Castagna, J. P. (1992). Shear-wave velocity estimation in porous rocks: Theoretical formulation, preliminary verification and applications. *Geophys. Prospect* 40 (2), 195–209. doi:10.1111/j.1365-2478.1992.tb00371.x

Hamilton, W. L., Ying, R., and Leskovec, J. (2017). "Inductive representation learning on large graphs," in Proc. Int. Conf. Neural syst, Long Beach, California, United States, December 4–9, 2017 (Neural Information Processing Systems (NIPS)), 1025–1035. doi:10.5555/3294771.3294869

Hampson, D., Schuelke, J. S., and Quirein, J. (2001). Use of multiattribute transforms to predict log properties from seismic data. *Geophysics* 66 (1), 220–236. doi:10.1190/1.1444899

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA.* 79 (8), 2554–2558. doi:10.1073/PNAS.79.8.2554

Huang, N. E., Shen, Z., Long, S. R., Wu, M. L. C., Shih, H. H., Zheng, Q., et al. (1998). The empirical mode decomposition and the Hilbert spectrum for non-linear and non-stationary time series analysis. *Proc. R. Soc. Lond.* 454, 903–995. doi:10.1098/rspa.1998.0193

Jiang, C., Zhang, D., and Chen, S. (2021). Lithology identification from well-log curves via neural networks with additional geologic constraint. *Geophysics* 86 (5), IM85–IM100. doi:10.1190/geo2020-0676.1

Jiang, S., Liu, X., Xing, J., and Liang, L. (2018). "Study of S-wave velocity prediction model in shale formations," in *SEG global meeting abstr* (Beijing, China: SEG Global Meeting), 1343–1345. doi:10.1190/IGC2018-329

Kingma, D. P., and Ba, J. (2015). Adam: A method for stochastic optimization. ArXiv abs/1412.6980.

Kipf, T., and Welling, M. (2017). Semi-supervised classification with graph convolutional networks. ArXiv abs/1609.02907.

Kumar, A., Irsoy, O., Su, J., Bradbury, J., English, R., Pierce, B., et al. (2016). Ask me anything: Dynamic memory networks for natural language processing. *Proc. Int. Conf. Neural Inf. Process. Syst.* 48, 1378–1387. doi:10.5555/3045390.3045536

Liu, F. T., Ting, K. M., and Zhou, Z. H. (2008). "Isolation forest," in IEEE Int. Conf. Data Min, Pisa, Italy, 15-19 December 2008 (IEEE), 413–422. doi:10.1109/ICDM.2008.17

Magner, A., Baranwal, M., and Hero, A. O. (2022). Fundamental limits of deep graph convolutional networks for graph classification. *IEEE Trans. Inf. Theory* 68 (5), 3218–3233. doi:10.1109/TIT.2022.3145847

Misra, D. (2019). Mish: A self regularized non-monotonic activation function. ArXiv abs/1908.08681.

Narayanan, A., Chandramohan, M., Venkatesan, R., Chen, L., Liu, Y., and Jaiswal, S. (2017). Graph2vec: Learning distributed representations of graphs. ArXiv abs/1707.05005.

Ni, W., Qi, L., and Tao, F. (2017). "Prediction of shear wave velocity in shale reservoir based on logging data and machine learning," in Proc. Int. Conf. Softw. Eng. Knowl. Eng, London, UK, 21-23 October 2017 (IEEE), 231–234. doi:10.1109/ICKEA.2017.8169935

Omovie, S. J., and Castagna, J. P. (2021). Estimation of shear-wave velocities in unconventional shale reservoirs. *Geophys. Prospect.* 69 (6), 1316–1335. doi:10.1111/1365-2478.13096

Pilikos, G., and Faul, A. C. (2019). Bayesian modeling for uncertainty quantification in seismic compressive sensing. *Geophysics* 84 (2), P15–P25. doi:10.1190/geo2018-0145.1

Refunjol, X. E., Wallet, B. C., and Castagna, J. P. (2022). Fluid discrimination using detrended seismic impedance. *Interpretation* 10 (1), SA15–SA24. doi:10.1190/INT-2020-0219.1

Reshef, D. N., Reshef, Y. A., Finucane, H. K., Grossman, S. R., Mcvean, G., Turnbaugh, P. J., et al. (2011). Detecting novel associations in large data sets. *Science* 334 (6062), 1518–1524. doi:10.1126/science.1205438

Reshef, D. N., Reshef, Y. A., Sabeti, P. C., and Mitzenmacher, M. (2018). An empirical study of the maximal and total information coefficients and leading measures of dependence. *Ann. Appl. Stat.* 12 (1), 123–155. doi:10.1214/17-AOAS1093

Reshef, Y. A., Reshef, D. N., Sabeti, P. C., and Mitzenmacher, M. (2015). Equitability, interval estimation, and statistical power. ArXiv abs/1505.022122015.

Reshef, Y. A., Reshef, D. N., Sabeti, P. C., and Mitzenmacher, M. (2016). Measuring dependence powerfully and equitably. *J. Mach. Learn. Res.* 17 (1), 7406–7468. doi:10.5555/2946645.3053493

Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., and Monfardini, G. (2009). The graph neural network model. *IEEE Trans. Neural Netw.* 20 (1), 61–80. doi:10.1109/TNN.2008.2005605

Skekely, G. J., Rizzo, M. L., and Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. *Ann. Stat.* 35 (6), 2769–2794. doi:10.1214/009053607000000505

Sukhbaatar, S., Szlam, A., Weston, J., and Fergus, R. (2015). End-to-end memory networks. *Proc. Int. Conf. Neural Inf. Process. Syst.* 2, 2440–2448. doi:10.5555/2969442.2969512

Sun, Y., and Liu, Y. (2020). Prediction of S-wave velocity based on GRU neural network. *Oil Geophys. Prospect.* 55 (3), 484–492. doi:10.13810/j.cnki.issn.1000-7210.2020.03.001-en

Szabó, N. P., and Dobroka, M. (2017). Robust estimation of reservoir shaliness by iteratively reweighted factor analysis. *Geophysics* 82 (2), D69–D83. doi:10.1190/geo2016-0393.1

Vernik, L., Castagna, J., and Omovie, S. J. (2018). S-wave velocity prediction in unconventional shale reservoirs. *Geophysics* 83 (1), MR35–MR45. doi:10.1190/geo2017-0349.1

Vernik, L., and Kachanov, L. M. (2010). Modeling elastic properties of siliciclastic rocks. *Geophysics* 75 (6), E171–E182. doi:10.1190/1.3494031

Wang, J., and Cao, J. (2021). Data-driven S-wave velocity prediction method via a deep-learning-based deep convolutional gated recurrent unit fusion network. *Geophysics* 86 (6), M185–M196. doi:10.1190/geo2020-0886.1

Xu, K., Hu, W., Leskovec, J., and Jegelka, S. (2019). How powerful are graph neural networks? ArXiv abs/1810.00826.

Xu, S., and White, R. E. (1995). A new velocity model for clay-sand mixtures. *Geophys. Prosp.* 43 (1), 91–118. doi:10.1111/j.1365-2478.1995.tb00126.x

Xue, Y., Cao, J., Du, H., Lin, K., and Yao, Y. (2016). Seismic attenuation estimation using a complete ensemble empirical mode decomposition-based method. *Mar. Pet. Geol.* 71, 296–309. doi:10.1016/J.MARPETGEO.2016.01.011

Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., and Hovy, E. (2016). "Hierarchical attention networks for document classification," in *Proc. North ame. Chapter assoc. Comput. Ling* (San Diego, California: Association for Computational Linguistics). doi:10.18653/v1/N16-1174

Yu, T., Yin, H., and Zhu, Z. (2018). "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," in Proc. Int. J. Artif. Intell, Stockholm, Sweden, July 13–19, 2018 (International Joint Conferences on Artificial Intelligence Organization (IJCAI)), 3634–3640. doi:10.24963/ijcai.2018/505

Yu, T., Zhao, Z., Yan, Z., and Li, F. (2016). "Robust L1-norm matrixed locality preserving projection for discriminative subspace learning," in Proc. Int. Jt. Conf. Neural Netw, Vancouver, BC, Canada, 24-29 July 2016 (IEEE), 4199–4204. doi:10.1109/IJCNN.2016.7727747

Zeng, L., Ren, W., and Shan, L. (2020). Attention-based bidirectional gated recurrent unit neural networks for well logs prediction and lithology identification. *Neurocomputing* 414 (13), 153–171. doi:10.1016/j.neucom.2020.07.026

Zhang, Y., Zhong, H., Wu, Z., Zhou, H., and Ma, Q. (2020). Improvement of petrophysical workflow for shear wave velocity prediction based on machine learning methods for complex carbonate reservoirs. *J. Pet. Sci. Eng.* 192, 107234. doi:10.1016/j.petrol.2020.107234

Zhao, L., Song, Y., Zhang, C., Liu, Y., Wang, P., Lin, T., et al. (2020). T-GCN: A temporal graph convolutional network for traffic prediction. *IEEE Trans. Intell. Transp. Syst.* 21 (9), 3848–3858. doi:10.1109/TITS.2019.2935152

Zhong, C., Geng, F., Zhang, X., Zhang, Z., Wu, Z., and Jiang, Y. (2021). "Shear wave velocity prediction of carbonate reservoirs based on CatBoost," in Int. Conf. Artif. Intell. Big Data, Chengdu, China, 28-31 May 2021 (IEEE), 622–626. doi:10.1109/ICAIBD51990.2021.9459061

Zhou, F., Cao, C., Zhang, K., Trajcevski, G., Zhong, T., and Geng, J. (2019a). "Meta-GNN: On few-shot node classification in graph meta-learning," in *Proc. Int. Conf. Inf. Knowl. Manag* (Washington, DC, USA: ACM), 2357–2360. doi:10.1145/3357384.3358106

Zhou, Y., Li, T., Shi, J. Y., and Qian, Z. (2019b). A CEEMDAN and XGBOOST-based approach to forecast crude oil prices. *Complex* 2019, 1–15. doi:10.1155/2019/4392785

frontiers | Frontiers in Earth Science

TYPE Original Research
PUBLISHED 23 February 2023
DOI 10.3389/feart.2023.1103522

OPEN ACCESS

EDITED BY
Peng Zhenming,
University of Electronic Science and Technology of China, China

REVIEWED BY
Shu Li,
Jishou University, China
Bibo Yue,
Southwest Petroleum University, China
Gulan Zhang,
Southwest Petroleum University, China

*CORRESPONDENCE
Lin Tian,
1610356358@qq.com

SPECIALTY SECTION
This article was submitted to Environmental Informatics and Remote Sensing, a section of the journal Frontiers in Earth Science

RECEIVED 20 November 2022
ACCEPTED 03 February 2023
PUBLISHED 23 February 2023

CITATION
Tian L and Qin S (2023), Reconstruction of seismic data based on SFISTA and curvelet transform.
Front. Earth Sci. 11:1103522.
doi: 10.3389/feart.2023.1103522

Reconstruction of seismic data based on SFISTA and curvelet transform

Lin Tian[1,2]* and Si Qin[1]

[1]School of Electronic Engineering, Yili Normal University, Yining, China, [2]Key Laboratory of Signal Detection and Control Technology, Yili Normal University, Yining, China

In seismic data processing, the reconstruction and interpolation of missing traces are essential tasks. To overcome the limitations of irregularly sampled seismic data, this paper proposes a seismic data interpolation method combining the smoothing fast iterative soft threshold algorithm (SFISTA) and the curvelet transform; this method uses the curvelet domain as the sparse domain. For comparison, the contourlet transform is used for different sparse domains, and the fast iterative shrinkage-thresholding algorithm (FISTA) is used for different optimization algorithms. Numerical modeling and real data show that the SFISTA method in the curvelet domain can give better reconstruction effects and higher reconstruction accuracy than those in the contourlet domain with the FISTA method; in addition, the SFISTA method in the curvelet domain can be used to reconstruct the missing traces of three-dimensional seismic data.

KEYWORDS

seismic data reconstruction, sparse constraint, curvelet transform, contourlet transform, SFISTA

1 Introduction

In complex environments, seismic data reconstruction has great significance as a recovery technique. Under external disturbance, irregularly sampled seismic data will affect further geological data processing such as migration imaging and data interpretation. In order to obtain high-quality seismic data, interpolation reconstruction is needed to approximate the original data. In recent years, under the compressive sensing theory, seismic data reconstruction methods based on sparse constraints have become more and more popular. It mainly consists of the sparse transform, measurement matrix, and reconstruction algorithm. The sparse transforms that are often used include the Fourier transform (Zhang et al., 2013; Ciabarri et al., 2014), curvelet transform (Hennenfent et al., 2010; Liu et al., 2015; Zhang et al., 2017; Zhang et al., 2019; Tian and Qin, 2022), contourlet transform (Eslami and Radha, 2006), and seislet transform (Liu W et al., 2016). Because the curvelet transform undergoes multi-scale and multi-direction analysis and can perform the optimal local decomposition of seismic data (Yang et al., 2017), the curvelet transform is employed in this paper as a sparse transform, and the contourlet transform is also used in this paper for comparison analysis.

A classical sparse recovery problem usually requires minimizing the L_0 norm, which is NP-hard. The L_1 norm is the approximate of the L_0 norm, which is a convex function, and can be solved by the convex optimization algorithms or tools; so, the L_0 norm is replaced by the L_1 norm for simplicity and effectiveness. The iterative soft threshold algorithm (ISTA) shows great advantages (Daubechies et al., 2004; Mohsin et al., 2015) for a convex optimization algorithm. Gradually, a faster ISTA algorithm (FISTA) has been developed. FISTA is more

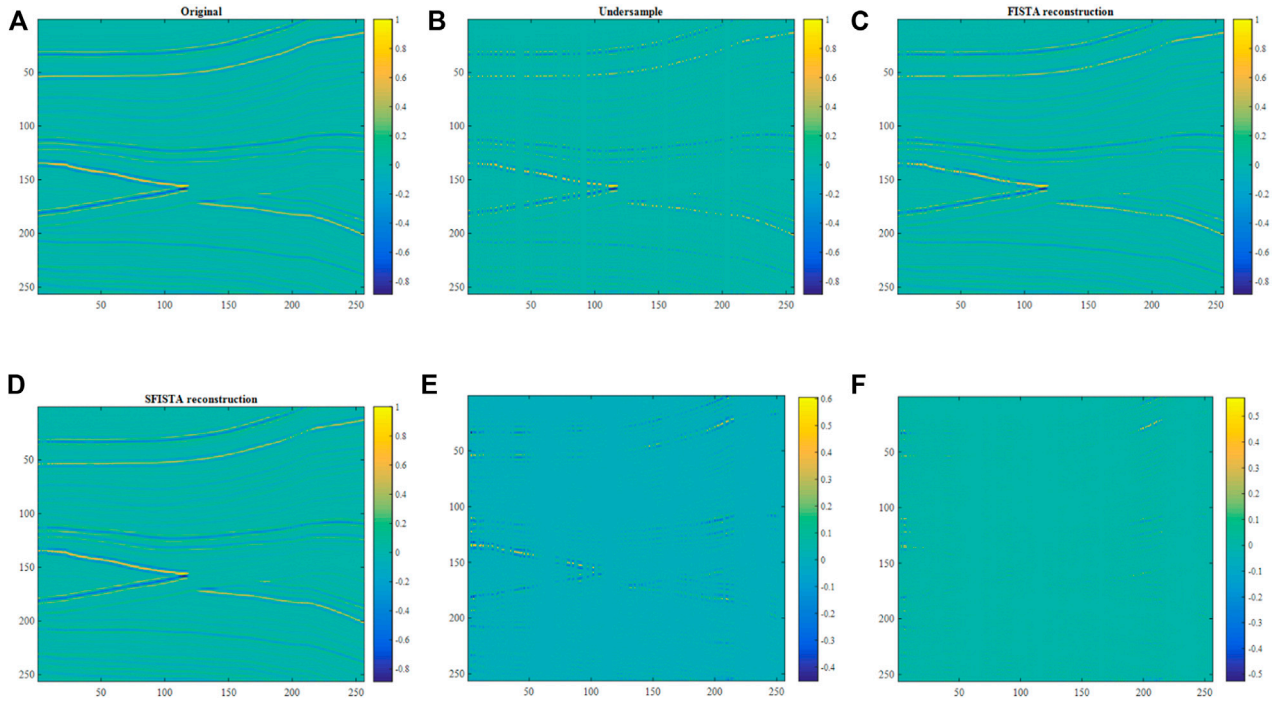


FIGURE 1
Reconstruction results of synthetic data in the contourlet domain. **(A)** Original data; **(B)** 50% of randomly missing data; **(C)** interpolated data obtained by the FISTA method; **(D)** interpolated data obtained by the SFISTA method; **(E)** reconstructed errors obtained by the FISTA method; **(F)** reconstructed errors obtained by the SFISTA method.

TABLE 1 Reconstruction error and SNR of different methods.

	FISTA	SFISTA
Reconstruction error	784.34	450.45
SNR	10.12	12.41

suitable the synthesis approach to sparse recovery. And for the analysis approach to sparse recovery, Tan et al. (2014) proposed a monotone version of the fast iterative shrinkage-thresholding algorithm, which is the smoothing fast iterative soft threshold algorithm (SFISTA). In this paper, we introduce the SFISTA method-based curvelet transform to the seismic data interpolation reconstruction problem (Tan et al., 2014; Liu Y et al., 2016; Pokala and Seelamantula, 2020; Shen et al., 2020). The theory is given in section 2, and experimental results are given in section 3.

2 Theory

Assuming that the observed seismic data is y and the downsampling matrix is U, the irregular missing seismic data can be modeled as follows:

$$y = Ux + n, \tag{1}$$

where n is a randomly generated noise and x is the original seismic data.

The interpolation problem can be expressed as follows:

$$\min_x \frac{1}{2}\|y - Ux\|_2^2 + \lambda\|\Psi x\|_1, \tag{2}$$

where $\|\cdot\|_2$ is the L_2 norm, $\|\cdot\|_1$ is the L_1 norm, λ is the regularization parameter, and Ψ is the analysis operator. Moreover, $\Psi^*\Psi = I$, where Ψ^* denotes the adjoint of the operator Ψ and I is the identity matrix. The sparse transform can be expressed as $x = \Psi\alpha$, where Ψ is the transform operator and α is the sparse coefficient. Let us denote the following equation:

$$f(x) = \frac{1}{2}\|Ux - y\|_2^2, \tag{3}$$

where $f(x)$ is the smooth part.

$$g(\Psi x) = \lambda\|\Psi x\|_1, \tag{4}$$

where $g(\Psi x)$ is the non-smooth part that needs to be replaced by the Moreau envelope with an approximately smooth form $g_\mu(\Psi x)$, where μ is the smooth approximate parameter.

$$\nabla g_\mu(\Psi x) = \frac{1}{\mu}\Psi^*\left(\Psi x - T_{\lambda\mu}(\Psi x)\right), \tag{5}$$

where ∇ is the gradient and $T_{\lambda\mu}$ is the soft threshold operator; the threshold of $T_{\lambda\mu}$ is the multiplication of λ and μ.

In this paper, the process of the smoothing fast iterative soft threshold algorithm (Shen et al., 2020) for optimization is expressed as follows:

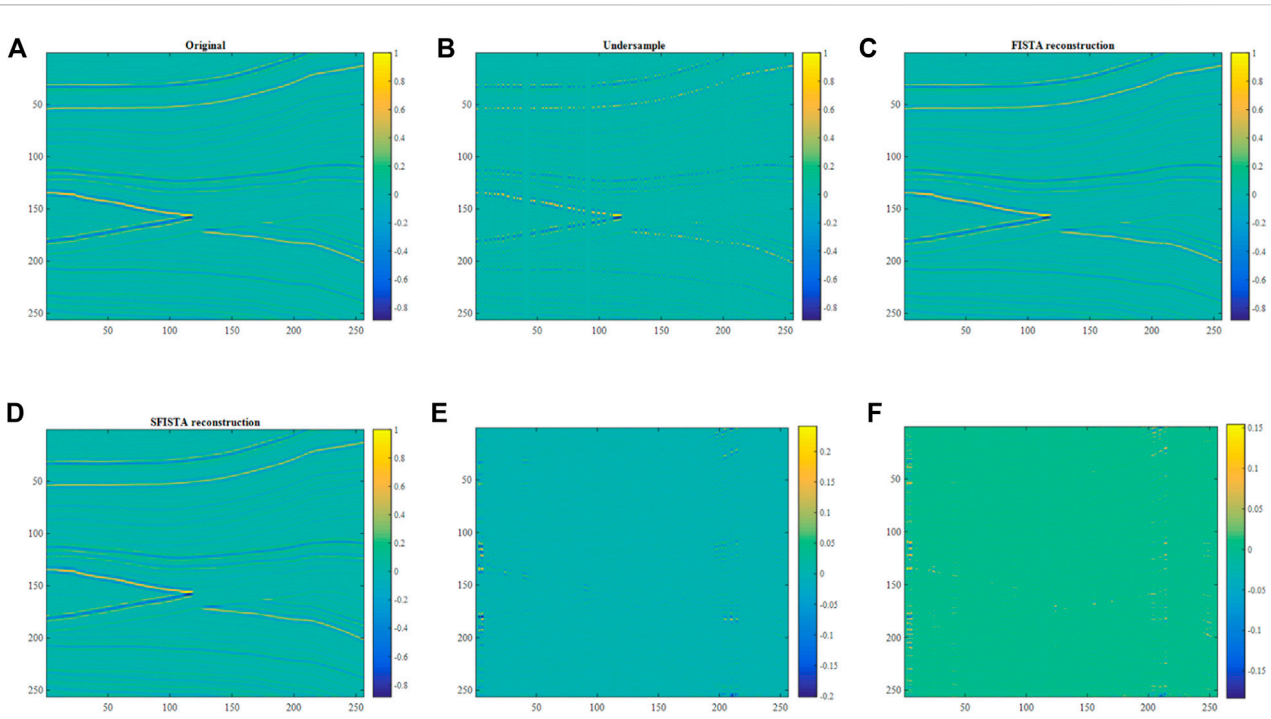


FIGURE 2
Reconstruction results of synthetic data in the curvelet domain. **(A)** Original data; **(B)** 50% of randomly missing data; **(C)** interpolated data obtained by the FISTA method; **(D)** interpolated data obtained by the SFISTA method; **(E)** reconstructed errors obtained by the FISTA method; **(F)** reconstructed errors obtained by the SFISTA method.

TABLE 2 Reconstruction error and SNR of different methods.

	FISTA	SFISTA
Reconstruction error	223.16	193.62
SNR	17.75	20.67

$$\nabla f(\hat{x}_k) = U^T(U\hat{x}_k - y), \tag{6}$$

where k is the number of iterations, and $\hat{x}_0 = x_0$ and $t_0 = 1$.

$$\nabla g_\mu(\Psi x_k) = \frac{1}{\mu}\Psi^*\left(\Psi x_k - T_{\lambda\mu}(\Psi x_k)\right). \tag{7}$$

The core iterations of the SFISTA are as follows:

$$x_{k+1} = \hat{x}_k - \frac{1}{L}\left(\nabla f(\hat{x}_k) + \nabla g_\mu(\Psi x_k)\right), \tag{8}$$

where L is the Lipschitz constant. At $0 < \gamma \le 1$, $L = 1/\gamma$. Substituting (6) and (7) in (8), the aforementioned equation is equivalent to the following equation:

$$x_{k+1} = \hat{x}_k - \frac{\gamma}{\mu}\left(x_k - \Psi^* T_{\lambda\mu}(\Psi x_k)\right) + \gamma U^T(y - U\hat{x}_k), \tag{9}$$

where γ is the iteration step size. t_{k+1} is expressed as follows:

$$t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}. \tag{10}$$

Then, we derive the following equation:

$$\hat{x}_{k+1} = x_{k+1} + \frac{t_k - 1}{t_{k+1}}(x_{k+1} - x_k). \tag{11}$$

The specific algorithm steps are as follows:

Parameters: $\lambda = 10^{-3}, \mu = 10^{-3}/\lambda = 1, \gamma = 1/(1 + 1/\mu)$
Initialization: $x_0, \hat{x}_0, t_0 = 1$
When not convergent, the following equations are used to calculate:
$x_{k+1} = \hat{x}_k - \frac{\gamma}{\mu}(x_k - \Psi^* T_{\lambda\mu}(\Psi x_k)) + \gamma U^T(y - U\hat{x}_k)$
$t_{k+1} = \frac{1+\sqrt{1+4t_k^2}}{2}$
$\hat{x}_{k+1} = x_{k+1} + \frac{t_k - 1}{t_{k+1}}(x_{k+1} - x_k)$
RLNE $= \frac{\|\hat{x} - x\|_2}{\|x\|_2}$
end if RLNE $< 10^{-6}$ or the maximum number of iterations is reached.
end
Output: x

Algorithm 1. SFISTA for seismic data interpolation

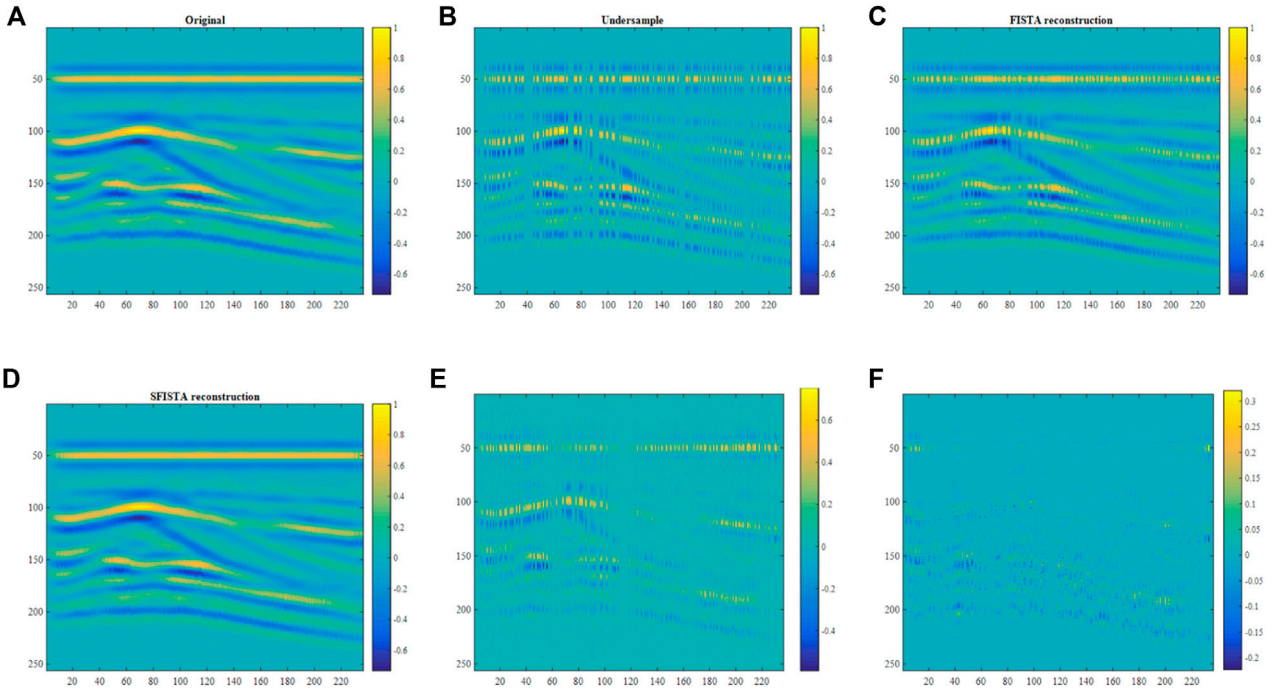


FIGURE 3
Reconstruction results of the coral reef model in the contourlet domain. **(A)** Original data; **(B)** 50% of randomly missing data; **(C)** interpolated data obtained by the FISTA method; **(D)** interpolated data obtained by the SFISTA method; **(E)** reconstructed errors obtained by the FISTA method; **(F)** reconstructed errors obtained by the SFISTA method.

TABLE 3 Reconstruction error and SNR of different methods.

	FISTA	SFISTA
Reconstruction error	1695.09	407.17
SNR	9.11	25.07

Parameters: $\lambda = 10^{-3}, \gamma = 1$
Initialization: $\alpha_0, \hat{\alpha}_0, t_0 = 1$
When not convergent, the following equations are used to calculate:
$\alpha_{k+1} = T_{\gamma\lambda}(\gamma\Psi U^T(y - U\Psi^*\alpha_k) + \hat{\alpha}_k)$
$t_{k+1} = \frac{1+\sqrt{1+4t_k^2}}{2}$
$\hat{\alpha}_{k+1} = \alpha_{k+1} + \frac{t_k - 1}{t_{k+1}}(\alpha_{k+1} - \alpha_k)$
RLNE $= \frac{\|\hat{\alpha} - \alpha\|_2}{\|\alpha\|_2}$
end if RLNE $< 10^{-6}$ or the maximum number of iterations is reached.
end
Output: α

Algorithm 2. FISTA for seismic data interpolation

3 Examples

In this section, we conduct numerical experiments with different seismic data to demonstrate the reconstruction performance of the SFISTA method in the curvelet domain. The algorithm performance is evaluated by interpolation results, the average amplitude spectrum, single-channel interpolation effect, reconstruction error, and signal-to-noise ratio. Numerical experiments are used to test the method. At last, the paper continues to test the 3D interpolation effect of the interpolation method of the proposed method. The experiments are conducted on a Millet computer running on the Windows 10 operating system and Inter Core m3-6Y30.

3.1 Synthetic examples

3.1.1 Seismic data interpolation in the contourlet domain

The regularization parameters of the FISTA and SFISTA methods are set to 10^{-3}. The step size of the FISTA is 1, and the step size of the SFISTA is set to 0.5 (Liu Y et al., 2016); the number of iterations is set as 500. The two-dimensional data tests have the same parameters.

Part data on the Marmousi2 model (Martin et al., 2006) are selected as test data; the data are shown in Figure 1A. Figure 1B shows the irregular data with 50% of the traces removed randomly, and this part's records are padded with zeros. The sparse domain is the contourlet domain. Figures 1C,D are interpolation results of Figure 1B using FISTA and SFISTA methods based on the contourlet transform. Figures 1E,F show the reconstruction errors that correspond to the FISTA and SFISTA methods, in which the

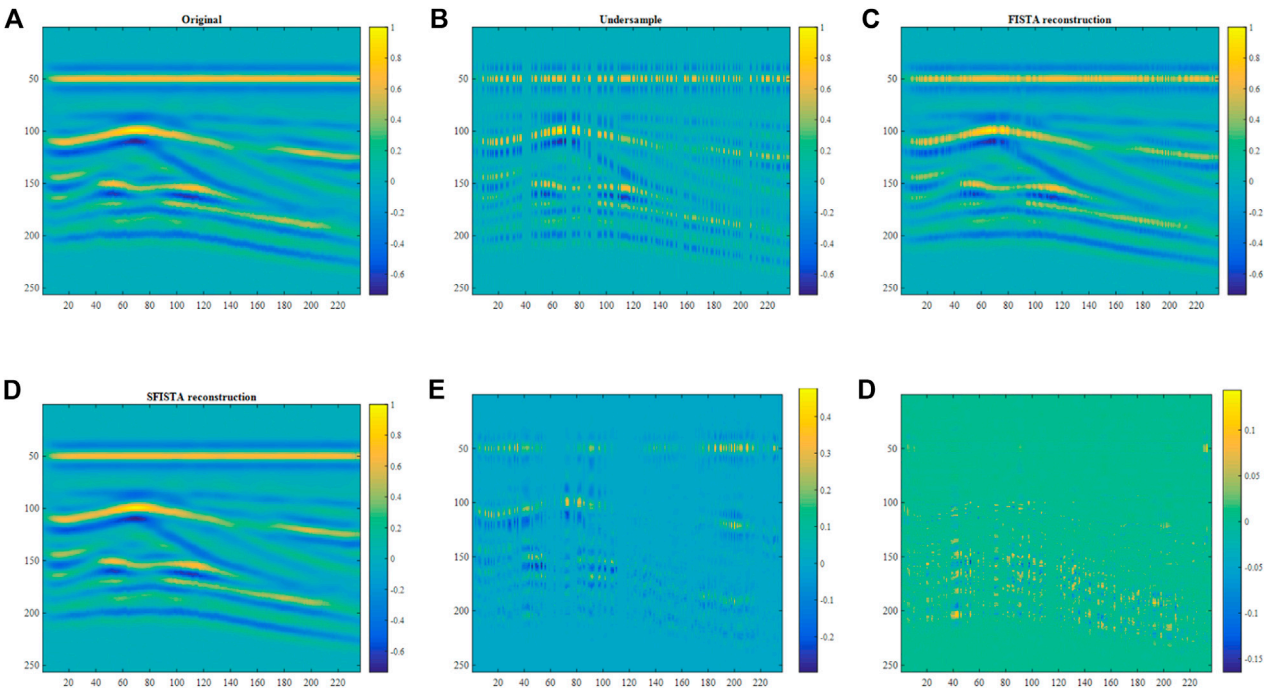


FIGURE 4
Reconstruction results of the coral reef model in the curvelet domain. **(A)** Original data; **(B)** 50% of randomly missing data; **(C)** interpolated data obtained by the FISTA method; **(D)** interpolated data obtained by the SFISTA method; **(E)** reconstructed errors obtained by the FISTA method; **(F)** reconstructed errors obtained by the SFISTA method.

TABLE 4 Reconstruction error and SNR of different methods.

	FISTA	SFISTA
Reconstruction error	902.74	323.55
SNR	15.45	27.72

reconstruction residuals of the SFISTA method have smaller amplitude ranges than those of the FISTA method. It is obvious that the reconstruction method of SFISTA based on the contourlet transform is more effective.

The performance of the proposed method in seismic data interpolation could be evaluated using different qualitative and quantitative analyzing tools. The reconstruction error is the general quantitative evaluation tool used in seismic data interpolation; the reconstruction error is defined to be $\sum |x - \hat{x}|$, where x denotes the original data and $\hat{x}$ denotes the reconstructed seismic data. If the reconstruction error is smaller, the reconstructed seismic data will be closer to the original data. SNR is the signal-to-noise ratio, which is defined as $\mathrm{SNR} = 10\log_{10}\frac{\|x\|_2^2}{\|x-\hat{x}\|_2^2}$. A higher SNR value means that the data have better reconstruction quality. The reconstruction error and signal-to-noise ratio are illustrated in Table 1. Table 1 shows that the SFISTA method has a smaller reconstruction error value and higher SNR. Compared to the two methods, the SFISTA method based on the contourlet transform has a lower reconstruction error and higher SNR, and therefore, the SFISTA method shows better performance.

3.1.2 Seismic data interpolation in the curvelet domain

The part of the Marmousi2 model is shown in Figure 2A. Figure 2B shows the irregular data with 50% of the traces missing randomly. Figures 2C, D show interpolation effects using FISTA and SFISTA methods based on the curvelet transform. Figures 2E, F show the reconstruction errors that correspond to the FISTA and SFISTA methods, in which the reconstruction residuals of the SFISTA method have a smaller magnitude than those of the FISTA method. It is obvious that the reconstruction method of SFISTA in the curvelet domain is more effective.

Comparing Figures 1, 2, the SFISTA method in the curvelet domain has lower reconstruction errors when using the qualitative analyzing method. Comparing Tables 1, 2, the SFISTA method in the curvelet domain has lower reconstruction errors and higher SNR when using the quantitative analyzing method. The SFISTA method shows better performance when using the curvelet domain as the sparse domain.

3.1.3 Coral reef model tests in the contourlet domain

The selected coral reef synthetic data are illustrated in Figure 3A. Figure 3B shows corrupted data with 50% of traces

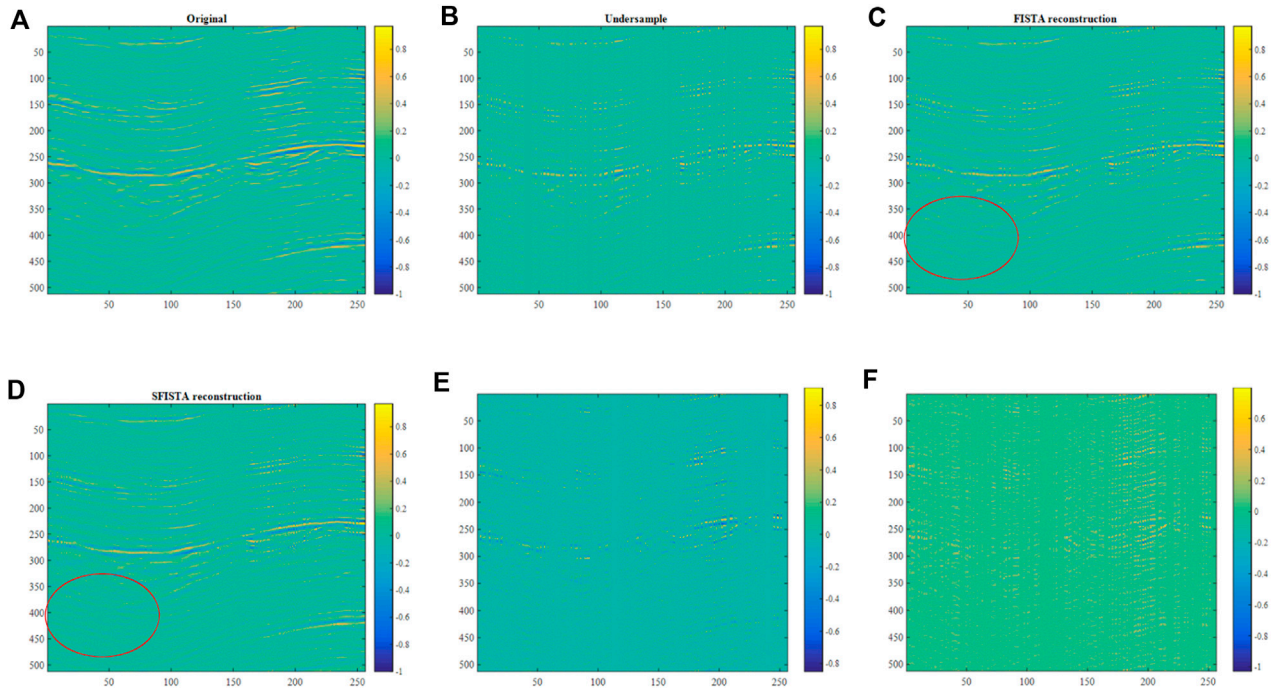


FIGURE 5
Reconstruction results of field data in the contourlet domain. **(A)** Original data; **(B)** 50% of randomly missing data; **(C)** interpolated data obtained by the FISTA method; **(D)** interpolated data obtained by the SFISTA method; **(E)** reconstructed errors obtained by the FISTA method; **(F)** reconstructed errors obtained by the SFISTA method.

TABLE 5 Reconstruction error and SNR of different methods.

	FISTA	SFISTA
Reconstruction error	3904.69	3633.50
SNR	6.43	7.16

missing randomly. Figures 3C, D show interpolation effects using FISTA and SFISTA methods in the contourlet domain. Figures 3E, F show the reconstructed errors between the reconstruction and original data; Figure 3F shows a smaller reconstructed error. Table 3 shows the reconstruction error and SNR, which demonstrates the validity of the SFISTA method in the contourlet domain.

3.1.4 Coral reef model tests in the curvelet domain

The selected Coral reef synthetic data are tested in the curvelet domain; the original data are shown in Figure 4A. Figure 4B shows corrupted data with 50% of traces missing randomly. Figures 4C, D show interpolation results using FISTA and SFISTA methods in the curvelet domain. Figures 4E, F show the reconstructed errors between the reconstruction and original data; Figure 4F shows a smaller reconstructed error. Table 4 shows the reconstruction error and SNR, which demonstrates the validity of the proposed method.

Comparing the detailed values in Tables 3, 4, the reconstruction performance of the SFISTA method in the curvelet domain is better than that in the contourlet domain.

3.2 Field examples

3.2.1 Seismic data interpolation in the contourlet domain

The field data are illustrated in Figure 5A. Figure 5B shows corrupted data with 50% of traces missing randomly. Figures 5C, D show interpolation effects using FISTA and SFISTA methods in the contourlet domain. By comparing the interpolation results of the ellipse regions, we learn that the reconstruction results of the SFISTA method in the contourlet domain are similar to the original data. Figures 5E, F show the reconstructed errors between the reconstruction and original data; Figure 5F shows a smaller reconstructed error. Table 5 shows the reconstruction error and SNR, which demonstrates the validity of the SFISTA method in the contourlet domain.

3.2.2 Seismic data interpolation in the curvelet domain

The field data are illustrated in Figure 6A. Figure 6B shows incomplete seismic data with 50% of traces missing randomly. Figures 6C, D show interpolation effects using FISTA and SFISTA methods in the curvelet domain. By comparing the interpolation results of the ellipse regions, we learn that the reconstruction data on the SFISTA method in the curvelet domain are similar to the original data. Figures 6E, F show the reconstructed errors between the reconstruction and original data; Figure 6F shows a smaller reconstructed error.

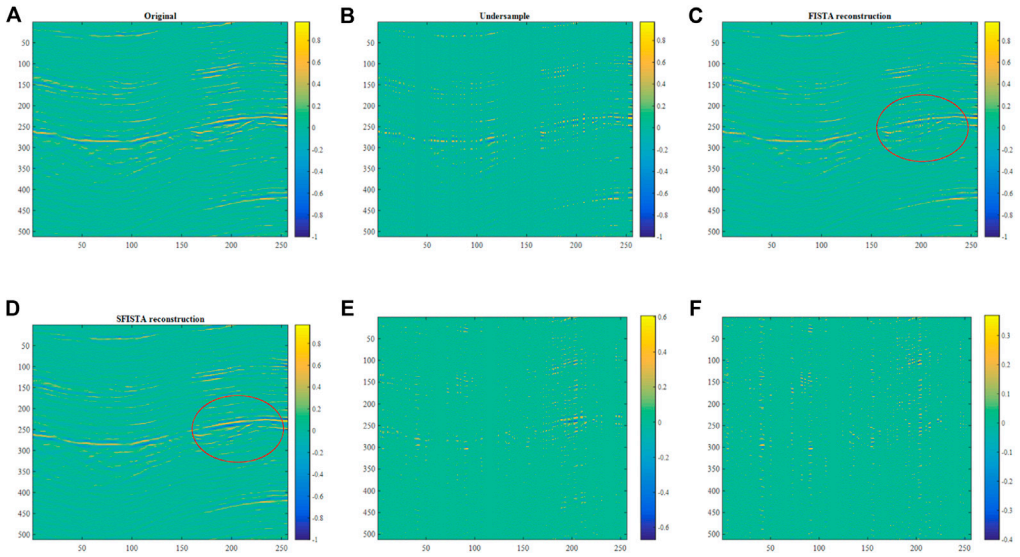


FIGURE 6
Reconstruction results of field data in the curvelet domain. **(A)** Original data; **(B)** 50% of randomly missing data; **(C)** interpolated data obtained by the FISTA method; **(D)** interpolated data obtained by the SFISTA method; **(E)** reconstructed errors obtained by the FISTA method; **(F)** reconstructed errors obtained by the SFISTA method.

TABLE 6 Reconstruction error and SNR of different methods.

	FISTA	SFISTA
Reconstruction error	2968.24	2220.78
SNR	9.18	12.91

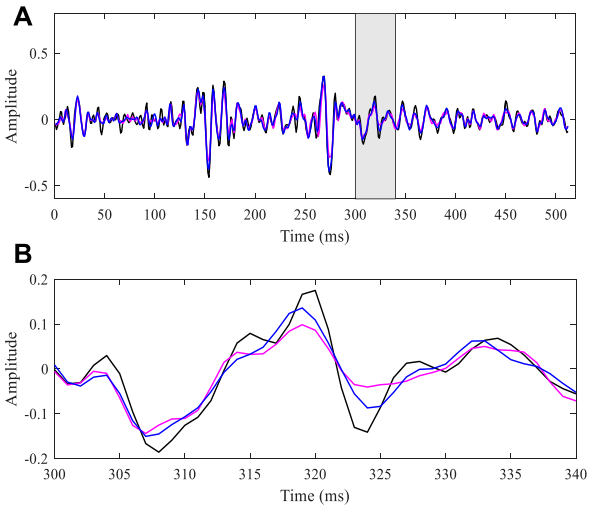


FIGURE 7
(A) Comparison of a missing single-trace amplitude interpolated with FISTA and SFISTA methods; **(B)** zoomed traces from the transparent gray window of **(A)**.

Table 6 shows the reconstruction error and SNR results of the two algorithms in the curvelet domain. Comparing the two methods, SFISTA performs better with a lower error and higher SNR.

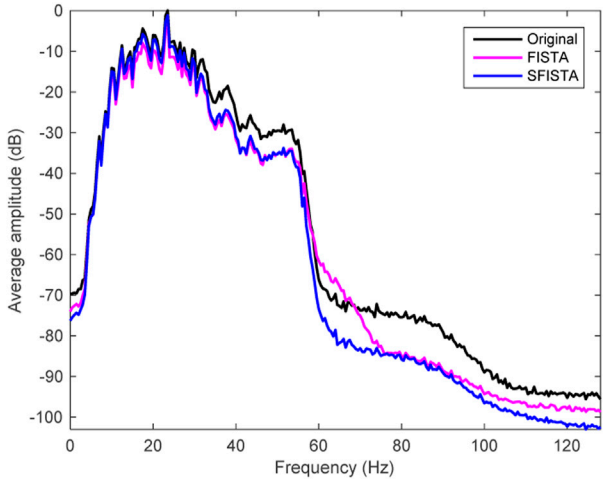


FIGURE 8
Average amplitude spectrum from original data, interpolated with FISTA and SFISTA methods.

Figure 7A plots the interpolation single trace of the missing trace with zero values. In Figures 7, the black, pink, and blue lines correspond to the original data and those obtained by FISTA and SFISTA methods, respectively. Figure 7B shows the zoomed traces from the transparent gray window of Figure 7A. A detailed comparison reveals that the blue line is similar to the black line. We can, therefore, affirm that the proposed SFISTA method can better restore the significant features of the useful signal than the FISTA method.

Comparing the detailed values in Tables 5, 6, the SFISTA method in the curvelet domain is better than that in the contourlet domain.

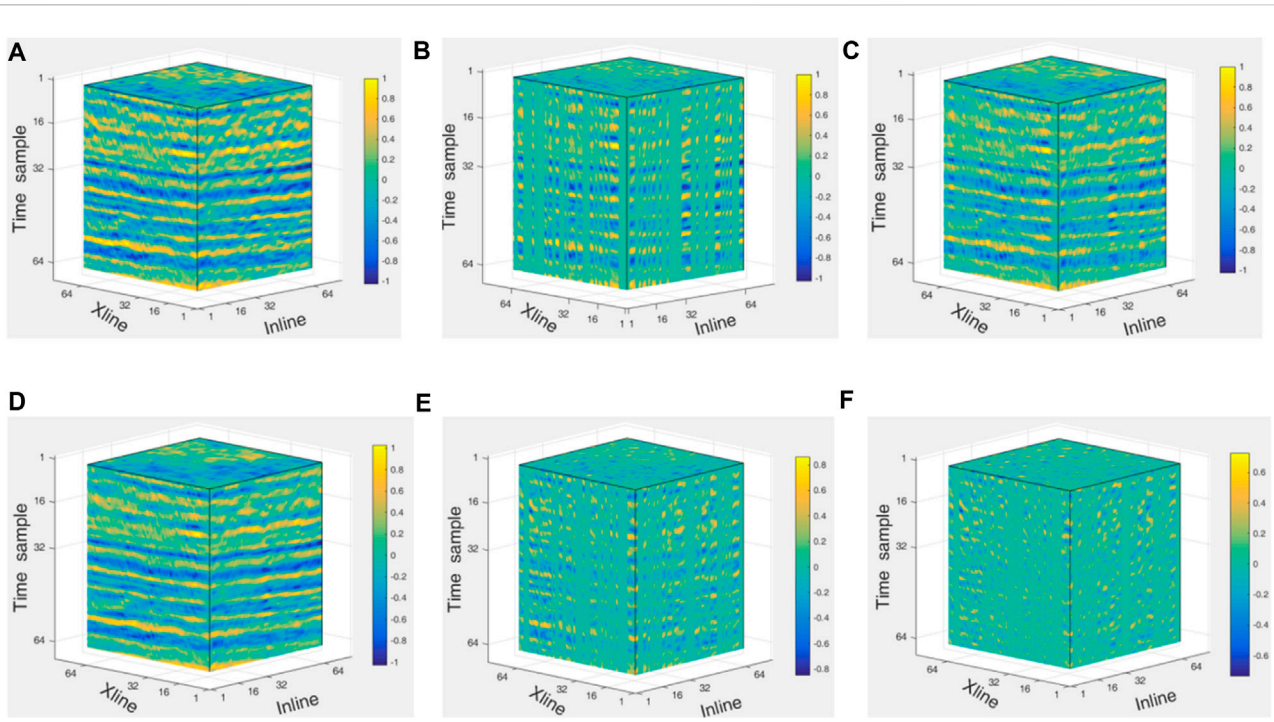


FIGURE 9
Reconstruction results of three-dimensional field data in the curvelet domain. **(A)** Original data; **(B)** 51% of randomly missing data; **(C)** interpolated data obtained by the FISTA method; **(D)** interpolated data obtained by the SFISTA method; **(E)** reconstructed errors obtained by the FISTA method; **(F)** reconstructed errors obtained by the SFISTA method.

The performance of these methods could be investigated more by comparing the average amplitude spectrum (Mafakheri et al., 2022). The average amplitude spectrum presents the original data and those obtained by FISTA and SFISTA methods, as shown in Figure 8 by the black, pink, and blue lines, respectively. The SFISTA method in the curvelet domain gives a closer spectrum to that of the original data, particularly in the range of 5–30 Hz. In this case, our method has better performance.

3.2.3 Three-dimensional seismic data tests

The experimental results of three sets of 2D data show that SFISTA based on the curvelet transform shows good performance. Next, this method will be tested for 3D seismic data reconstruction. The 3D data (size: 64 × 64 × 64) are from the software package of MathGeo 2020 (https://gitee.com/sevenysw/MathGeo2020). The 3D discrete curvelet transform method comes from the reference of Ying et al. (2005).

The original data are shown in Figure 9A. Figure 9B shows corrupted data with 51% of the traces removed randomly; the iteration number of the FISTA and SFISTA methods is 1000. Figures 9C, D show the interpolation results using the FISTA and SFISTA methods in the curvelet domain. Figures 9E, F show reconstructed errors obtained by the FISTA and SFISTA methods. The quantitatively recovered reconstruction error and SNR are shown in Table 7; Table 7 illustrates that the SFISTA method is better than the FISTA method.

Figure 10A plots the interpolation single trace from Figures 9C, D. The black, pink, and blue lines represent the original data and those obtained by FISTA and SFISTA methods, respectively. Figure 10B shows the zoomed traces from the transparent gray window of Figure 10A. A detailed comparison reveals that the blue line is similar to the black line. We can, therefore, affirm that the proposed SFISTA method can better restore the significant features of the useful signal than the FISTA approach.

TABLE 7 Reconstruction error and SNR of different methods.

	FISTA	SFISTA
Reconstruction error	27781.58	18396.99
SNR	7.32	10.79

The performance of these methods could be investigated more by comparing the average amplitude spectrum. The average amplitude spectrum of the original data and FISTA and SFISTA methods are shown in Figure 11 by the black, pink, and blue lines, respectively. Interestingly, in this case, the reconstruction data obtained by our method are very similar to the original data.

4 Conclusion

In this paper, we proposed a new seismic data reconstruction method combining the curvelet transform and the SFISTA method.

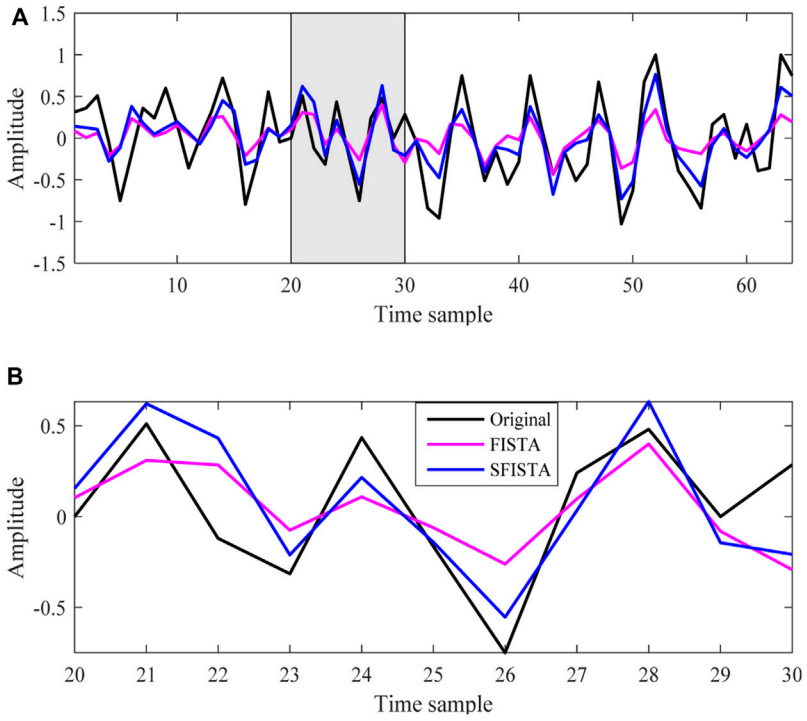


FIGURE 10
(A) Comparison of a missing single-trace amplitude interpolated with FISTA and SFISTA methods; **(B)** zoomed traces from the transparent gray window of **(A)**.

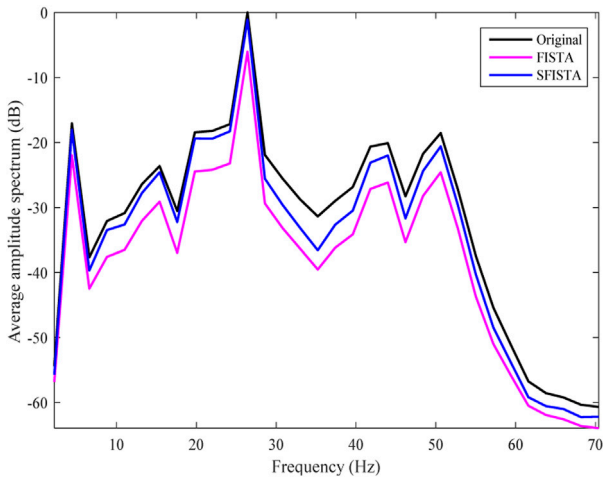


FIGURE 11
Average amplitude spectrum from original data, interpolated with FISTA and SFISTA methods.

Comparing the results obtained when using the contourlet and curvelet domains as the sparse domain, it can be concluded that the optimization algorithm in the curvelet domain has better performance. In the same sparse domain, comparing the FISTA and SFISTA methods, it can be concluded that SFISTA shows better performance. The seismic data reconstruction effects of SFISTA based on the curvelet transform have been demonstrated by quantitative and qualitative comparisons with several sets of data. The proposed method can be used for 2D and 3D seismic data reconstruction.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding author.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Funding

This work was supported by the Yili Normal University Foundation under Grant 22XKZZ22, the Talent Position Project of Yili Normal University under Grant YSXSGG22006, and the National Natural Science Foundation of China under Grant 61761043.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Ciabarri, F., Mazzotti, A., Stucchi, E., and Bienati, N. (2014). Appraisal problem in the 3D least squares Fourier seismic data reconstruction. *Geophys. Prospect.* 63 (2), 296–314. doi:10.1111/1365-2478.12192

Daubechies, I., Defrise, M., and Mol, C. D. (2004). An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Commun. Pure Appl. Math.* 57 (11), 1413–1457. doi:10.1002/cpa.20042

Eslami, R., and Radha, H. (2006). Translation-invariant contourlet transform and its application to image denoising. *IEEE Trans. Image Process* 15 (11), 3362–3374. doi:10.1109/tip.2006.881992

Hennenfent, G., Fenelon, L., and Herrmann, F. J. (2010). Nonequispaced curvelet transform for seismic data reconstruction: A sparsity-promoting approach. *GEOPHYSICS* 75 (6), 203–210. doi:10.1190/1.3494032

Liu, W., Cao, S., Gan, S., Chen, Y., Zu, S., and Jin, Z. (2016). One-step slope estimation for dealiased seismic data reconstruction via iterative seislet thresholding. *IEEE Geoscience Remote Sens. Lett.* 13 (10), 1462–1466. doi:10.1109/lgrs.2016.2591939

Liu, W., Cao, S., Li, G., and He, Y. (2015). Reconstruction of seismic data with missing traces based on local random sampling and curvelet transform. *J. Appl. Geophys.* 115, 129–139. doi:10.1016/j.jappgeo.2015.02.009

Liu, Y., Zhan, Z., Cai, J.-F., Guo, D., Chen, Z., and Qu, X. (2016). Projected iterative soft-thresholding algorithm for tight frames in compressed sensing magnetic resonance imaging. *IEEE Trans. Med. Imaging* 35 (9), 2130–2140. doi:10.1109/TMI.2016.2550080

Mafakheri, J., Kahoo, A. R., Anvari, R., Mohammadi, M., Radad, M., and Monfared, M. S. (2022). Expand dimensional of seismic data and random noise attenuation using Low-Rank estimation. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* 15, 2773–2781. doi:10.1109/jstars.2022.3162763

Martin, G. S., Wiley, R., and Marfurt, K. J. (2006). Marmousi2: An elastic upgrade for Marmousi. *Lead. Edge* 25 (2), 156–166. doi:10.1190/1.2172306

Mohsin, Y. Q., Ongie, G., and Jacob, M. (2015). Iterative shrinkage algorithm for patch-smoothness regularized medical image recovery. *IEEE Trans. Med. Imaging* 34 (12), 2417–2428. doi:10.1109/tmi.2015.2398466

Pokala, P., and Seelamantula, C. (2020). "Accelerated weighted ℓ_1-minimization for MRI reconstruction under tight frames in complex domain," in International Conference on Signal Processing and Communications (SPCOM), Bangalore, India, 19-24 July 2020. doi:10.1109/SPCOM50965.2020.9179611

Shen, L., Chu, Z., Tan, L., Chen, D., and Ye, F. (2020). Improving the sound source identification performance of sparsity constrained deconvolution beamforming utilizing SFISTA. *Shock Vib.* 2020, 1–9. doi:10.1155/2020/1482812

Tan, Z., Eldar, Y. C., Beck, A., and Nehorai, A. (2014). Smoothing and decomposition for analysis sparse recovery. *IEEE Trans. Signal Process.* 62 (7), 1762–1774. doi:10.1109/tsp.2014.2304932

Tian, L., and Qin, S. (2022). "Seismic data interpolation by the projected iterative soft-threshold algorithm for tight frame," in 2022 4th International Conference on Image Processing and Machine Vision (IPMV), Hong Kong China, March 25 - 27, 2022. doi:10.1145/3529446.3529460

Yang, H., Long, Y., Lin, J., Zhang, F., and Chen, Z. (2017). A seismic interpolation and denoising method with curvelet transform matching filter. *Acta Geophys.* 65 (5), 1029–1042. doi:10.1007/s11600-017-0078-x

Ying, L., Demanet, L., and Candes, E. (2005). 3D discrete curvelet transform. *Proc. SPIE - Int. Soc. Opt. Eng.* 50 (3), 351–361. doi:10.1117/12.616205

Zhang, H., Chen, X.-H., and Wu, X.-M. (2013). Seismic data reconstruction based on CS and Fourier theory. *Appl. Geophys.* 10 (2), 170–180. doi:10.1007/s11770-013-0375-3

Zhang, H., Chen, X.-H., and Zhang, L.-Y. (2017). 3D simultaneous seismic data reconstruction and noise suppression based on the curvelet transform. *Appl. Geophys.* 14 (1), 87–95. doi:10.1007/s11770-017-0607-z

Zhang, H., Diao, S., Chen, W., Huang, G., Li, H., and Bai, M. (2019). Curvelet reconstruction of non-uniformly sampled seismic data using the linearized Bregman method. *Geophys. Prospect.* 67 (5), 1201–1218. doi:10.1111/1365-2478.12762

frontiers | Frontiers in Earth Science

TYPE Original Research
PUBLISHED 27 February 2023
DOI 10.3389/feart.2023.1104488

OPEN ACCESS

EDITED BY
Peng Zhenming,
University of Electronic Science and Technology of China, China

REVIEWED BY
Gulan Zhang,
Southwest Petroleum University, China
Bibo Yue,
Southwest Petroleum University, China
Cai Hanpeng,
University of Electronic Science and Technology of China, China

*CORRESPONDENCE
Shu Li,
shuli@jsu.edu.cn

SPECIALTY SECTION
This article was submitted to Environmental Informatics and Remote Sensing, a section of the journal Frontiers in Earth Science

RECEIVED 21 November 2022
ACCEPTED 13 February 2023
PUBLISHED 27 February 2023

CITATION
Ning J, Li S, Wei Z and Yang X (2023), Multichannel seismic impedance inversion based on Attention U-Net. *Front. Earth Sci.* 11:1104488. doi: 10.3389/feart.2023.1104488

Multichannel seismic impedance inversion based on Attention U-Net

Juan Ning[1], Shu Li[1,2]*, Zong Wei[1] and Xi Yang[1]

[1]School of Communication and Electronic Engineering, Jishou University, Jishou, China, [2]School of Biomedical Engineering, Guangzhou Medical University, Guangzhou, China

Recently, seismic inversion has made extensive use of supervised learning methods. The traditional deep learning inversion network can utilize the temporal correlation in the vertical direction. Still, it does not consider the spatial correlation in the horizontal direction of seismic data. Each seismic trace is inverted independently, which leads to noise and large geological variations in seismic data, thus leading to lateral discontinuity. Given this, the proposed method uses the spatial correlation of the seismic data in the horizontal direction. In the network training stage, several seismic traces centered on the well-side trace and the corresponding logging curve form a set of training sample pairs for training, to enhance the lateral continuity and anti-noise performance. Additionally, Attention U-Net is introduced in acoustic impedance inversion. Attention U-Net adds attention gate (AG) model to the skip connection between the encoding and decoding layers of the U-Net network, which can give different weights to different features, so the model can focus on the features related to the inversion task and avoid the influence of irrelevant data and noise during the inversion process. The performance of the proposed method is evaluated using the Marmousi2 model and the SEAM model and compared with other methods. The experimental results show that the proposed method has the advantages of high accuracy of acoustic impedance value inversion, good transverse continuity of inversion results, and strong anti-noise performance.

KEYWORDS

Attention U-Net, acoustic impedance inversion, spatial correlation, deep learning, multichannel inversion

1 Introduction

Seismic inversion can be defined as the process of obtaining subsurface model parameters, such as formation velocity, density, or impedance, from seismic data by comprehensively available geological and logging data (Treitel and Lines, 2001). For conventional seismic inversion methods, i.e., model-driven inversion methods, the mathematical theory is based on the convolution model or other mathematical and physical models. The convolution model is essentially a simplification and approximation of the seismic wave transmission process. The subsurface structure is usually very complex, and errors will inevitably arise when describing the wave propagation with the convolution model, which leads to inaccurate inversion results. On the other hand, in order to get a good inversion result, the model-driven method needs a better initial model and an accurate wavelet. In practical applications, it is usually challenging to obtain good initial models and accurate wavelets. In addition, problems such as limited

data bandwidth, data noise, and incomplete data coverage cause various troubles for model-driven inversion methods.

Unlike traditional model-driven seismic inversion, deep learning is a data-driven approach that can learn complex non-linear mappings between inputs and outputs based on training datasets, and the parameters are adjustable. Deep learning is a subset of machine learning that has recently made breakthroughs in image classification (Krizhevsky et al., 2017), object detection (Ren et al., 2015), image segmentation (Chen et al., 2017), image and video captioning (Vinyals et al., 2022), speech recognition (Graves et al., 2013), and machine translation (Cho et al., 2014). The success of deep learning in the fields of computer vision and natural language processing has led to widespread interest among scholars in data-driven intelligent seismic inversion methods. This class of methods does not require an initial model and does not require the estimation of seismic wavelets. Using the powerful learning ability of deep neural networks to establish non-linear mapping relationships between seismic data and parameters to be inverted has become a trendy research direction in the field of seismic inversion.

Currently, the application of deep learning methods in the field of seismic inversion is expanding, involving acoustic impedance inversion, pre-stack elastic and lithological parameter inversion, full waveform inversion, and so on. Recently, seismic inversion has made extensive use of supervised learning methods. Alfarraj and AlRegib (2018) used recurrent neural networks for petrophysical parameter estimation. Das et al. (2019) and Wu et al. (2020) trained the convolutional neural networks (CNNs) to invert seismic impedance using synthetic seismic records on the earth model constrained by petrophysical relationships. The results show that the type of sediment phase and source wavelet parameters used in the training dataset affect the inversion process of the network. Mustafa et al. (2019) used the temporal convolution network (TCN) to estimate the acoustic impedance. This method not only successfully captured the long-term trend but also preserved the local patterns while overcoming the gradient disappearance problem in the inversion of recurrent neural network (RNN) and the overfitting problem in convolutional neural networks. Du et al. (2019) proposed SeisInv-ResNet for pre-stack seismic inversion to obtain p-wave impedance, s-wave impedance, and other petrophysical parameters. Aleardi and Salusti (2021) proposed an elastic pre-stack seismic inversion method based on CNN.

Although the above inversion networks based on deep learning can well utilize the temporal correlation in the vertical direction, they do not consider the spatial correlation of seismic data in the horizontal direction, and each seismic trace is inverted independently. However, in subsurface seismic profiles, adjacent traces are highly correlated. The inversion method based on trace by trace does not exploit the spatial correlation in the horizontal direction, which may lead to poor horizontal continuity of inversion results. To improve the continuity, Wu et al. (2021) proposed a 2D network-based inversion method.

Traditional CNN networks take a long time to train and need a lot of labeled data. To address these drawbacks of classical CNN networks, Ronneberger et al. (2015) proposed the U-Net network in their study of biomedical image segmentation problems. Their research shows that U-Net can reduce the need for labeled data to a certain extent while improving training efficiency. Seismic inversion also faces the problem of a small number of labels (few logging data) and a very large amount of seismic data. In view of this, Cao et al. (2022) proposed an inversion network consisting of a U-Net combined with three fully connected networks and named it the UCNN, which was used to predict elastic parameters from pre-stack seismic data. To further reduce the reliance on labeled data, they use Sequential Gaussian Co-Simulation and Elastic Distortion algorithms to generate adequate and diverse pre-stack seismic inversion datasets. Similarly, Wang et al. (2020) proposed a closed-loop CNN structure with a U-Net network as the main body to make CNN less dependent on the amount of labeled data in seismic inversion. The proposed closed-loop CNN can simulate both seismic forward and inversion processes from the training dataset.

Given the excessive and repeated extraction and utilization of similar features for each cascaded CNN structure in U-Net, this results in a significant computational effort and network parameter scale. Oktay et al. (2018) proposed the Attention Gate (AG) model and integrated it into U-Net to obtain the Attention U-Net network. The AG model can implicitly learn to emphasize prominent features that are helpful for inversion while suppressing irrelevant regions in the input data. In addition, AG is easily integrated into standard CNN architectures such as U-Net, which can reduce the computational overhead while improving the sensitivity and prediction accuracy of the network.

In conclusion, this paper proposes a multichannel acoustic impedance inversion based on Attention U-Net to address the issues with conventional deep learning inversion networks, such as poor continuity of inversion results and susceptibility to noise due to the trace-by-trace inversion method. The horizontal spatial correlation is applied to the inversion network by mapping multiple seismic traces to one logging curve. Under the supervision of limited logging data, the inversion network is trained. The training samples consist of several seismic traces centered on the well-side traces and associated well-logging curves. The inversion network simultaneously performs the duties of predicting acoustic impedance and forwarding seismic data. This paper is structured as follows: In Section 2, the theory and network structure of Attention U-Net are briefly introduced, and then the architecture of the inversion network consisting of three modules and their specific internal parameter settings are presented. In Section 3, the experimental results of the inversion of two typical seismic models (the Marmousi2 model and the SEAM model) are presented, analyzed, and discussed. The experimental results are compared with other deep learning inversion methods, and the noise immunity of the inversion network is discussed in this paper. Finally, Section 4 concludes this paper.

2 Methods

2.1 Inversion framework

Geological structures are spatially correlated. The closer the distance, the stronger the correlation, and conversely, the weaker the correlation. The correlation of seismic data is reflected in the temporal correlation in the vertical direction of seismic traces

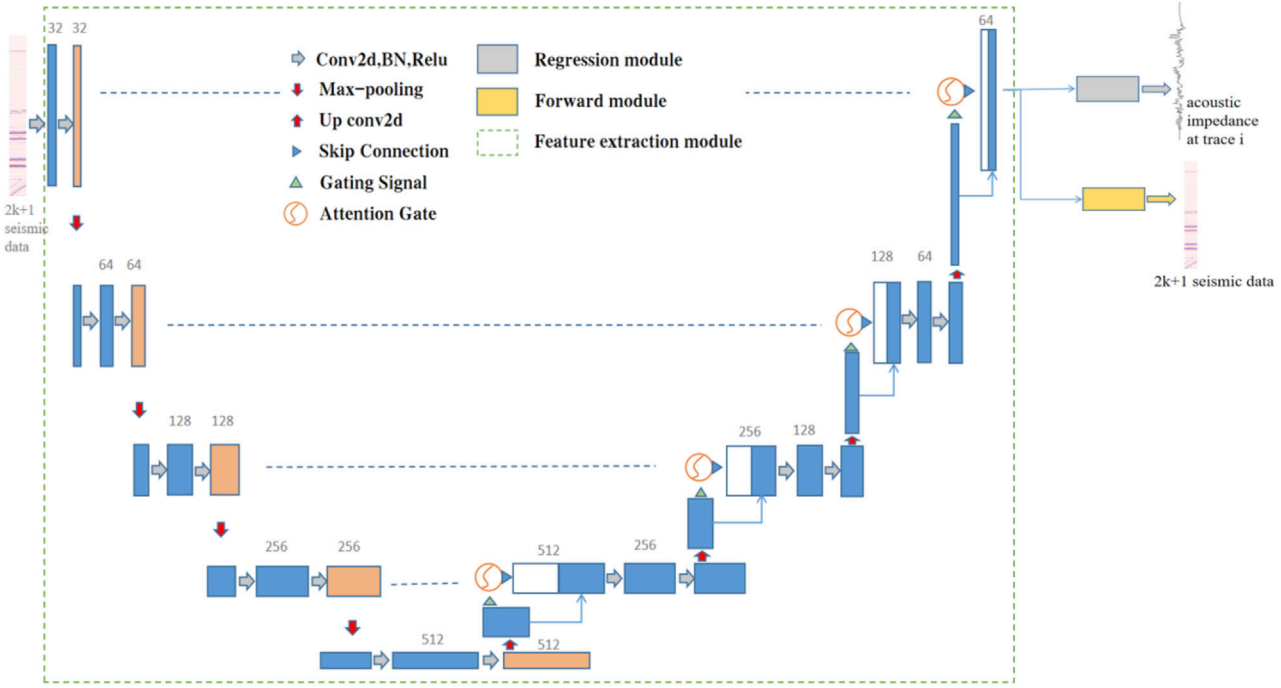


FIGURE 1
Structure of inversion network.

and the spatial correlation in the horizontal direction between the central trace and the adjacent traces. Based on the spatiotemporal characteristics of the seismic data, the inversion framework in Figure 1 is constructed using a supervised learning approach. The inversion framework shown in Figure 1 consists of three main modules: the feature extraction module, the regression module, and the forward module.

In the training phase, the input of the inversion network is the seismic data of the well-side trace and the 2 k nearby seismic data centered on it. The feature extraction module extracts the temporal and spatial features of the seismic data of the well-side trace and the 2 k nearby seismic traces by Attention U-Net. The regression module is used to map the data from the feature domain (spatiotemporal feature series) to the target domain (predicted acoustic impedance), while the forward module is used to map the data from the feature domain to the target domain (forward 2 k + 1 traces seismic data). Referring to the structure of the multi-task learning of Mustafa et al. (2021), the inverse network learns two tasks simultaneously: the predicted acoustic impedance data and forward seismic data. By sharing representations between the two tasks, especially if they are related to each other, we bias the network to learn more generalizable features.

2.2 Network model

2.2.1 Feature extraction module

The Attention U-Net is used as a feature extraction module to extract spatial and temporal features of seismic data. The input of the feature extraction module is the seismic data of the well-side trace and the nearby 2 k traces centered on it, and the output feature size is the same as the input size. Attention U-Net is improved by using U-Net as the base framework, as shown in Figure 1, adding AG at the jump connection between the encoding-decoding layers of the U-Net network, so that the originally up-sampled features are connected with the encoded layer AG-processed signal. By assigning different weights to different features, the model is better able to pay attention to the features relevant to the inversion task, which improves the sensitivity and prediction accuracy of the model.

Attention U-Net is divided into an encoding part and a decoding part, as shown in Figure 1. The encoding part of the Attention U-Net framework used in this paper contains four downsampling layers. The downsampling layer includes two consecutive convolutional blocks and a 2 × 1 max-pooling layer, and each convolutional block consists of a 3 × 3 two-dimensional convolutional layer (Conv2d), a batch normalization layer (BN) (Ioffe and Szegedy, 2022), and a rectified linear unit (ReLU) (Nair and Hinton, 2010) activation function. Batch normalization is used to accelerate the convergence of the network, and ReLU is used to enhance the non-linear approximation capability of the model. The decoding part corresponds to the encoding part, and the decoding part also contains four upsampling layers. Each upsampling layer consists of a 4 × 3 deconvolution layer, an AG model, and two convolution blocks.

The input of AG is the feature in the encoding part and the feature after deconvolution in the decoding part. The specific structure of AG is shown in Figure 2. The features extracted from the decoding part after deconvolution are used as the

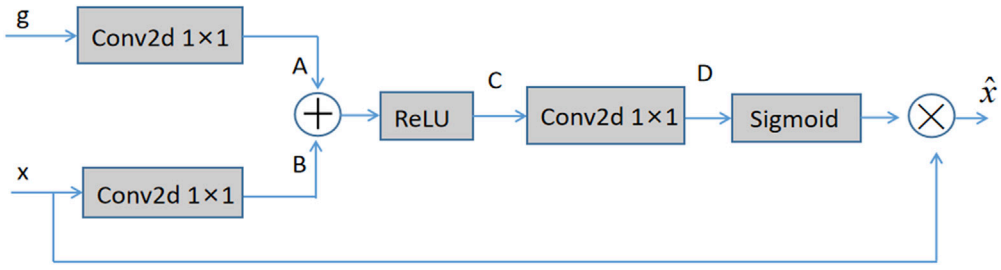


FIGURE 2
Attention gating model.

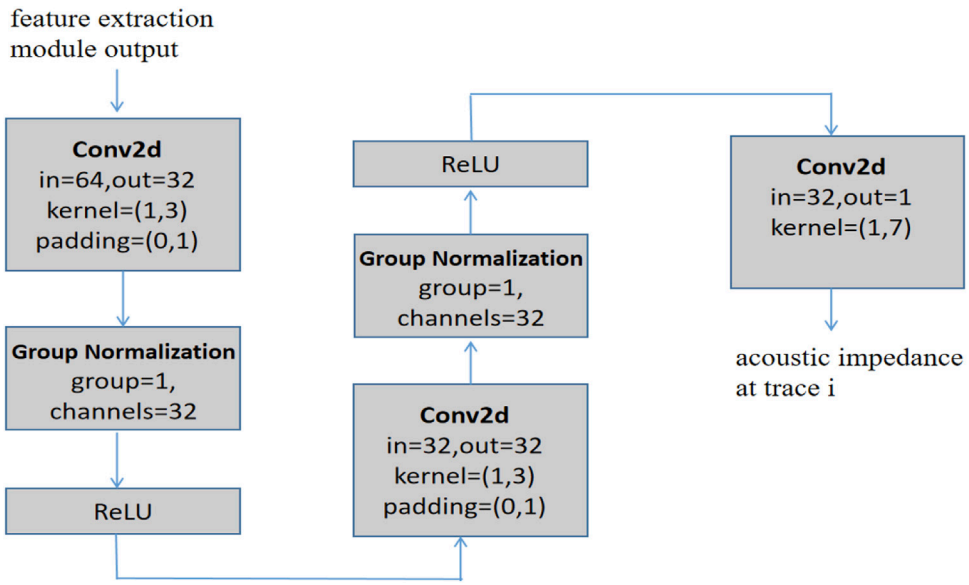


FIGURE 3
Block diagram of the regression module.

gating signal g, and the features from the matching layer's coding portion are used as x. The 1 × 1 convolution is done for g and x, and the two results A and B are added element by element to highlight the features. Then, the non-linear ability of the added result is increased by the ReLU activation function to obtain C, and the channel of C is reduced to 1 channel by a convolution operation. D is processed using a sigmoid activation function such that its value falls within the range of (0, 1), and the result is an attention weight that is the same size as the input feature and has one channel. Finally, the attention weight is multiplied by x.

2.2.2 Regression module

The regression module maps the output of the feature extraction module from the feature domain to the target domain. The regression module's structure, as shown in Figure 3, consists of two convolutional blocks and a 2D convolutional layer. Each convolutional block consists of a 2D convolutional layer, a group normalization layer, and the ReLU activation function. Group normalization groups the outputs of the convolutional layers and normalizes each group using the learned mean and standard deviation, which have been shown to reduce covariate bias in the learned features and speed up learning (Wu and He, 2012).

As shown in Figure 3, the input of the regression module is the output of the feature extraction module, and the output is the predicted acoustic impedance. Calculate the mean square error between the actual acoustic impedance and the output of the regression module. In other words, the mean square error between the predicted and the actual acoustic impedance data is calculated to update the learnable parameters in the feature extraction module and the regression module. The following Eq. 1 illustrates this:

$$l_1 = MSE(m_{i,t}, \hat{m}_{i,t}) \tag{1}$$

where $m_{i,t}$ is the actual acoustic impedance, $\hat{m}_{i,t}$ is the predicted acoustic impedance, and MSE is as in Eq. 4.

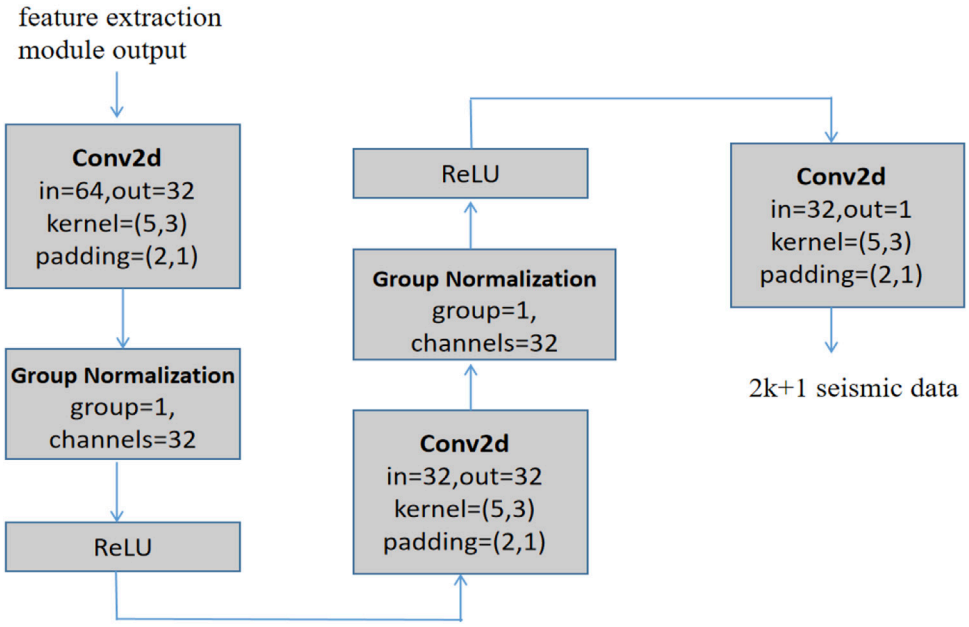


FIGURE 4
Block diagram of the forward module.

2.2.3 Forward module

The forward module maps the output of the feature extraction module from the feature domain to the target domain. As shown in Figure 4, the input of the forward module is the output of the feature extraction module, and the output is the predicted well-side trace and 2 k nearby seismic data. The structure of the forward module consists of two convolutional blocks plus a 2D convolutional layer. Each convolutional block consists of a 2D convolutional layer, a group normalization layer, and a ReLU activation function to achieve reconstruction.

Calculate the mean square error between the feature extraction module's input and the forward module's output. To put it another way, the mean square error between the 2 k + 1 seismic data in the well-side trace and nearby traces and the predicted 2 k + 1 seismic data is calculated in order to update the learnable parameters in the feature extraction module and the forward module. The following Eq. 2 illustrates this:

$$l_2 = MSE\left(x_{i,t}, \hat{x}_{i,t}\right) \tag{2}$$

where $x_{i,t}$ is the seismic data of the well-side and nearby 2 k traces, and $\hat{x}_{i,t}$ is the predicted 2 k + 1 seismic data.

2.3 Loss function

The loss of the entire inversion network is the mean square error between the predicted acoustic impedance data and the actual acoustic impedance (l_1), and the mean square error between the 2 k seismic data in and around the well-side traces and the predicted 2 k + 1 seismic data (l_2), with the total loss shown in Eq. 3:

$$l = \alpha l_1 + \beta l_2 \tag{3}$$

where α and β are weighting factors that control the effects of acoustic impedance losses and seismic losses, respectively.

2.4 Evaluation of inversion results

The inversion results are evaluated quantitatively by calculating the mean square error (MSE) and the coefficient of determination (R^2) of the actual and predicted acoustic impedance.

Mean Squared Error (MSE): MSE is the average of the squared sum of the errors of the corresponding points of the predicted data and the real data, and the smaller the value indicates that the predicted data fits better with the original data, which is defined as:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left\|y_i - \hat{y}_i\right\|_2^2 \tag{4}$$

where y_i, $\hat{y}_i$ denote the actual acoustic impedance and predicted impedance, respectively, and N is the number of data.

Determination Coefficient (R^2): R^2 is a measure of the goodness of fit between variables that takes into account the mean square error between predicted and actual data. Its range of values is [0, 1], and the larger the value, the better the fit between the variables, the more the independent variable explains the dependent variable, and the more the independent variable contributes to the overall variation. It is defined as:

$$R^2\left(y, \hat{y}\right) = 1 - \frac{\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{N}\left(y_i - \mu_y\right)^2} \tag{5}$$

where y_i, $\hat{y}_i$, and μ_y represent the actual acoustic impedance, predicted acoustic impedance, and the average of the actual

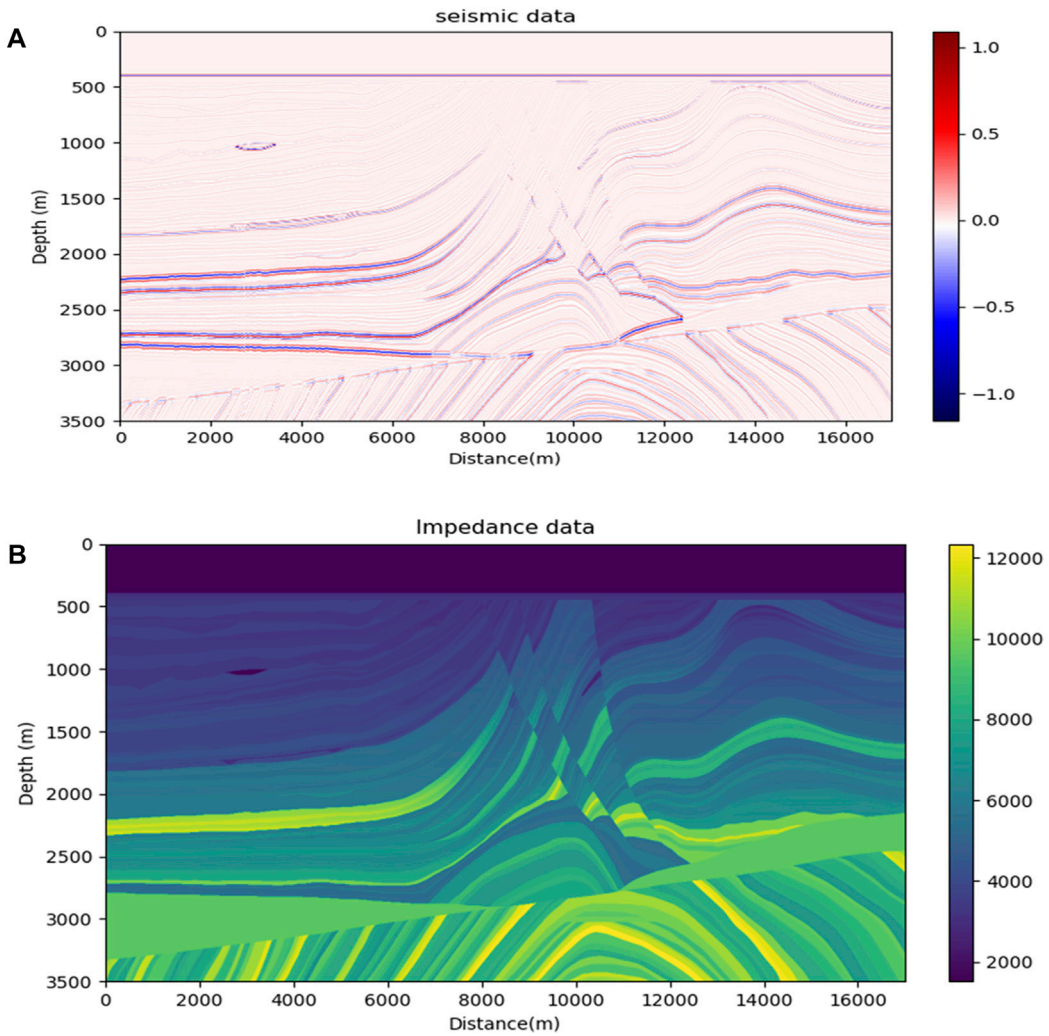


FIGURE 5
The Marmousi2 model. **(A)** Seismic data profile; **(B)** real acoustic impedance profile.

acoustic impedance, respectively. When R^2 is closer to 1, the stronger the correlation between the predicted and actual acoustic impedance is.

3 Experiments

The Marmousi2 and the SEAM models are widely used to validate the performance of deep learning inversion methods. This subsection will use these two models to validate the performance of the inversion network architecture proposed in this paper for acoustic impedance inversion.

3.1 Marmousi2 model inversion experiments

The Marmousi2 model is an extension of the original Marmousi model for amplitude variation with offset (AVO) analysis (Martin et al., 2002). The original Marmousi model has been widely used to validate inversion and imaging algorithms. The researchers added more complex structures representing hydrocarbon regions to the model and increased the number of strata, resulting in a new model, the Marmousi2 model, which has a width of 17 km and a depth of 3.5 km. The model is accompanied by synthetic seismic data, which are obtained by convolutional forward simulations of the model's reflection coefficients using seismic wavelets.

The acoustic impedance model was obtained by multiplying the density and p-velocity models of the Marmousi2 data. The seismic data and acoustic impedance profiles are shown in Figures 5A, B, with 2,721 traces and 688 sampling points per trace in the seismic profile and 2,721 traces and 688 sampling points per trace in the acoustic impedance profile. The colors in Figure 5A represent seismic amplitude values, and the colors in Figure 5B represent the acoustic impedance values. Twenty traces of acoustic impedance are uniformly extracted as pseudo-well data, and for each pseudo-logging curve, 2 k + 1 seismic traces centered on the well-side trace and with k as the radius will be obtained. This paper sets k to 3, and each pseudo-logging curve corresponds to 7 seismic traces with a

10.3389/feart.2023.1104488

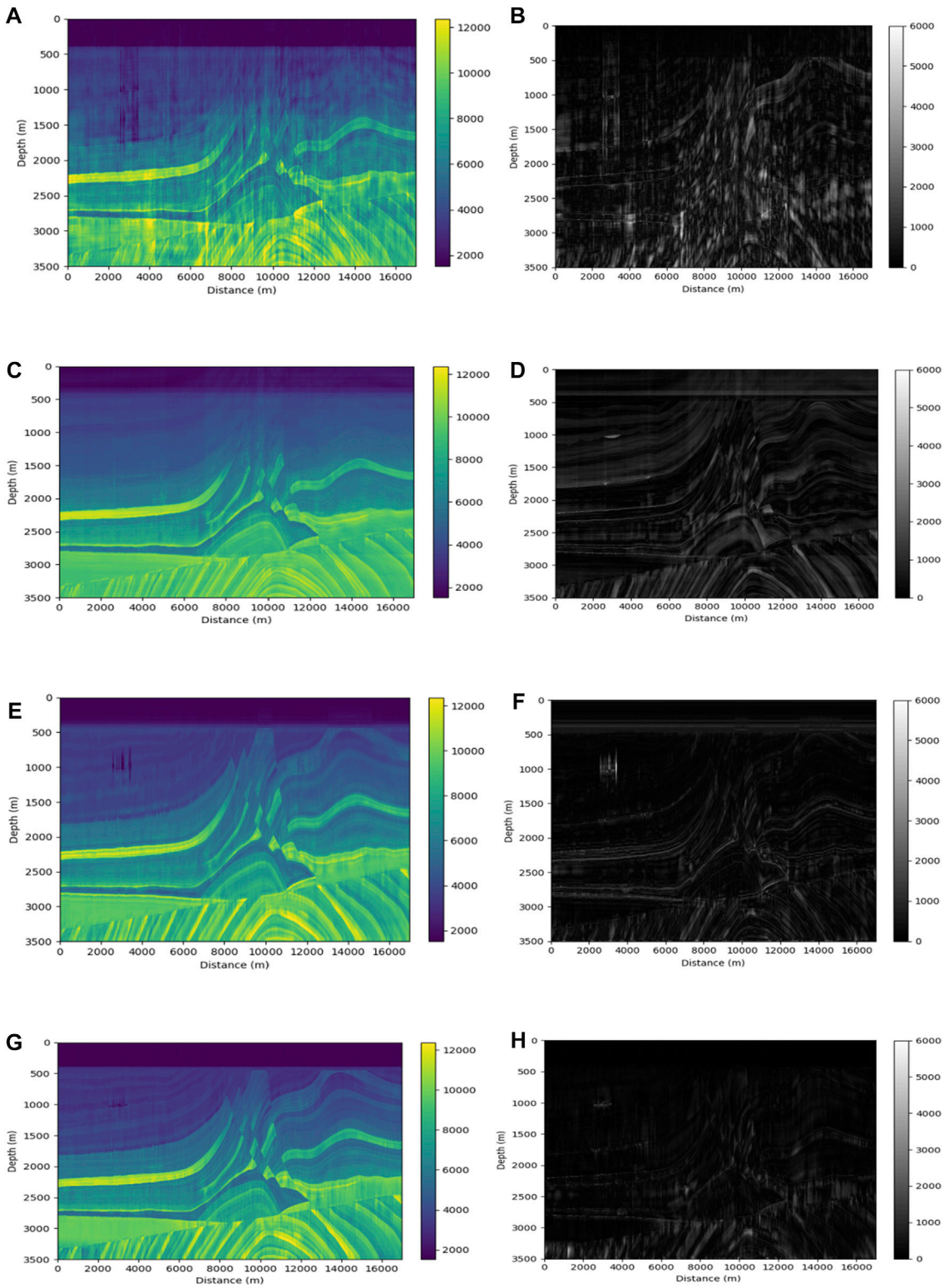


FIGURE 6
Acoustic impedance inversion profiles and residual profiles on the Marmousi2 model. **(A)** Inversion result of CNN method and its residual **(B)**. **(C)** Inversion result of TCN method and its residual **(D)**. **(E)** Inversion result of U-Net method and its residual **(F)**. **(G)** Inversion result of the method in this paper and its residual **(H)**.

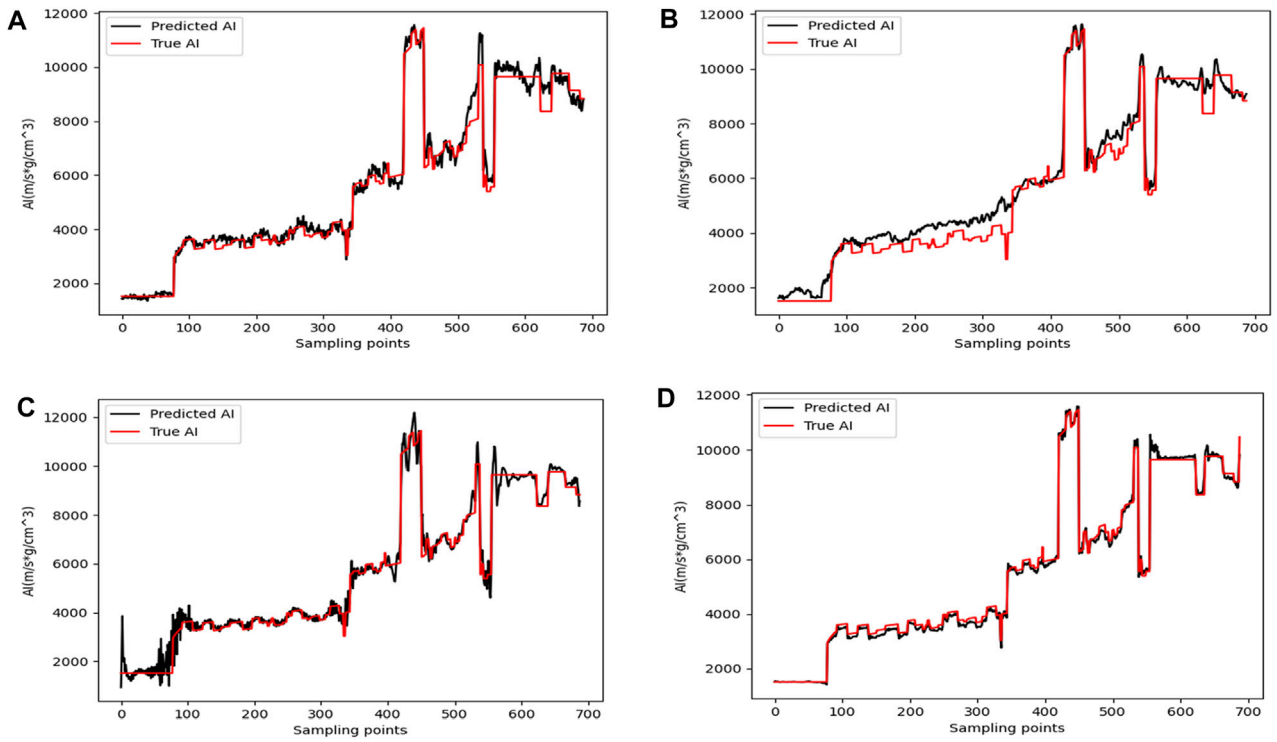


FIGURE 7
Acoustic impedance inversion results of trace no. 570. **(A)** Inversion result of the CNN method. **(B)** Inversion result of the TCN method. **(C)** Inversion result of the 1D U-Net method. **(D)** Inversion result of the method in this paper.

depth of 688 sampling points. The inverse network is trained using seismic data and pseudo-well data, the training epoch is set to 700, and the batch size is 20 for each iteration. In each training iteration, the weight coefficients in the loss function α and β are set to 1. The total training loss of the previously described inverse network is calculated and back-propagated through the network.

The Marmousi2 model has a complex stratigraphic structure and contains many different subsurface layered media models. The mean square error function is chosen as the loss function to measure the mean square error of the predicted and real acoustic impedance. ADAM is chosen as the optimizer, and ADAM adaptively sets the learning rate during training, with the initial learning rate set to 0.001. A weight decay of 0.0001 is chosen to limit the L_2 norm of the weights from becoming too large, reducing the risk of overfitting the network. The network's training is implemented in the PyTorch framework, and GPUs are applied to accelerate the computation. Finally, the trained inverse network is used for acoustic impedance inversion.

In order to prove the effectiveness of this paper's method, the inversion results of this paper's inversion method are compared with the inversion results of the commonly used deep learning inversion methods, including the inversion method based on CNN (Das et al., 2019), the inversion method based on 1D TCN (Mustafa et al., 2019), and the inversion method based on 1D U-Net. This 1D U-net model is constructed into the same network structure as the U-net proposed in this paper, but it lacks an attention mechanism. These inverse networks are set up with the same training conditions, training data, and hyperparameters. The inversion result of the method based on CNN is shown in Figure 6A, the inversion result of the method based on 1D TCN is shown in Figure 6C, the inversion result of the method based on 1D U-Net is shown in Figure 6E, and the inversion result of the method proposed in this paper is shown in Figure 6G. Figures 6B, D, F, H correspond to the residual difference between each network's inverse acoustic impedance and the real acoustic impedance.

As shown in Figure 6, the inversion results shown in Figures 6E, G have a higher similarity to the real model than the inversion results in Figures 6A, C. Moreover, Figure 6G has stronger horizontal continuity and weaker visible jitter in both horizontal and vertical directions for the inverse acoustic impedance profile than Figure 6E, the water layer at the top of the figure also clearly shows a relatively better inversion of Figure 6G. The partition interface and fault location in different strata are the main locations where the inversion results show errors, according to the residual profiles. In comparison to other figures in Figure 6, the inversion method in this paper can also invert the convolution structure in the model well, and the inversion results are more continuous and closer to the actual acoustic impedance, as well as more accurate in predicting the location of the faults. In most locations, the error is lower than that of other inversion methods. This is due to the effective use of the inversion network proposed in this paper for the spatial correlation of seismic data's horizontal direction.

In order to compare the details of the inversion results of different methods from the microscopic level, the representative Trace No. 570 (corresponding to the position around x = 3,565 m) and Trace No. 1400 (corresponding to the position around x = 8,747 m) are selected for inversion.

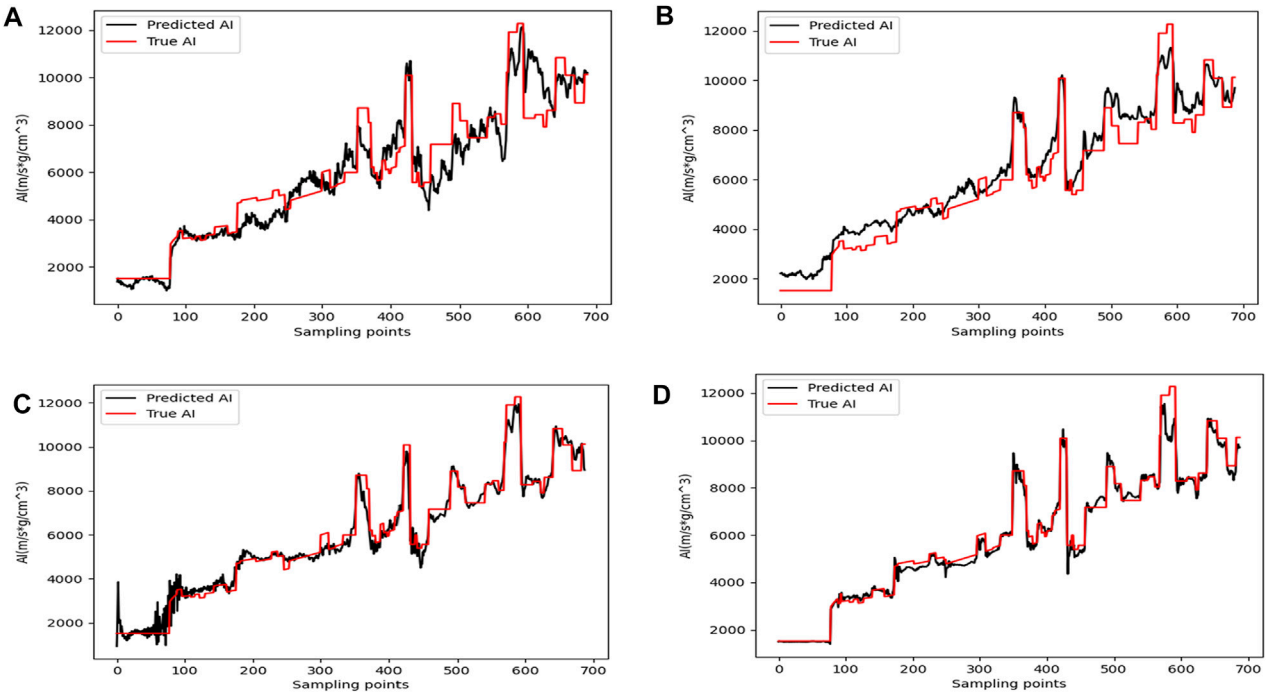


FIGURE 8
Acoustic impedance inversion results of trace no. 1400. **(A)** Inversion result of the CNN method. **(B)** Inversion result of the TCN method. **(C)** Inversion result of the 1D U-Net method. **(D)** Inversion result of the method in this paper.

TABLE 1 MSE, R^2 between inversion results and actual acoustic impedance.

Methods	MSE	R^2
CNN	0.0897	0.9090
TCN	0.0540	0.9452
U-Net	0.0343	0.9653
Attention U-Net	0.0199	0.9800

At these two locations, the acoustic impedance values obtained by four inversion methods were compared. Figures 7A–D shows the inversion results of Trace No. 570 using the conventional CNN inversion method, the 1D TCN inversion method, the 1D U-Net inversion method, and the method proposed in this paper, with the red and black lines representing the true impedance and acoustic impedance inversion results, respectively. Similar to the inversion results of the four networks mentioned above for all seismic traces, the inversion result of the method in this paper is relatively better. The inversion result in Figure 7A has a large inversion error at a large depth, the inversion result in Figure 7B is very different from the true value, and the inversion result in Figure 7C changes too drastically, whereas the difference between the inversion result and the actual acoustic impedance in Figure 7D is very small, with the two curves almost overlapping.

Figures 8A–D corresponds to the inversion results of the above four methods for Trace No. 1400 seismic trace, respectively, and the conclusions are consistent with Figure 7. The inversion results of Figures 8A, B in the figure deviate more from the true values. The inversion results of Figures 8C, D are in better agreement with the actual curves, but between sampling points 0 and 100, the inversion of Figure 8D is better, while the curve change of the inversion result of Figure 8C is too drastic. This further validates the performance of the inversion network proposed in this paper.

In order to objectively and quantitatively evaluate the reliability of the inversion results of the four methods, the coefficients R^2 and MSE are used as evaluation criteria. Table 1 shows the MSE and R^2 between the acoustic impedance inversion results of different methods in Figure 6 and the actual acoustic impedance.

Table 1 shows that this paper employs multichannel inversion, and the method of acoustic impedance inversion by Attention U-Net using spatial correlation performs best in terms of MSE and R^2, demonstrating the method's efficacy.

Gaussian noise of 4%, 8%, and 12% was added to the seismic data to test the adaptability of the method proposed in this paper to noise. Table 2 shows the quantitative evaluation of the inversion results obtained from the different inversion networks in Figure 6 under different noise conditions. As shown in Table 2, the performance of each method's inversion results decreases as noise increases relative to a noiseless environment, but the performance index of the method proposed in this paper decreases the least. For example, when the noise of the seismic data increases from 4% to 12%, the R^2 coefficients of the inversion results of CNN, TCN, U-Net, and the proposed method decreased by 7.43%, 2.85%, 4.83%, and 1.38%, respectively. Observing the changes in MSE data leads to a similar conclusion. It can be seen that the proposed method in this paper has better noise immunity performance compared with other methods.

TABLE 2 MSE, R^2 between inversion results and actual acoustic impedance under different noise conditions.

Indicator、SNR / methods	MSE			R^2		
	4%	8%	12%	4%	8%	12%
CNN	0.1146	0.1404	0.1787	0.8836	0.8571	0.8180
TCN	0.0622	0.0742	0.0886	0.9375	0.9253	0.9108
U-Net	0.0454	0.0729	0.0911	0.9540	0.9261	0.9079
Attention U-Net	0.0244	0.0285	0.0375	0.9751	0.9711	0.9616

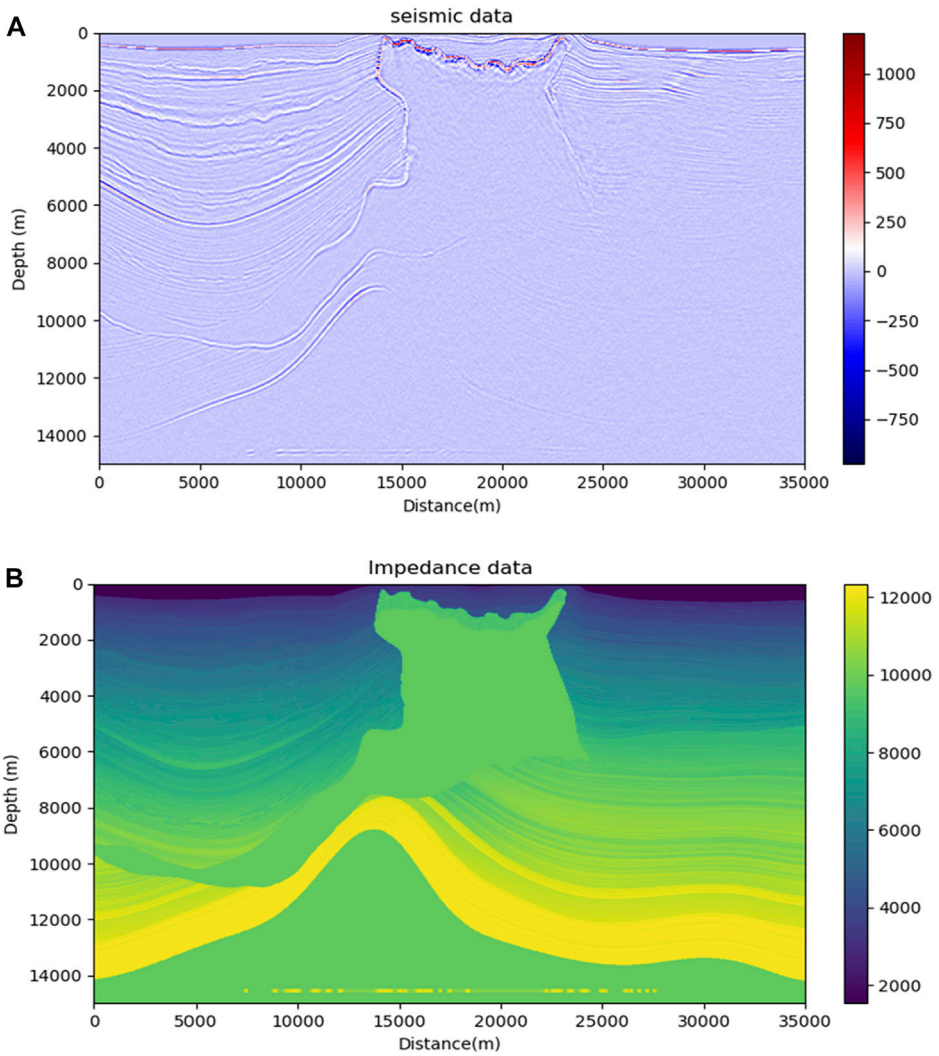


FIGURE 9
SEAM model. **(A)** Seismic data profile. **(B)** Real acoustic impedance profile.

3.2 SEAM model inversion experiments

To further verify the feasibility of the method, this paper conducts experiments with the SEAM model. The SEAM model is open source and also widely used for the validation of deep learning inversion methods (Mustafa et al., 2021). The SEAM model is a 3D seismic survey with very drastic lateral variations in density and longitudinal wave velocity, which is challenging for the inversion algorithm. The SEAM model is constructed based on basic rock properties, such as the volume of shale and sand. It follows the changing trend of shale porosity characteristics in the Gulf of Mexico, which is a better simulation of the actual geological conditions. The density of the SEAM model and the

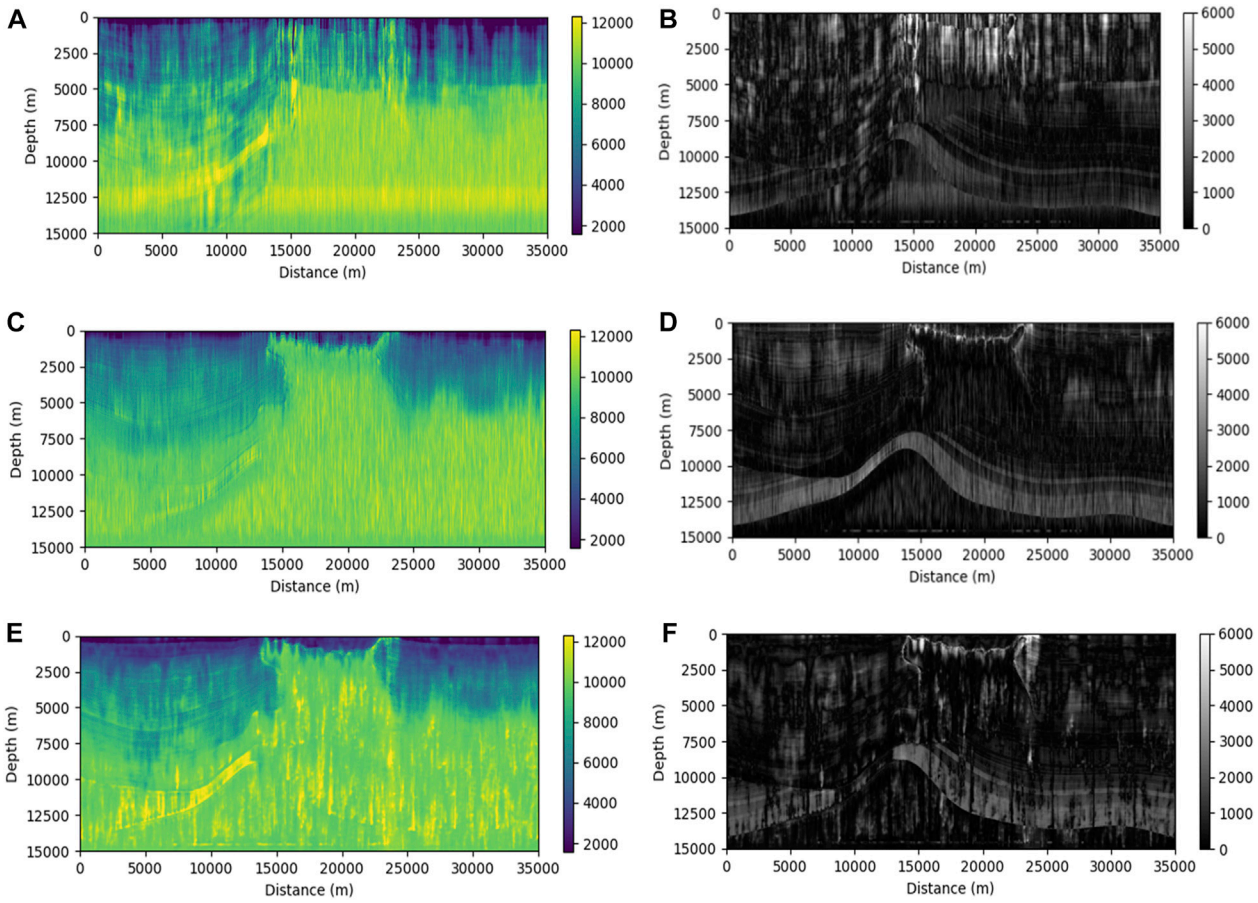


FIGURE 10
Acoustic impedance inversion profiles and residual profiles on the SEAM model. **(A)** Inversion result of CNN method and its residual **(B)**. **(C)** Inversion result of U-Net method and its residual **(D)**. **(E)** Inversion result of the method in this paper and its residual **(F)**.

longitudinal wave velocity model are multiplied to obtain the real acoustic impedance model. The seismic data and the real acoustic impedance profiles are shown in Figures 9A, B, respectively, with 501 traces and 688 sampling points per trace in the seismic profile and 501 traces and 688 sampling points per trace in the acoustic impedance profile. 12 traces of acoustic impedance are uniformly extracted from the acoustic impedance model as pseudo-well data, and k is also set to 3, so that each pseudo-logging curve corresponds to 7 seismic traces with a depth of 688 sampling points. The training epoch is set to 400, and the batch size is 12 for each iteration. The network is then trained in the same way as the Marmousi2 model, and the trained network is used to perform acoustic impedance inversion on all seismic traces.

The inversion results are shown in Figure 10. Figures 10A, C, E correspond to the results of the inversion based on the conventional CNN inversion method, the 1D U-Net inversion method, and the inversion of the proposed method in this paper, respectively. Figures 10B, D, F correspond to the residuals between the acoustic impedance and the real acoustic impedance inverted by each method, respectively. As can be seen from the figure, compared with Figures 10C, E has a better effect in displaying the stratigraphic interface in the left half of the depth range of 10,000 m to 14,000 m, and the strata are clearer. Some thin stratigraphic variations can be clearly observed in the upper left part of Figure 10E diagram between 5,000 and 9,000 m depth. For example, at 2,500 m depth in the real model, there is a thin arc-shaped stratigraphy that can be seen more clearly in Figure 10E, whereas it is difficult to see in Figures 10A, C, and Figure 10A does not outline the central uplifted area in the real model better. Although the method in this paper has some errors in the inversion of the SEAM model, the overall effect is better than the other two methods.

Trace No. 179 (corresponding to the vicinity of x = 12,500 m) was selected for the inversion experiment, and the acoustic impedance inversion results of the three inversion methods are shown in Figure 11. Figures 11A–C shows the inversion results of Trace No. 179 using the conventional CNN inversion method, the 1D U-Net inversion method, and the method proposed in this paper, with the red and black lines representing the true impedance and acoustic impedance inversion results, respectively. The proposed method has better inversion results compared with other methods. From Figure 11C, we can see that the inversion result obtained by the proposed method almost completely overlaps with the true impedance, while the inversion result of the CNN deviates from the true value, and the result obtained by the 1D U-Net inversion method also has large deviations, with a large deviation at a small depth.

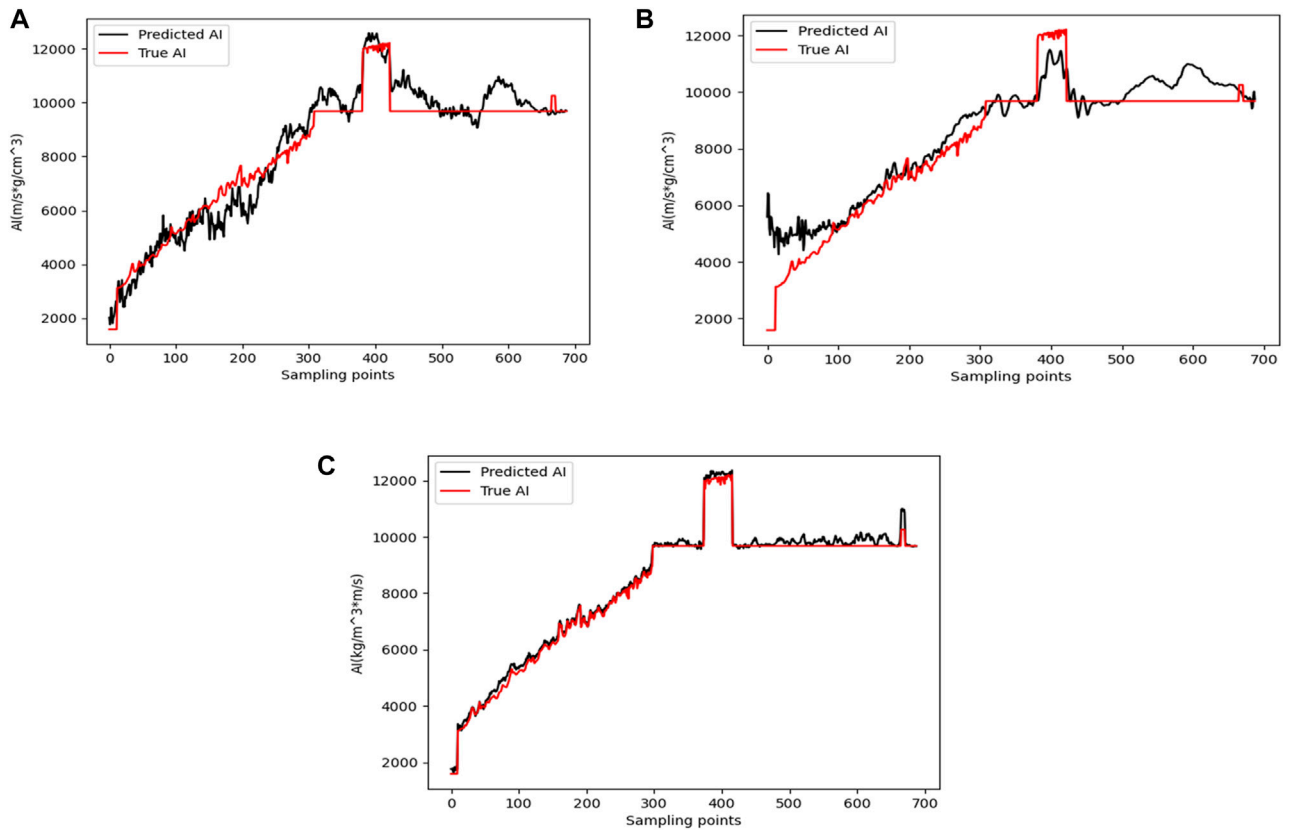


FIGURE 11
Acoustic impedance inversion results of trace no. 179. **(A)** Inversion result of the CNN method. **(B)** Inversion result of the 1D U-net method. **(C)** Inversion result of the method in this paper.

TABLE 3 MSE, R^2 between inversion results and actual acoustic impedance.

Methods	MSE	R^2
CNN	0.2659	0.5436
U-Net	0.1549	0.7991
Attention U-Net	0.1182	0.8250

To quantitatively evaluate the performance of the method proposed in this paper, the MSE and the R^2 between the acoustic impedance inversion results and the true acoustic impedance are calculated and presented in Table 3. The data are the MSE and R^2 between the acoustic impedance and the true acoustic impedance obtained by the inversion of different inversion methods in Figure 10. The data in the table show that the inversion result of the proposed method performs best in terms of MSE and R^2, which verifies the effectiveness of the method.

4 Conclusion

This paper proposes a multichannel seismic acoustic impedance inversion method based on the Attention U-Net network. Different from the conventional supervised learning inversion method, this inversion method applies the spatial correlation in the horizontal direction to the inversion network, and trains the network with 2 k + 1 seismic traces centered on the well-side trace and the corresponding logging curve to enhance the lateral continuity. In addition, the Attention U-Net network is used as a feature extraction module in the inversion network, and the attention gating model is added to the traditional U-Net-based inversion network. The AG is used to implicitly learn to suppress irrelevant regions in the input data while emphasizing salient features useful for inversion results, and it can be easily integrated into the standard CNN architecture to reduce computational overhead while improving the model's sensitivity and prediction accuracy. The method's performance is evaluated using the Marmousi2 and SEAM models, and it is also compared to several other commonly used deep learning inversion methods. The results show that the inversion results of the method proposed in this paper are more consistent with the actual acoustic impedance values, and the anti-noise performance is the best. In the SEAM model, where the lateral velocity and density vary drastically, the proposed method can better obtain the stratigraphic structure and details in the true model. These are attributed to the combined application of the attention gating model and methods such as multichannel simultaneous inversion.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://github.com/amustafa9/Geophysics-2021-Joint-learning-for-spatial-context-based-inversion/blob/master/data.zip.

Author contributions

JN responses for the experiments of the work and drafting papers. SL responses for the concept and design of the work and revisions to the paper. ZW and XY responses for important revisions to the papers.

Funding

This research was funded by the National Natural Science Foundation of China (Nos. 42164006 and 62161012), the Hunan Provincial Natural Science Foundation of China (No. 2022JJ30474), the China Postdoctoral Science Foundation (No. 2021M700682) and the Scientific Research Fund of Hunan Provincial Education Department (No. 21B0507).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Aleardi, M., and Salusti, A. (2021). Elastic prestack seismic inversion through discrete cosine transform reparameterization and convolutional neural networks. *Geophysics* 86 (1), R129–R146. doi:10.1190/geo2020-0313.1

Alfarraj, M., and AlRegib, G. (2018). "Petrophysical property estimation from seismic data using recurrent neural networks," in *SEG technical program expanded abstracts 2018* (United States: Society of Exploration Geophysicists). .

Cao, D., Su, Y., and Cui, R. (2022). Multi-parameter pre-stack seismic inversion based on deep learning with sparse reflection coefficient constraints. *J. Petroleum Sci. Eng.* 209, 109836. doi:10.1016/j.petrol.2021.109836

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. pattern analysis Mach. Intell.* 40 (4), 834–848. doi:10.1109/tpami.2017.2699184

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., et al. (2014). "Learning phrase representations using RNN encoder-decoder for statistical machine translation,". arXiv preprint arXiv:1406.1078.

Das, V., Pollack, A., Wollner, U., and Mukerji, T. (2019). Convolutional neural network for seismic impedance inversion. *Geophysics* 84 (6), R869–R880. doi:10.1190/geo2018-0838.1

Du, J., Liu, J., Zhang, G., Han, L., and Li, N. (2019). "Pre-stack seismic inversion using SeisInv-ResNet," in *SEG technical program expanded abstracts 2019* (United States: Society of Exploration Geophysicists).

Graves, A., Mohamed, A.-r., and Hinton, G. (2013). "Speech recognition with deep recurrent neural networks," in IEEE international conference on acoustics, speech and signal processing, Singapore, 22-27 May 2022.

Ioffe, S., and Szegedy, C. (2022). "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in International conference on machine learning: PMLR, USA, 25-27 July 2022, 448–456.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Commun. ACM* 60 (6), 84–90. doi:10.1145/3065386

Martin, G. S., Larsen, S., and Marfurt, K. (2002). *Marmousi-2: An updated model for the investigation of AVO in structurally complex areas*. Salt Lake City: SEG Annual Meeting: OnePetro.

Mustafa, A., Alfarraj, M., and AlRegib, G. (2019). "Estimation of acoustic impedance from seismic data using temporal convolutional network," in *SEG technical program expanded abstracts 2019* (United States: Society of Exploration Geophysicists). .

Mustafa, A., Alfarraj, M., and AlRegib, G. (2021). Joint learning for spatial context-based seismic inversion of multiple data sets for improved generalizability and robustness. *Geophysics* 86 (4), O37–O48. doi:10.1190/geo2020-0432.1

Nair, V., and Hinton, G. E. (2010). "Rectified linear units improve restricted Boltzmann machines," in Proceedings of the 27th International Conference on Machine Learning, Haifa, Israel, June 21-24, 2010.

Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., et al. (2018). "Attention u-net: Learning where to look for the pancreas,". arXiv preprint arXiv:1804.03999.

Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. neural Inf. Process. Syst.* 28, 1. doi:10.1109/TPAMI.2016.2577031

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-net: Convolutional networks for biomedical image segmentation," in International Conference on Medical image computing and computer-assisted intervention, Germany, October 8-12, 2023.

Treitel, S., and Lines, L. (2001). Past, present, and future of geophysical inversion—a new millennium analysis. *Geophysics* 66 (1), 21–24. doi:10.1190/1.1444898

Vinyals, O., Toshev, A., Bengio, S., and Erhan, D. (2022). ""Show and tell: A neural image caption generator," in Proceedings of the IEEE conference on computer vision and pattern recognition, San Juan, PR, USA, 17-19 June 1997.

Wang, Y., Ge, Q., Lu, W., and Yan, X. (2020). Well-logging constrained seismic inversion based on closed-loop convolutional neural network. *IEEE Trans. Geoscience Remote Sens.* 58 (8), 5564–5574. doi:10.1109/tgrs.2020.2967344

Wu, B., Meng, D., Wang, L., Liu, N., and Wang, Y. (2020). Seismic impedance inversion using fully convolutional residual network and transfer learning. *IEEE Geoscience Remote Sens. Lett.* 17 (12), 2140–2144. doi:10.1109/lgrs.2019.2963106

Wu, X., Yan, S., Bi, Z., Zhang, S., and Si, H. (2021). Deep learning for multidimensional seismi impedance inversion. *Geophysics* 86 (5), R735–R745. doi:10.1190/geo2020-0564.1

Wu, Y., and He, K. (2012). "Group normalization," in Proceedings of the European conference on computer vision (ECCV)), Florence Italy, October 7 - 13, 3–19.,

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 09 March 2023
DOI 10.3389/fmars.2023.1112955

OPEN ACCESS

EDITED BY
Peng Zhenming,
University of Electronic Science and Technology of China, China

REVIEWED BY
Tianlei Ma,
Zhengzhou University, China
Yun Zhang,
Harbin Institute of Technology, China
Zhengzhou Li,
Chongqing University, China

*CORRESPONDENCE
Vinay Mohan
vinay.mohan@colostate.edu
Steven J. Simske
steve.simske@colostate.edu

SPECIALTY SECTION
This article was submitted to Ocean Observation, a section of the journal Frontiers in Marine Science

RECEIVED 06 December 2022
ACCEPTED 13 February 2023
PUBLISHED 09 March 2023

CITATION
Mohan V and Simske SJ (2023) Cross-sensor vision system for maritime object detection. *Front. Mar. Sci.* 10:1112955. doi: 10.3389/fmars.2023.1112955

Cross-sensor vision system for maritime object detection

Vinay Mohan* and Steven J. Simske*

Department of Systems Engineering, Colorado State University, Fort Collins, CO, United States

Accurate and automated detection of maritime vessels present in aerial images is a considerable challenge. While significant progress has been made in recent years by adopting neural network architectures in detection and classification systems, these systems are usually designed specific to a sensor, dataset or location. In this paper, we present a system which uses multiple sensors and a convolutional neural network (CNN) architecture to test cross-sensor object detection resiliency. The system is composed of five main subsystems: Image Capture, Image Processing, Model Creation, Object-of-Interest Detection and System Evaluation. We show that the system has a high degree of cross-sensor vessel detection accuracy, paving the way for the design of similar systems which could prove robust across applications, sensors, ship types and ship sizes.

KEYWORDS

deep learning, vessel detection system, maritime vessel, optical satellite system, object detection, convolutional neural network, synthetic aperture radar

Introduction

From the advent of passenger ships in the late 19th century to container-revolutionized maritime transport in the 1970s, there has been increasing interest in monitoring, tracking and identifying vessels at sea. Before the first artificial earth satellite was placed into orbit in the mid 1950s, vessels were primarily tracked using either primitive cooperative systems such as inter-ship radio transmission or rudimentary non-cooperative systems such as coastal or on-board RADAR. Human interests at this time – revolving around safety & rescue, fishing and passenger transport - were largely satisfied by these systems.

More recently, effective understanding of the global maritime domain – or Maritime Domain Awareness (MDA) – has exploded in importance around the world with a significant number of commercial, defense and other government applications. There has been increasing attention given to exclusive economic zones (EEZ) and governance of a country's natural resources with state interests including maritime security, monitoring of marine traffic, illegal fishing, smuggling and maritime search & rescue. Commercial interests have expanded to include drilling and exploration of ocean floors, the management of fisheries, maritime piracy and cargo transportation. Private entities and NGOs have interests ranging from forecasting weather to the protection of ecology and sea health.

A number of these applications use knowledge of position and behavior of vessels as their cornerstone with MDA being enabled by information from land, sea, air and/or space systems and in some cases, vessel information repositories (Dekker et. al, 2013). These systems can broadly be classified into one of two types – Cooperative and Non Cooperative Systems - based on whether the system is employed by vessels to communicate information about themselves or whether they are observation systems which function independently of vessel cooperation (Table 1). Information captured usually includes the vessel type, cargo, position, velocity, route as well as other identifying and tracking data.

Cooperative systems are rarely used for comprehensive MDA. Most small (<300) ton vessels are not required to carry either an Automatic Identification System (AIS) or a Long-Range Identification and Tracking System (LRIT) while fishing vessels – regardless of size – are not required to carry a Vessel Monitoring System (VMS). Additionally, illegally operating vessels rarely carry or operate their systems accurately. Some vessels turn off their systems while others spoof their mandatory position reports. Operations such as search and rescue can't be carried out effectively if one is to rely solely on cooperative reports either. These reasons make non-cooperative systems among the most beneficial sources of information for a number of the MDA applications outlined above. In particular, Synthetic Aperture Radar (SAR) and Optical Imaging Satellite Systems have several advantages such as their remote access, global reach, frequency of information updates and the high amount of data they can collect and process.

While the last century saw incremental progress made in computer-generated detection and classification of objects in images, the creation of the first convolutional neural network in the 1980s and GPU-accelerated training in the 2000s enabled significant strides in machine learning approaches to detection, segmentation and localization of objects in images.

However, a number of distinct challenges exist which prevent the robust detection of vessels at sea. Sea surfaces can be complex, and variations in weather and vessel reflectivity can lead to a loss of system precision. Small, densely packed and blurry vessels – as well as vessels very close to land – have all proven challenging for detection systems. While traditional systems are inefficient and generally have lower accuracy, modern systems have been time-consuming to build and require large amounts of labeled data. Lastly, no single technique has proved robust across sensors, leading to piecemeal solutions for various sensors, datasets and locations.

This paper proposes a vision system which can provide robust target detection across disparate sensor types. The system is comprised of the following subsystems – Image Capture Subsystem, Image Processing Subsystem, Model Creation Subsystem, Object-of-Interest Detection Subsystem and System Evaluation Subsystem - and provides functionality for object detection using distinct independent data sources for model creation and object detection.

Related work

LandSat-1, launched in 1972, was the first civil optical satellite. Since then, hundreds of optical satellites with varying resolutions have been launched with many continuing to orbit our planet. Recent VHR additions like the WorldView and GeoEye series have expanded spectral and spatial resolutions while others like QuickBird and IKONOS have a higher radiometric resolution as well. An increasing number of optical satellite sensors now also provide more frequent coverage of Earth. At the turn of the century, there was a significant increase in the availability of commercial VHR sensor data and with it an explosion in the number of publications exploring the viability of maritime vessel detection using satellite systems.

Some of the earliest systems for maritime vessel detection used a number of pre-processing steps prior to target detection. Sea-land separation was considered crucial for accurate detection of vessels in harbors (Willhauck et al., 2005) as well as reducing the high number of false positives generated when vessel detection systems were applied to land (Corbane at. al, 2008). Consequently, coastline data was either incorporated from existing GIS data (Lavalle at al., 2011) or land masks were created from the images themselves (Dong et al, 2013). Similarly, key environmental effects – cloud coverage, waves and sunlight – were usually minimized using cloud masks (ESA, 2015), texture discrimination (Yang et. al, 2014) or Fourier transform algorithms (Buck et al., 2007; Jin and Zhang, 2015).

Vessel Detection and Classification methods ranged from simple geometrical feature detection (Lin et al, 2012; Heiselberg, 2016) to machine learning techniques. Prior to the advancements made in object detection systems which used neural networks, support vector machines (SVM) – a supervised classification

TABLE 1 Cooperative vs Non Cooperative Systems.

Cooperative Systems	Non-cooperative Systems		
Examples	Type	Examples	Spectrum
Automatic Identification System (AIS)	Visual Sighting	Shore or Ship-based Sighting	Varies
Long-Range Identification & Tracking System (LRIT)	Optical Camera	Optical Imaging Satellites	Panchromatic, Multispectral
Vessel Monitoring System (VMS)	Infrared Camera	FLIR, Weather Satellites	Thermal Infrared
	Radar	Real Radar , SLAR , SAR	Electromagnetic

model – dominated publications (Bi et al., 2010; Bi et al., 2012; Kumar and Selvi, 2011; Li and Itti, 2011; Xia et al., 2011; Guo and Zhu, 2012; Satyanarayana and Aparna, 2012; Song et al., 2014). Other classifiers for vessel detection include the Bayesian classifier (Antelo et al., 2009), random forest models (Johansson, 2011) and Fisher classification (Zhang et al., 2012).

More recently, neural network-based systems have taken the world of image recognition and object recognition by storm. AlexNet, a convolutional neural network architecture designed by Alex Krizhevsky in 2012 achieved a top-5 error of 15.3%, a full 10 percentage points lower than that of the runner up and paved the way for significant strides in image classification, segmentation and object detection.

In Ramani et al., 2019 the authors build a vessel detection system for real-time maritime applications. The system employs a Mask R-CNN architecture to segment and classify 30 images every 30 seconds.

In Gallego et al. (2018), results from a Convolutional Neural Network are passed to a k-NN model to improve detection performance.

In Zhang et al. (2019), pre-processing of satellite images is performed using a support vector machine framework following which variations of the Faster R-CNN neural network architecture are applied to measure each system's performance on different sizes and types of vessels. The authors are able to identify a framework which performs reasonably well for both offshore and inland vessel detection.

Chen et al., 2020 also used CNNs to create an end-to-end detection system capable of detecting both inshore and offshore ships with an accuracy >90%. Their detection speed was 72 fps and their system intentionally balanced accuracy against speed of detection on the SAR Ship Detection Dataset (SSDD).

Li et al. (2017) used a CNN architecture-based detection system on a custom dataset consisting of ships of various sizes as well a variety of environmental and sea conditions. Their paper established a higher precision with the custom framework than an equivalent Faster R-CNN system applied on the same dataset.

In contrast to common SAR and Optical Satellite Systems used in other publications, Yang et al. (2018) uses a remote sensing system which captured and segmented Google Earth images which were then used for vessel detection. The authors also used a custom neural network framework with a Feature Pyramid Network (FPN) to minimize false positives in images consisting of densely packed ships.

When we examine a collection of approaches used to build vessel detection systems, we observe a number of underlying trends:

- -Neural networks have gained popularity in recent publications due to the largely scripted/automated approach to building highly accurate detection systems.
- -Classification of vessels by vessel type has proven very challenging regardless of the type and resolution of the sensor(s) used.
- -Most publications have built and tested their systems using a homogenous dataset of images collected from either a single sensor or a set of sensors, thereby failing to establish robustness of their system across sensor types.

In this paper, we tackle the challenge of building a system robust enough to collect and use images from one sensor to detect objects in images collected from a second, disparate sensor. Such a system would be tunable, adaptable and re-purposable across applications. In the Object-of-Interest Detection subsystem, we evaluate several state-of-the-art algorithms as well as create a custom model architecture from scratch.

While we do not intend to recommend a winning algorithm to solve cross-sensor vessel detection, we show that the designed system has a high degree of cross-sensor vessel detection accuracy, paving the way for future research in tunable, adaptable and re-purposable systems which could prove robust across applications, sensors, ship types and ship sizes.

System overview

Image capture subsystem

The Image Capture Subsystem (Figure 1) uses two satellite sensor feeds along with two XML file feeds to obtain and provide data to the consequent subsystems. The XML files contain annotated image information for the corresponding image feeds.

The first input is an optical aerial image feed of maritime scenes on the visible spectrum. The images are sourced in the RGB color scheme, and can contain zero, one or multiple maritime vessels in varying weather and lighting conditions. The images contain scenes from different regions of the world including Africa, Europe and Asia and different water bodies including the Mediterranean Sea as well as the Atlantic and Pacific Oceans. While the images are of different sizes, the average image has a spatial resolution of 512 x 512 pixels.

The second input is a synthetic aperture radar (SAR) generated feed of maritime scenes (sea waves, shallow sea topography, coastal zones, maritime vessels etc.) with a spatial resolution between 1-500m. This feed provides images of 256 pixels in both range and azimuth, and the vessels in these images have distinct scales and backgrounds. A given image can contain a single vessel, multiple vessels or none.

For each image feed, annotations are provided in the Pascal VOC format. The Pascal VOC format is a common annotation format for images which stores annotations in the XML file format with a separate XML annotation file for each image. Optionally, bounding box information is included in the [x-top-left, y-top-left, x-bottom-right, y-bottom-right] format.

The image feeds are tagged with their source before being merged together into a single stream and sent to the Image Processing Subsystem. The two streams of XML annotations comprise the other outputs of this system.

Image processing subsystem

The inputs to the Image Processing Subsystem (Figure 2) are

- -a single stream of images tagged with their source, &

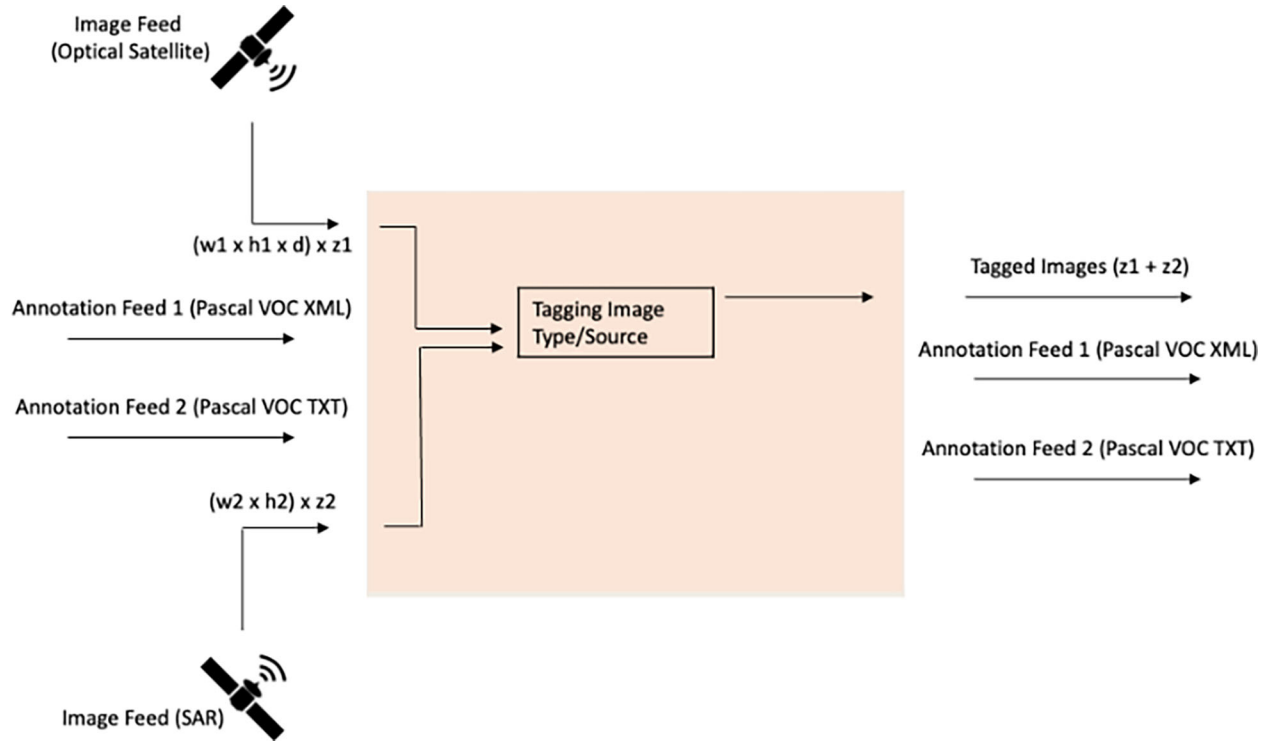


FIGURE 1
Image Capture Subsystem.

-two annotation feeds corresponding to the respective image streams.

First, the images are re-sized for uniformity across input streams and to match the dimensions of the input layer in the Model Creation Subsystem. The pixels in the image stream are then converted to the float datatype following which each image is normalized. Normalization scales the pixel values down from a range of (0,255) to a range of (0,1). Lastly, the image streams are split based on their source, annotations are appended and the output of the subsystem consists of two tagged and annotated image streams. Convolutional Neural Network Architectures like AlexNet and GoogleNet perform various image chopping and feature extraction steps which, in conjunction with pooling layers make them translation (and to a large degree, rotation) invariant model architectures.

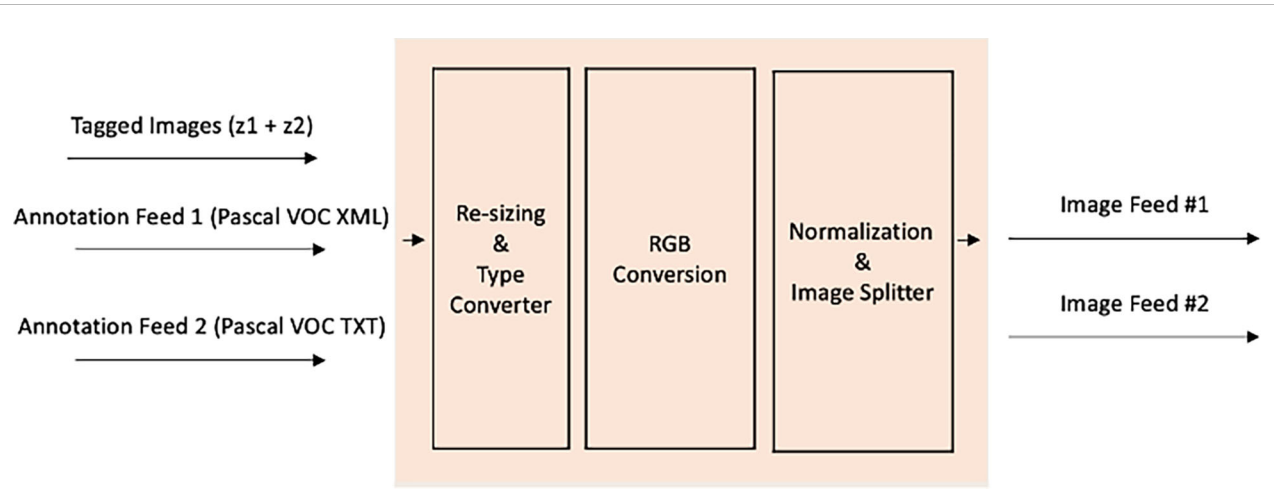


FIGURE 2
Image Processing Subsystem.

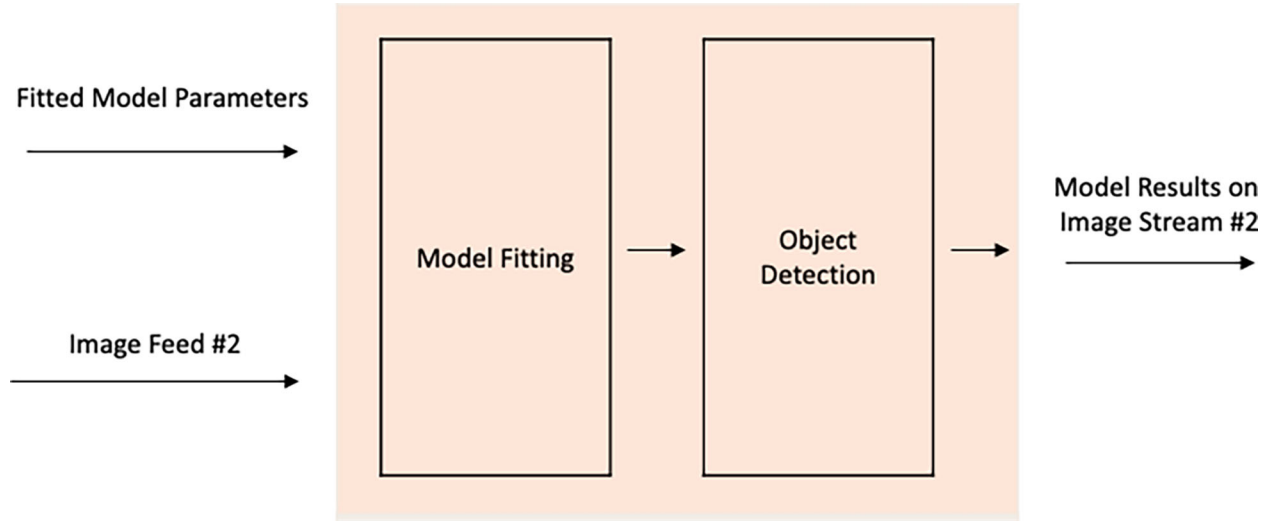


FIGURE 3
Model Creation Subsystem.

Model creation subsystem

The input to the Model Creation Subsystem (Figure 3) is a single annotated image feed. This feed is used to train a binary classification model to detect the presence of a vessel in an image using a combination of pre-defined model frameworks and hyperparameters. The models employed are (a) a custom convolutional neural network architecture, defined and trained from scratch, and (b) transfer learning and benchmarking using four common computer vision model architectures. For the latter, we re-define and fine-tune the last layers for our specific task while leaving the architecture and weights of other layers as is. Model parameters for each fitted model comprise the output of the Model Creation subsystem as well as each model's predictions on the input image feed.

Object-of-interest detection subsystem

The inputs to the Object-of-Interest Detection Subsystem (OOIDS) (Figure 4) are the fitted model parameters and the second image feed on which OOI Detection is to be performed. The fitted model parameters can either be the hyperparameters of the model – in which case the model will need to be re-fit on the original dataset – or a fit model, as we have assumed here. The model is applied ('scored') on the second image stream producing predictions indicating the presence or absence of maritime vessels. The output of this subsystem are the model results on the second image stream. As a reminder, this is an image feed the model itself has not been exposed to, and is an attempt to measure the model's power on a disparate and independent data source.

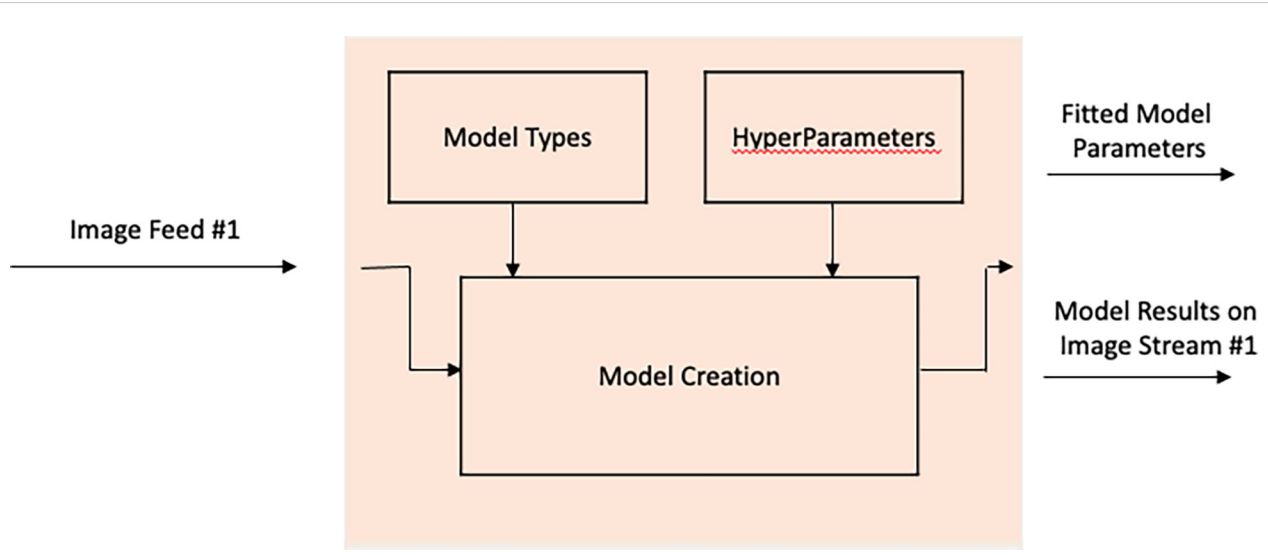


FIGURE 4
Object-of-Interest Detection Subsystem.

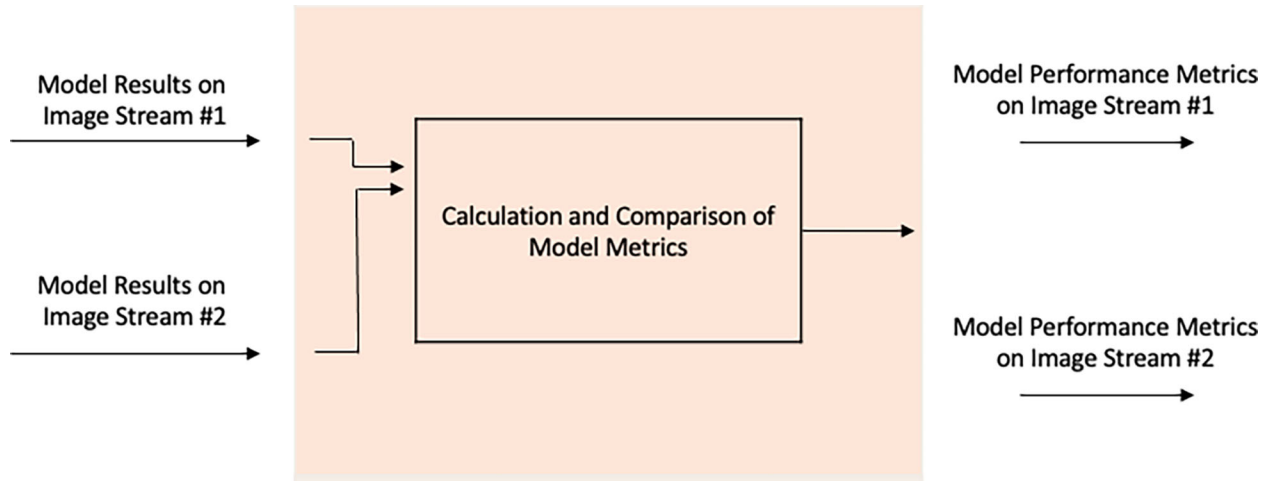


FIGURE 5
System Evaluation Subsystem.

System evaluation subsystem

The System Evaluation Subsystem (Figure 5) calculates and produce model metrics which measure the performance of the model on the dependent ('training') and independent ('test') data sources. The inputs to this subsystem are the model results on both the training (image feed #1) and test (image feed #2) datasets. Using these model results, model metrics such as accuracy, precision and recall can be calculated, dependent on the number and frequency of classes in each dataset. The metrics indicate the overall performance of the system at performing object detection using different sensors and sensor types, i.e SAR and Optical Satellite sensors.,,,,

Methodology

The System Block Diagram is shown in Figure 6.

To test system functionality and gauge performance, we use the MASATI (Maritime Satellite Imagery Dataset) and Sentinel datasets as inputs to the Image Capture Subsystem. The images contained in these datasets contain maritime scenes in the visible spectrum using optical aerial cameras and SAR-based radio waves, respectively. These datasets mimic and satisfy the earlier outlined assumptions regarding the two satellite image feeds (III.A) and are accompanied by annotations indicating the presence/absence of maritime vessels which are treated as ground truth in the subsequent model design and evaluation subsystems.

The datasets are tagged with their source name, re-sized to standardized dimensions, normalized and the bits converted to the float datatype. The Model Creation Subsystem uses the Keras Deep Learning API with the Tensorflow backend to fit four pre-defined convolutional neural network architectures and one custom architecture on the MASATI dataset. The MASATI dataset consists of 1027 (48%) images containing one or more maritime vessels and 1132 (52%) images with none. In addition to a custom model trained from scratch, the 4 pre-defined architectures include

-VGG-16, proposed by Karen Simonyan and Andrew Zisserman of Oxford University in 2014, the '16' in the

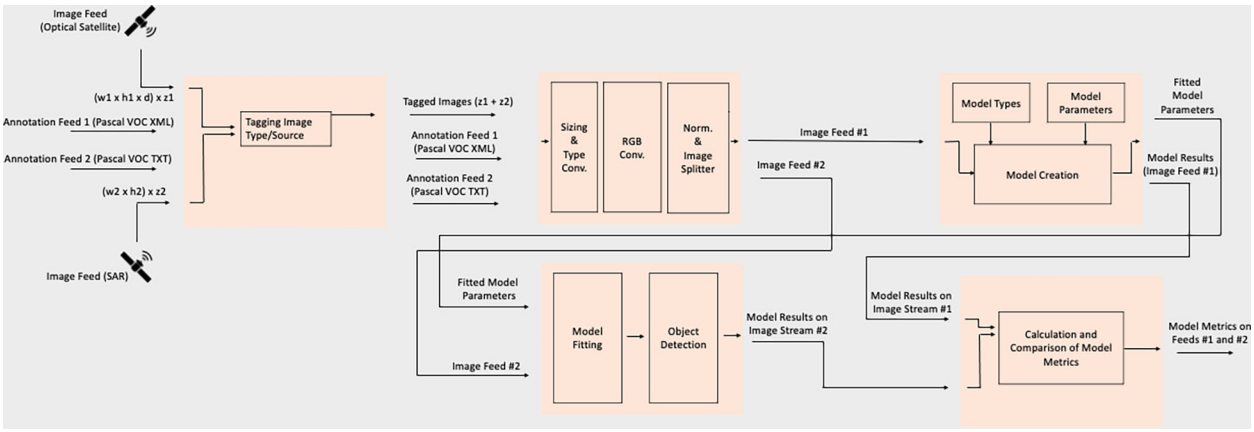


FIGURE 6
System Block Diagram.

name indicating the number of layers with weights (Simonyan and Zisserman, 2014).

-InceptionV3, originally a module for GoogleNet in 2015 (Szegedy et al., 2016)

-ResNet50, a variant of the ResNet Model consisting of 48 convolution layers and residual blocks, introduced in 2015, and

-XCeption, a deep convolutional neural network architecture involving Depthwise Seperable Convolutions, introduced by Francois Chollet in 2016

Since (a) one of the primary goals behind most maritime object detection systems is real-time processing, and (b) our primary goal is to develop a system capable of using data from one sensor to detect objects in incoming data from a second sensor, each of the 5 models is trained for <=5 epochs. There is no minimum stopping threshold or other optimization criteria since we want flexible models which aren't overfit or optimized on the MASATI dataset alone.

While the custom model is trained from scratch, each pre-defined architecture has the following changes:

-The input layer is altered to match the dimensions of the incoming data stream,

-The output layer is altered to a softmax function with two classes of interest, and

-while the fitted weights of most layers stay the same, the last five layers are re-trained for the purpose of optimizing detection of our classes of interest

Each model is trained using specific values of hyperparameters following which fitted parameter values are saved and transferred as outputs to the Object-of-interest Detection Subsystem. The OOI Detection Subsystem re-fits models and uses the fit models to predict the presence or absence of a maritime vessel in the second (SAR) input stream. The results of the 5 models on the SAR image stream is an input to the System Evaluation Subsystem which calculates, compares and displays metrics for the system's user(s).

Model Configuration and hyperparameters for each model are shown in Table 2.

Results and discussion

The System Evaluation Subsystem calculates each model's accuracy in detecting ships on the two datasets – MASATI and Sentinel, referred to as the training data and test data, respectively. The number of images in each dataset as well as the time to train and score each model on the respective datasets is also calculated. These results are shown in Table 3.

As we can see, the results are interesting and varied.

-While the custom model – trained from scratch – has a low accuracy, recall and F-Score, it has high precision and beats larger architectures like ResNet50 across the board when trained for only a few epochs.

-ResNet50, as we can see in Table 2 also has the highest number of parameters of all the architectures indicating that the extra learning potential of this network likely requires additional parameter tuning and in its current form results in overfitting on the training data.

-Most pre-trained models performed better than the custom model indicating that the extra layers, learning capacity and learned features in these models aided in our binary classification task, despite being designed for larger and more complex image classification and object detection tasks. In addition to the higher F-Scores, InceptionV3 and XCeption have much lower training times than the custom architecture.

-Despite the datasets being collected from different sensors and sensor types, many of the models are successfully able to

TABLE 2 Parameters for Models 1-5.

Architecture	*Custom*	*VGG-16*	*InceptionV3*	*ResNet50*	*Xception*
Train-Val Ratio	*N/A (Train = 1.0)*	*N/A (Train = 1.0)*	*N/A (Train = 1.0)*	*N/A (Train = 1.0)*	*N/A (Train = 1.0)*
Image Width	*200*	*200*	*200*	*200*	*200*
Image Height	*200*	*200*	*200*	*200*	*200*
Number of Images	*2,159*	*2,159*	*2,159*	*2,159*	*2,159*
Number of Classes	*2*	*2*	*2*	*2*	*2*
Number of Epochs	*5*	*5*	*1*	*1*	*1*
Learning Rate	*0.001*	*0.001*	*0.001*	*0.001*	*0.001*
Decay Factor	*0.9*	*0.9*	*0.9*	*0.9*	*0.9*
Optimizer	*RMSProp*	*RMSProp*	*RMSProp*	*RMSProp*	*RMSProp*
Total Parameters	*1,483,010*	*14,751,554*	*21,868,322*	*23,735,298*	*21,062,186*
Trainable Parameters	*1,483,010*	*2,396,674*	*65,538*	*200,706*	*204,802*

TABLE 3 System Evaluation Subsystem Results.

		Custom Model	VGG-16	InceptionV3	ResNet50	Xception
Training Data	*Number of Images*	2,159	2,159	2,159	2,159	2,159
	Time to Train	88 seconds	536 seconds	31 seconds	65 seconds	57 seconds
	Training Data - Overall Accuracy	0.92	0.99	0.91	0.56	0.96
Test Data	*Number of Images*	19,500	19,500	19,500	19,500	19,500
	Time to Score	49 seconds	918 seconds	208 seconds	576 seconds	482 seconds
	Precision (Ship)	0.73	0.83	0.97	0.46	0.72

identify maritime vessels in one using data solely from the other with both high precision and recall despite the unbalanced Sentinel dataset.

-While most modern systems built on underlying neural network architectures require sufficiently large (a) computing power, (b) time, and/or (c) data to perform well, a dataset of ~2K images was used to sufficiently capture between 62% - 94% of vessels in a dataset 10x as large (21,682 images) with training and testing times of <=10 minutes.

-While many systems require significant tuning and selection of hyperparameters to optimize object detection, limited fine-tuning resulted in respectable vessel detection results.

While we have not examined incorrect classification results further to discover potential underlying trends, future research in cross-sensor vessel detection could prove robustness across ship and sensor types with longer training times, other model architectures and/or further hyperparameter tuning.

We propose the following guidelines which similar studies could consider:

-Using multiple sensors for both system training and testing

-Verification of algorithms on varied maritime scenes

-Validation of accuracy and false detection rates across different ship sizes and difficult conditions

-Introduction of ship classification algorithms for specific applications

Given that earth observation is a rapidly growing field with an increasing availability of open data and new satellite technology, cross-sensor vessel detection systems would be more adept than traditional systems and could prove less cost-sensitive for new applications. Future research using small datasets and low system processing times may also lead to rapid detection rates, thereby aiding real-time maritime applications including safety, logistics and transportation.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Antelo, J., Ambrosio, G., Gonzalez, J., and Galindo, C. (2009). "Ship detection and recognition in high-resolution satellite images. In 2009 IEEE International Geoscience and Remote Sensing Symposium (IEEE) 4, 514–517. doi: 10.1109/IGARSS.2009.5417426

Copernicus Open access hub. Available at: https://scihub.copernicus.eu/.

Sentinel data access overview. Available at: https://sentinel.esa.int/web/sentinel/sentinel-data-access.

U.I for computer research – the MASATI dataset. Available at: https://www.iuii.ua.es/datasets/masati/.

Bi, F., Liu, F., and Gao, L. (2010). "A hierarchical salient-region based algorithm for ship detection in remote sensing images," in *Advances in neural network research and applications*. Eds. Z. Zeng and J. Wang (Berlin Heidelberg: Springer), 729–738.

Bi, F., Zhu, B., Gao, L., and Bian, M. (2012). A visual search inspired computational model for ship detection in optical satellite images. *IEEE Geosci. Remote Sens. Lett.* 9, 749–753. doi: 10.1109/LGRS.2011.2180695

Buck, H., Sharghi, E., Bromley, K., Guilas, C., and Chheng, T. (2007). "Ship detection and classification from overhead imagery," in *Applications of Digital Image Processing XXX* (SPIE) 6696, 522-536. doi: 10.1117/12.754019

Chen, Y., Duan, T., Wang, C., Zhang, Y., and Huang, Mo (2020). End-to-End ship detection in SAR images for complex scenes based on deep CNNs. *J. Sensors* 2021, 8893182, 19. doi: 10.1155/2021/8893182

Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1251–1258.

Corbane, C., Pecoul, E., Demagistri, L., and Petit, M. (2008). "Fully automated procedure for ship detection using optical satellite imagery," in R.J. Frouin, S. Andrefouet, H. Kawamura, M.J. Lynch, D. Pan, T. Platt, et al *Remote Sensing of Inland, Coastal, and Oceanic Waters* (SPIE) 7150, 13. doi: 10.1117/12.805097

Dekker, R. J., Bouma, H., Breejen, E., den, Broek, A. C., van den, Hanckmann, P., Hogervorst, M. A., et al. (2013). "Maritime situation awareness capabilities from satellite and terrestrial sensor systems," in *Proc. Maritime Systems and Technologies MAST Europe*.

Dong, L., Yali, L., Fei, H., and Shengjin, W. (2013). "Object detection in image with complex background," in *3rd International Conference on Multimedia Technology (ICMT-13)* (Atlantis Press.), 471-478. doi: 10.2991/icmt-13.2013.58

ESA. (2015). Sentinel-2 User Handbook. 2015. Available at: https://scholar.google.com/scholar_lookup?title=Sentinel-2%20User%20Handbook&author=ESA&publication_year=2015.

Gallego, A.-J., Pertusa, A., and Gil, P. (2018). Automatic ship classification from optical aerial images with convolutional neural network. *Remote Sens.* 10 (4). doi: 10.3390/rs10040511

Guo, J., and Zhu, C. R. (2012). A novel method of ship detection from spaceborne optical image based on spatial pyramid matching. *Appl. Mech. Mater.* 190–191, 1099–1103. doi: 10.4028/www.scientific.net/AMM.190-191.1099

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778. doi: 10.1109/CVPR.2016.90

Heiselberg, H. (2016). A direct and fast methodology for ship recognition in sentinel-2 multispectral imagery. *Remote Sens* 8. doi: 10.3390/rs8121033

Jin, T., and Zhang, J. (2015). "Ship detection from high-resolution imagery based on land masking and cloud filtering," in *Seventh International Conference on Graphic and Image Processing (ICGIP 2015)* (SPIE) 9817, pp. 268-273. doi: 10.1117/12.2228219

Johansson, P. (2011). "Small vessel detection in high quality optical satellite imagery," in *Tech. report chalmers university of technology sweden. JR* (European Commission - Joint Research Centre, Ispra, Italy: Integrated Maritime Policy for the EU, Working Document III on Maritime Surveillance Systems).

Kumar, S. S., and Selvi, M. U. (2011). Sea Object detection using colour and texture classification. *Int. J. Comput. Appl. Eng.* 1, 59–63.

Lavalle, C., Rocha Gomes, C., Baranzelli, C., and Batista e Silva, F. (2011). *Coastal zones: Policy alternatives impacts on European coastal zones 2000–2050* (Joint Research Centre, Institute for Environment and Sustainability: European Commission).

Li, Z., and Itti, L. (2011). Saliency and gist features for target detection in satellite images. *IEEE Trans. Image Process.* 20, 2017–2029. doi: 10.1109/TIP.2010.2099128

Li, J., Qu, C., and Shao, J. (2017). Ship detection in SAR images based on an improved faster R-CNN. In *2017 SAR in Big Data Era: Models, Methods and Applications (BIGSARDATA)* (IEEE), 1–6. doi: 10.1109/BIGSARDATA.2017.8124934

Lin, J., Yang, X., Xiao, S., Yu, Y., and Jia, C. (2012). "A line segment based inshore ship detection method," in *Future control and automation*. Ed. W. Deng (Berlin Heidelberg: Springer-Verlag), 261–269.

Ramani, S., Prabakaran, N., Kannadasan, R., and Rajkumar, S. (2019). Real time detection and segmentation of ships in satellite images. *Int. J. Sci. Technol. Res.* 8 (12).

Satyanarayana, M. S., and Aparna, G. (2012). A method of ship detection from spaceborne optical image. *Int. J. Adv. Comput. Math. Sci.* 3, 535–540.

Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv* 1409, 1556.

Song, Z., Sui, H., and Wang, Y. (2014). "Automatic ship detection for optical satellite images based on visual attention model and LBP," in *2014 IEEE Workshop on Electronics, Computer and Applications*. 722–725. doi: 10.1109/IWECA.2014.6845723

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. B. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2818-2826. doi: 10.1109/CVPR.2016.308

Willhauck, G., Caliz, J. J., Hoffmann, C., Lingenfelder, I., and Heynen, M. (2005) Object oriented ship detection from VHR satellite images (Accessed 6th Geomatic Week Conference).

Xia, Y., Wan, S., and Yue, L. (2011). "A novel algorithm for ship detection based on dynamic fusion model of multi-feature and support vector machine," in *6th International Conference on Image and Graphics (ICIG)*. 521–526. doi: 10.1109/ICIG.2011.147

Yang, G., Li, B., Ji, S., Gao, F., and Xu, Q. (2014). Ship detection from optical satellite images based on sea surface analysis. *IEEE Geosci. Remote Sens. Lett.* 11, 641–645. doi: 10.1109/LGRS.2013.2273552

Yang, X., Sun, H., Fu, K., Yang, J., Sun, X., Yan, M., et al. (2018). Automatic ship detection in remote sensing images from Google earth of complex scenes based on multiscale rotation dense feature pyramid networks. *Remote Sens.* 10, 132. doi: 10.3390/rs10010132

Yang, X., Sun, H., Fu, K., Yang, J., Sun, X., Yan, M., et al. (2011). A sea-land segmentation scheme based on statistical model of sea. *4th International Congress on Image and Signal Processing. CISP* 2011, 1155–1159. doi: 10.1109/CISP.2011.6100503

Zhang, W., Bian, C., Zhao, X., and Hou, Q. (2012). "Ship target segmentation and detection in complex optical remote sensing image based on component tree characteristics discrimination," in *Proc. SPIE 8558, Optoelectronic Imaging and Multimedia Technology II*, 85582F, pp. 9. doi: 10.1117/12.2000688

Zhang, S., Wu, R., Xu, K., Wang, J., and Sun, W. (2019). R-CNN-Based ship detection from high resolution remote sensing imagery. *Remote Sens.* 11 (6). doi: 10.3390/rs11060631

frontiers | Frontiers in Earth Science

TYPE Original Research
PUBLISHED 10 March 2023
DOI 10.3389/feart.2023.1073211

OPEN ACCESS

EDITED BY
Peng Zhenming,
University of Electronic Science and Technology of China, China

REVIEWED BY
Fengchao Xiong,
Nanjing University of Science and Technology, China
Peng Zhenming,
University of Electronic Science and Technology of China, China
Shu Li,
Jishou University, China

*CORRESPONDENCE
Gerard Schuster,
j.schuster@utah.edu

SPECIALTY SECTION
This article was submitted to Environmental Informatics and Remote Sensing, a section of the journal Frontiers in Earth Science

RECEIVED 18 October 2022
ACCEPTED 09 February 2023
PUBLISHED 10 March 2023

CITATION
Shi Y, Ballesio M, Johansen K, Trentman D, Huang Y, McCabe MF, Bruhn R and Schuster G (2023), Semi-universal geo-crack detection by machine learning.
Front. Earth Sci. 11:1073211.
doi: 10.3389/feart.2023.1073211

Semi-universal geo-crack detection by machine learning

Yongxiang Shi[1,2], Marco Ballesio[2], Kasper Johansen[2], Daniel Trentman[2], Yunsong Huang[2], Matthew F. McCabe[2], Ronald Bruhn[3] and Gerard Schuster[2,3]*

[1]School of Earth and Space Science, Peking University, Beijing, China, [2]King Abdullah University of Science and Technology, Thuwal, Makkah, Saudi Arabia, [3]Geology and Geophysics Department, University of Utah, Salt LakeCity, UT, United States

Introduction: Cracks are a key feature that determines the structural integrity of rocks, and their angular distribution can be used to determine the local or regional stress patterns. The temporal growth of cracks can be monitored in order to predict impending failures of materials or structures such as a weakened dam. Thus, cracks and their spatial-temporal distributions should be automatically monitored for assessing their structural integrity, the associated stress patterns and their potential for failure.

Method: We show that the U-Net convolutional neural network, semantic segmentation and transfer learning can be used to accurately detect cracks in drone photos of sedimentary massifs. In this case, the crack distributions are used to assess the safest areas for tunnel excavation. Compared to the coarse performance of ridge detection, the U-Net accuracy in identifying cracks in images can be as high as 98% when evaluated against human identification, which is sufficient for assessing the general crack properties of the rock faces for the engineering project.

Result: Based on approximately 100 h of manual cracks labeling in 127 drone photos and 20 h of network training, the U-Net was able to successfully detect cracks in 23,845 high-resolution photographs in less than 22 h using two Nvidia V100 GPUs. Meanwhile, the network was able to detect more than 80% of the observable cracks of a volcanic outcrop in Idaho without additional training. With a modest amount of extra labeling on photos of the volcanic outcrop and transfer training, we found that the accuracy significantly improved. The surprising outcome of this research is that the U-Net crack detector laboriously trained on photos of sedimentary rocks can also be effectively applied to photos of volcanic rock faces. This can be important for real-time assessment of geological hazards and lithology information for dam inspection and planetary exploration by autonomous vehicles. For another application, we accurately detected fractures and faults with a scale of tens of kilometers from Martian photographs.

Conclusions: In summary, our methodology of using CNN with transfer training suggests that it can be used as a semi-universal detector of cracks in across a range of diverse geological settings.

KEYWORDS

geo-crack detection, rock cracks, U-Net, convolutional neural network, transfer learning, machine learning

1 Introduction

1.1 Crack detection and deep learning

All solids weaken over time and develop a reduction in their mechanical strength. A sign of this weakness is the development of cracks, a quasi-linear physical separation of material on the surface and in the interior of a solid. For siting buildings on a rock foundation (Wyllie, 1999), assessing dam hazards (Herbert, 2011), avoiding drilling hazards or for mining excavation (Dyskin and Germanovich, 1993), it is critical to assess the density and distribution of fractures in the rock mass. To aid in this task, thousands of aerial photos of a rock area can easily be obtained by cameras mounted on Unmanned Aerial Vehicles (UAVs). As introduced in (Bemis et al., 2014) and (Vasuki et al., 2014), UAVs can be programmed to photograph an area of interest to a centimeter accuracy, no matter how large the rock mass. Then, methods from photogrammetry can create a mosaic of these images and merge them with real topograhy (Mikhail et al., 2001).

In tradition, surface cracks were labeled by experts and their distribution was cataloged (Sanderson et al., 2019) to give engineers an estimate of the integrity and stability of the rock mass (Aydan et al., 2014). However, if there are thousands of images then the manual interpretation of cracks is both time consuming and error prone because of variable lighting, shadows, non-crack erosional features, rock spall, and complex rock surfaces. In addition, the accuracy of manual crack interpretation depends on the expertise of the interpreter, and the interpretation criteria (Hillier et al., 2015); (Sander et al., 1997). Therefore, there is a growing demand to develop tools that can automatically detect and catalogue cracks in a more efficient and accurate way.

Because the shape of cracks is long and narrow, many crack detection methods are based on edge-detection algorithms. For example (Pereira and Pereira, 2015), successfully applied edge-detection algorithms to UAV photographs to detect cracks in building structures. Their algorithm applied Sobel operator (Sobel, 1990) and particle filters (Thrun, 2002) for detecting cracks in building facades. For fractured outcrop images, Prabhakaran et al. (2019) used the complex shearlet transform to automatically extract fracture ridge realizations from images. After getting the features of ridges, post-processing image analysis algorithms were then used to vectorize the fracture traces in an automated manner. The resulting fracture detection maps are generally consistent with the human-labeled fractures in photos taken over a geological outcrop in Parmelan, France.

One problem with an image processing approach is that it cannot easily distinguish abrupt changes in a photo's intensity caused by, for example, a skyline in a photo (Mohan and Poobal, 2018). This type of noise is easily distinguished by an interpreter, but manual detection is labor intensive when thousands of photos must be analyzed. To incorporate humans into the automated decision process, convolutional neural network (CNN) architectures were introduced to detect cracks. In the earliest applications, a variety of CNN networks including AlexNet (Dorafshan et al., 2018) and VGG16 (WilsonLeal da Silva and de Lucena, 2018) were used to detect cracks in concrete structures (Kim and Cho, 2018); (Cao and Anh, 2019). Their limitation is that they are inefficient for high-resolution semantic segmentation, which is the task of classifying all pixels in the input image. Such CNN methods cannot localize cracks at the pixel level. In the work of (Cha et al., 2017), a method based on CNN can only locate cracks in a box which has the same size as the CNN's input samples. Sufficiently large samples are required by this CNN to ensure its accuracy, which limits its resolution.

Developed from previous CNN architectures, the standard U-Net CNN (Ronneberger et al., 2015) mitigates the main flaw by classifying each pixel in the input image with a very high precision. As a convolutional neural network, U-Net can provide a high accuracy and excellent resolution for semantic segmentation. This is because the U-Net is a deep network with short-connections between the encoder and decoder structure to enable an accurate semantic segmentation: the short connections ease information propagation in training and compensate details for high level semantic features (Zhang et al., 2018). This high-level accuracy makes the U-Net architecture favored by medical personnel for analyzing MRI, CT, and ultrasound scans (Zhou et al., 2018); (Han and Ye, 2018); (Yang et al., 2019a); (Yap et al., 2017); (Behboodi and Hassan, 2019) and signal denoising (Sun et al., 2020); (Xiong et al., 2022). In addition, the high resolution of U-Net has been successfully used for crack detection in concrete structures (Cheng et al., 2018); (Liu et al., 2019); (Li et al., 2020). For scenes with a complex background, U-Net has been used to detect prominent features such as roads in images taken by satellites and performs better than traditional CNN methods (Constantin et al., 2018), (Yang et al., 2019b). The convolutional blocks in U-Net can be replaced with high-performing blocks with residual branches (Yang et al., 2019b) and short-connections can be modified by including several convolutional layers (Zou et al., 2021) to improve detection accuracy and resolution in complex backgrounds. In geophysics, U-Net architectures are used to identify faults in seismic images and delineate their 3D patterns (Wu et al., 2019), (Guitton, 2018). U-Net usually requires the support from powerful hardware because of the millions of parameters, so that it is limited in clusters rather than directly implemented on embedded devices.

1.2 Project background

Sandstones in the Middle East study area are cut by a NW-trending, roughly vertical set of long cracks that dissect the rocks, forming blocky massifs and elongated rock fins that are several tens to hundreds of meters wide, up to a kilometer or more in length and several tens of meters in height. NE-trending vertical cracks form a second but less prominent set of vertical cracks in the study area. Cliff faces or facades are marked by traces of both vertical and sub-horizontal cracks, with the latter aligned parallel to bedding. Discontinuous vertical cracks located within the rocks terminate at or near bedding surfaces suggesting that mechanical changes across bedding locally inhibited crack growth (Pollard and Aydin, 1988). Bedding-parallel cracks are relatively short, discontinuous features caused mostly by failure along bedding surfaces exposed in vertical facades. Consequently, failure along bedding surfaces should be anticipated in the design of underground excavations. This type of failure is common in crack (joint) bounded rock fins, where failure and collapse along bedding surfaces leads to the development of natural arches.

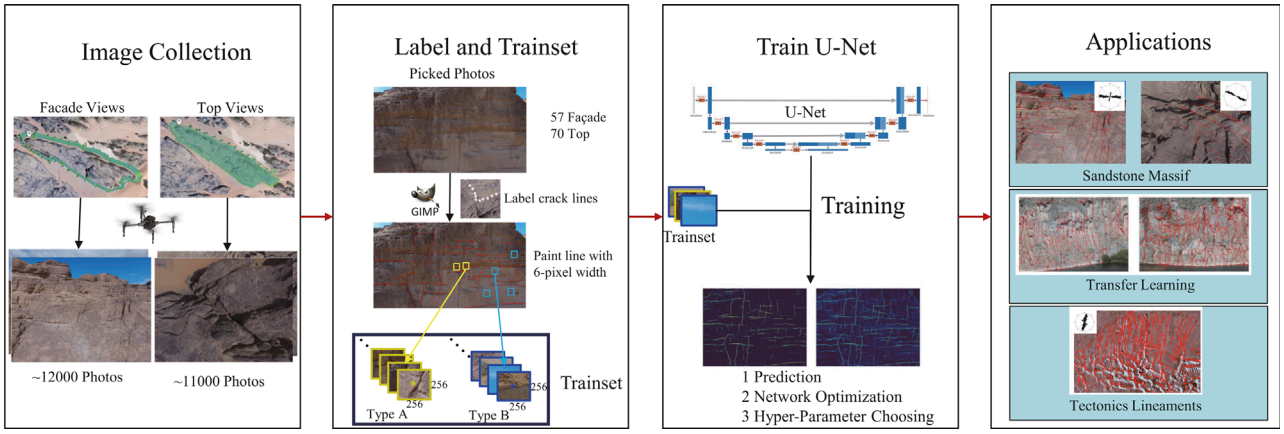


FIGURE 1
Workflow of the procedure for training and labeling of cracks in drone photos.

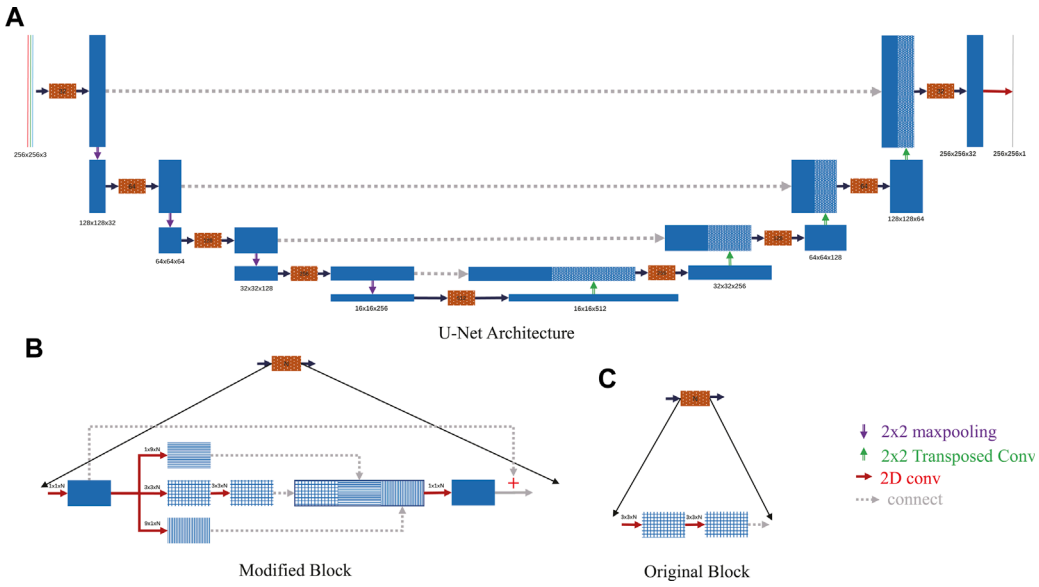


FIGURE 2
U-Net architecture used for detecting cracks in drone images. In **(A)** and **(B)** each arrowed tier represents a layer; the blue boxes represent features after the calculations from that layer. The input is a 256×256 RGB image. After five convolutional blocks (brown dotted block in a) and down-sampling (violet downside arrow), the feature map has the size 16×16. This is followed by up-sampling of features with transposed layers (green upside arrow), concatenation (gray dotted arrow) with the former output of the same size, and a final convolution. The final output size matches the 256×256 input to allow individual pixel classification of the input. Each convolutional block contains 6 separate convolutional layers (red dash arrow) with a final input identity operation (+) summed with the residual function block to form the desired output. **(C)** is the block of the original U-Net in (Ronneberger et al., 2015) in comparison with **(B)**.

Large cracks, especially those longer than 1 m, pose a problem for engineers who must drill into portions of the sandstones. Prior to drilling, a strict safety assessment must be carried out to access the rock integrity. Mechanical integrity is related to the density and distribution of large cracks, where drilling into the massif with a high density of cracks must be avoided or extra precautions should be taken. To quantitatively estimate the crack density, more than 23,000 drone photos were taken of the study area. Based on the success of the U-Net architecture, we now use it to detect cracks in photographs of rock faces. In our work, we add special convolutional modules which include residual shortcuts and two additional branches to better match crack orientations. More than 100 h were used to manually label large cracks in 127 high-resolution training images, which were then used to train the architecture. Applying the trained U-Net to new input images achieved an accuracy of 98%, which is sufficient to assess rock integrity prior to drilling into the massif.

In addition we used transfer training to generalize the U-Net crack detector so it is applicable to different rock types.

We describe our results in four sections. Following the introduction, we present the methodology of our CNN procedure for crack detection. The workflow for training and testing of the U-Net is shown in **Figure 1**. After this, we describe the training of the network and present results from the validation set. The next section presents the numerical results of applying the trained U-Net to more than 23,000 unlabeled drone images of our survey area. To test the generalizability of the trained U-Net, we use it to detect cracks in photos of volcanic rocks near the Teton dam site in Idaho and do transfer learning and fine-tuning (Shin et al., 2016) to improve accuracy. Finally we test the trained U-Net model on lineaments in photos of Mars taken by a Mars orbiter. We draw conclusions in the last section.

2 Network and loss function choosing

2.1 U-net architecture

To better match the reqiurements of crack detection, our U-Net is designed and different from standard U-Net. The U-Net architecture is diagrammed in **Figure 2**. It consists of a series of contracting encoders followed by expanding decoders weighted with features from the contraction path. In addition to the cross-connected paths, each convolution block implements an internal encoder/decoder with three convolution branches to enhance the detection of specific orientations of cracks. The convolution block forms the residual function to which an identity operation is added (He et al., 2016). The dimensions of the final U-Net output are the same as the input images but there is only one output channel of binary values to represent the probability of a pixel to be crack.

The U-Net design was chosen for the following reasons.

- **Additional convolution branches**: This idea comes from the Inception Network (Szegedy et al., 2016) because most of the cracks in our drone images are approximately horizontal or vertical. The two additional convolution branches with filter sizes 1×9 and 9×1 enable the efficient extraction of crack features by focusing on a specified dimension without the need for a square filter. A square filter with many parameters can detect a wide variation of crack orientations, but it is computationally inefficient if the cracks are confined to just a few orientations. This is a data specific enhancement.
- **Residual function connection**: In deep learning models, the convergence rate and accuracy can become degraded with an increase in the number of layers (He and Sun, 2015). Including a residual operation to each convolution block helps to improve accuracy in deep CNN models (He et al., 2016), (Zhang et al., 2018).

Our U-Net has 9 convolutional blocks, including 54 convolution layers, 4 maxpooling layers, and 4 transposed layers. For each convolution layer, we include batch normalization (Ioffe and Szegedy, 2015) and ReLU activation (Eckle and Schmidt-Hieber, 2019). Each convolution block is followed by a dropout layer with a rate of 0.5 (a random choice, half are dropped) for additional regularization. The final layer uses a sigmoid function to constrain the output of U-Net between 0 and 1.

The labeling targets are cracks that tend to form less than 1% of the entire set of images. Therefore, there is an imbalanced set of equations due to most labels having a label value of 0 (background). This typically leads to poor convergence and large errors in inference labeling. To overcome the imbalance problem, we test the performance of two weighted loss functions.

1. Weighted cross entropy (WCE) (Sousa Aurelio et al., 2019):

$$WCE(y,p) = -\sum_i \left[\lambda y^{(i)} \log p^{(i)} + (1-\lambda)\left(1-y^{(i)}\right)\log\left(1-p^{(i)}\right)\right] \quad (1)$$

where $y^{(i)}$ is our assigned label for the i^{th} example, $p^{(i)}$ is the U-Net prediction, and λ is a scalar chosen to improve performance.

2. Focal loss (FL) (Lin et al., 2017):

$$FL(y,p) = -\sum_i \left[\lambda y^{(i)} \times \left(1-p^{(i)}\right)^{\gamma} \log p^{(i)} + (1-\lambda) \times \left(1-y^{(i)}\right) \times \left(p^{(i)}\right)^{\gamma} \log\left(1-p^{(i)}\right)\right] \quad (2)$$

where we use the constant $\gamma = 2$.

WCE is a typical loss function used in classification problems where λ provides for a rebalancing of the possibly underrepresented positive pixels in the loss calculation. FL is used to address class imbalance by down-weighting the contribution of correctly identified pixels to focus on the loss due to misidentified pixels. We used a fixed value of $\gamma = 2$, varying only λ for all comparisons of each loss function. The FL formula reduces to the WCE formula when $\gamma = 0$. Our U-Net architecture is implemented using Keras 2.2.3 and Tensorflow-GPU 1.14.

2.2 Labeling, training and validation

Although the facade and top images have different crack features of the massifs, we only describe U-Net training of the facade images because the workflow is the same for the top images.

2.2.1 Labeling and picking of sub-samples

We select 57 photos from the facade view to be manually labeled for training and validation, which contain typical cracks examples. 54 photos are used for training and 3 are used for validation. The public-domain editing software GIMP (GNU Image Manipulation Program) is used to label the crack lines. We use the pen tool in GIMP to mark crack lines with many anchor points. Anchor points are required to locate the center of the crack along the normal direction to ensure the accuracy of labels. To convert labeled lines to pixels, we paint the line with a width of 6-pixels to closely match the crack size of interest. Image labeling is a time intensive task that is necessary for the training of the CNN. Labeling requires approximately 30–60 minutes per photo for a human, depending on the number of cracks. The size of each image is $4,000 \times 3,000$ or $4,000 \times 2000$ so that more than 100 hours (including labeling the 70

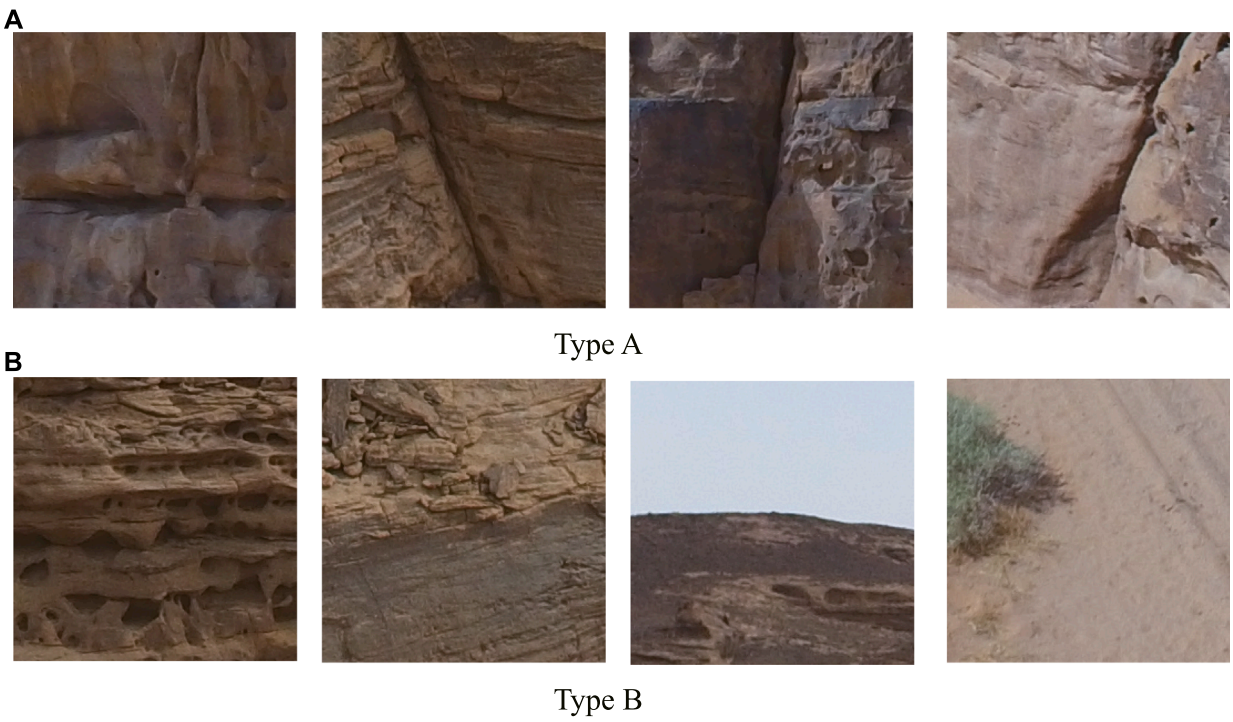


FIGURE 3
Different dataset samples. We use the two types of subimages to balance cracks in the dataset. **(A)** type A sub-samples that contain cracks; **(B)** type B sub-samples that only contain the background and some have confusing and crack-like features. The size of each sample image is 256 × 256.

top images) are needed to manually label the photos for training and validation.

The biggest problem in labeling is defining what constitutes a crack of interest. In the drone photos, there is a large variation in the size of cracks and not all cracks are of interest. For these data, geologists and building-site engineers are only interested in identifying cracks longer than 1 m, which are long enough to indicate possible instability in the building. Soil and small-rock sections covering parts of cracks are not labeled, resulting in discontinuous crack labels. Another problem is the mixture of horizontal cracks and some special bedding contacts. Some bedding contacts are strongly eroded so that there are some troughs along bedding surfaces. It is challenging to discriminate them from cracks. So we could include some troughs into the "horizontal cracks" label.

There are two types of subimages selected from the labeled photos: subimages of type A contain labeled cracks (Figure 3A); type B subimages contain the background rock, sky, sand, and/or bare rocks, which are devoid of labeled cracks (Figure 3B). Some samples in type B includes items which complicate network training, including line-like shapes of rain traces, shadow edges, and trees. Although the U-Net output is a binary classification, identifying types for inclusion in training allows for selecting a good balance of cracks and non-cracks.

For samples of type A, labels are indexed with small random shifts in the choice of index positions along the cracks to avoid cracks are always in the center of training samples. Type B subimages are randomly sampled. To reduce the bias of samples being too close to each other, a minimal spacing distance was set for center points of types A and B as 70 and 150 pixels, respectively. We select a count of 100 A types and 200 B types in each photo for balanced training.

2.2.2 Network training

Subsampling the 57 photos resulted in a training dataset consisting of 16,200 subimages and a validation dataset consisting of 900 subimages. The networks are trained using two Nvidia v100 GPU cards with a batch size of 20 and an initial learning rate of 0.001. The data are augmented in each batch of processed training samples by adding copies with horizontal, vertical, or 0°–45° rotation using a reflected sample to fill in the boundary space created by rotation. The data augmentation is performed within Keras.

The maximum number of epochs is set to 100 and training is stopped when either the maximum is reached or when loss in the validation set does not decrease for 30 epochs (Figure 4). Loss values in both the training and validation sets do not decrease after around epoch number 50, so training is terminated at epoch 80. Our modified U-Net provides lower loss values for both the training and validation sets, compared to the standard U-Net architecture in Figure 2C.

2.2.3 Validation and confusion matries after skeletonization

We used two loss functions and a range of λ values to train the U-Net. After trial-and-error tests, an output value threshold of 0.5 is used for all pixels in the U-Net for classification as either a crack or background. Cracks in the output are marked as red in the figures. For comparison with the U-Net, the automatic fracture

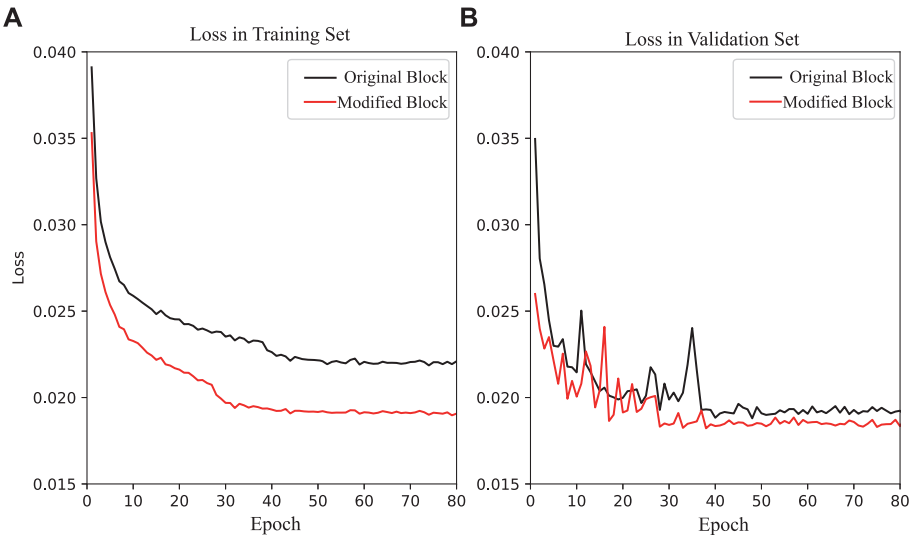
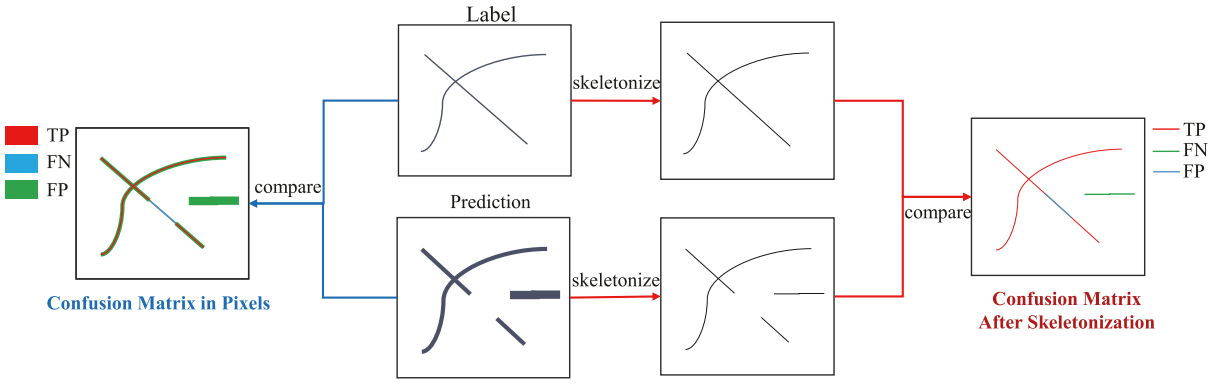


FIGURE 4
Learning curves for the training data. **(A,B)** show the loss values in training set and validation set. The loss function uses WCE and λ = 0.85. Displayed two lines show results of U-Net with original and modified blocks in **Figure 2**.

TABLE 1 **Confusion matrix definition and results for the validation images. a) is the definition of the confusion matrix and related values: recall (R), precision (P) and accuracy (A). b) is the workflow to get the matrix both in pixels and after skeletonization. The values in c) and d) are examples from results of validation photos in which U-Net uses WCE and λ = 0.85.**

		Actual class		
		Crack	Non-crack	
Prediction	Positive	True positive (TP)	False positive (FP)	$P=\frac{TP}{TP+FP}$
	Negative	False negative (FN)	True negative (TN)	
		$R=\frac{TP}{TP+FN}$		$A=\frac{TP+TN}{TP+FP+FN+TN}$

(a) Definition of Confusion Matrix



(b) Confusion Matrix Illumination

	Crack	Non-crack	
Positive	288228	443882	P=39%
Negative	142541	35221349	
	R=67%		A=98%

(c) Confusion Matrix in Pixels

	Crack	Non-crack	
Positive	52419	53399	P=50%
Negative	6858		
	R=88%		

(d) Confusion Matrix After Skeletonization

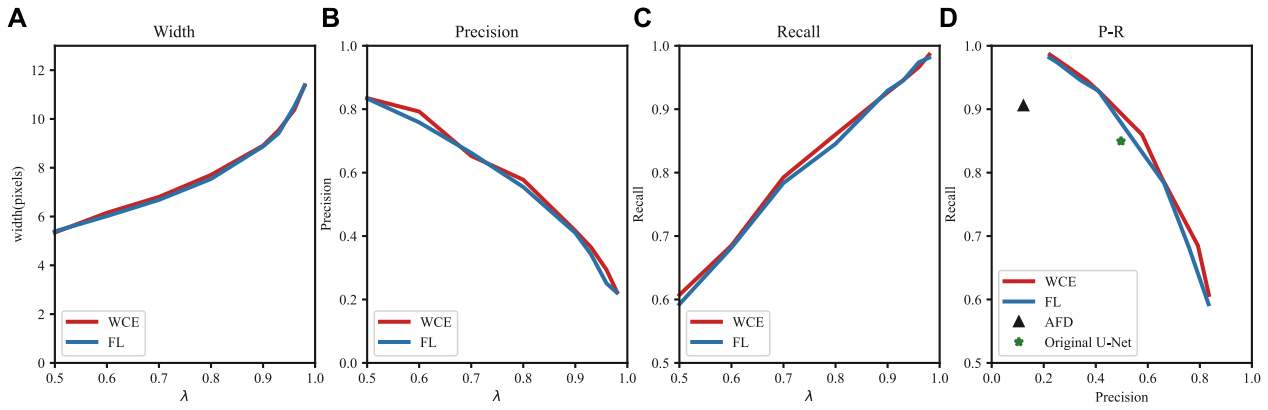


FIGURE 5
Perfomance metrics of the modified U-Net plotted against epoch number for the training data. We set eight λ from 0.5 to 0.98 in the loss functions and display the average width, precision, recall of predictions in **(A–C)**. The P and R values are calculated by the confusion matrix after skeletonization. In **(D)**, the relationship between P and R is for both the standard and modified U-Net architectures, and we add the performances of AFD method as comparison.

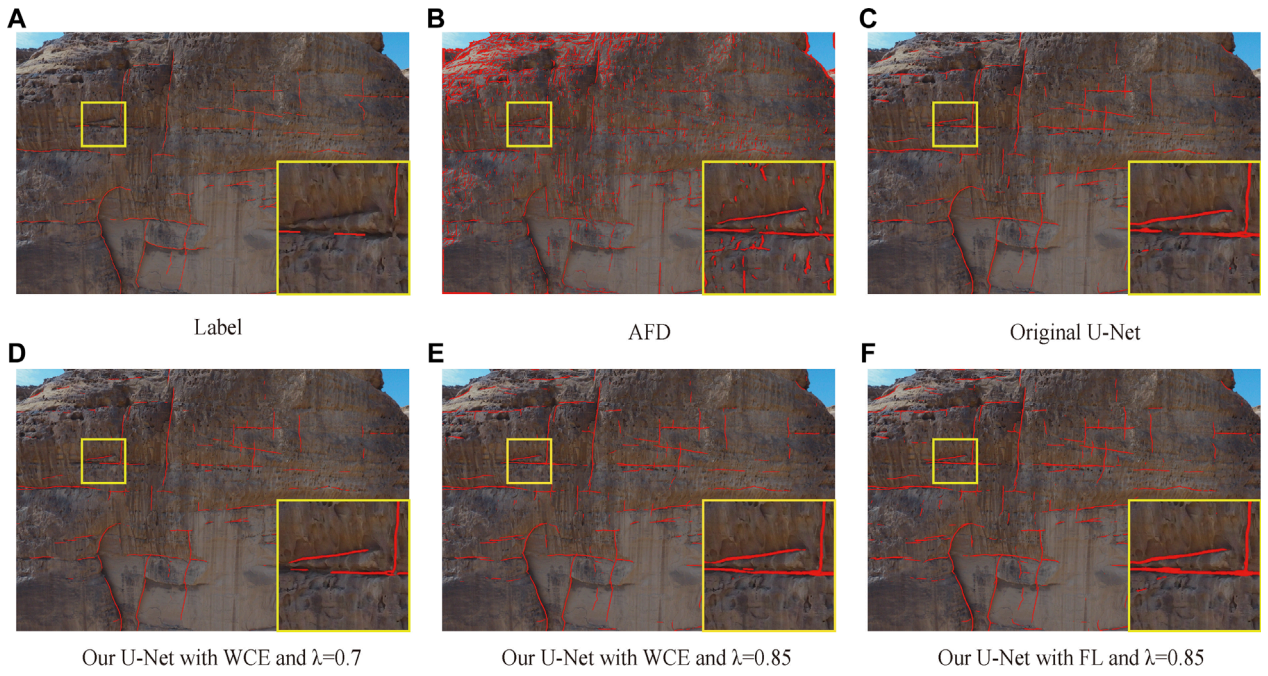


FIGURE 6
Results of different algorithms applied to two validation images. Photos are 4,000 × 3,000 pixels each. Cracks marked in red color overlie on raw photos. Human labeled images are shown in **(A)** and we only label those that we are quite confident about; **(B)** is calculated by automatic fracture detection (AFD) in (Prabhakaran et al., 2019); **(C–E)** are trained with weighted cross entropy (WCE) using different values of λ; **(F)** focal loss (FL) shows output from one value of λ.

detection (AFD) code in (Prabhakaran et al., 2019) is used as a ridge-detection method for the validation images. We used typical 108 shearlet systems with a threshold of 0.52 for AFD.

The confusion matrix is necessary to evaluate the performance of the U-Net for different parameter values. To normalize the crack widths, we skeletonize the labeled and predicted cracks to be a same width of one pixel (Zhang and Suen, 1984), (van der Walt et al., 2014). After skeletonization, the labeled crack is only 1 pixel wide. **Table 1** compares the ground-truth labels with the predicted ones using TP, FP and FN values. The identification accuracy is 98% but the P and R parameters usually are the most important ones for comparison purposes.

As shown in **Figures 5**, **6**, λ is the key parameter which controls the performance of the U-Net. With an increase in the value of λ,

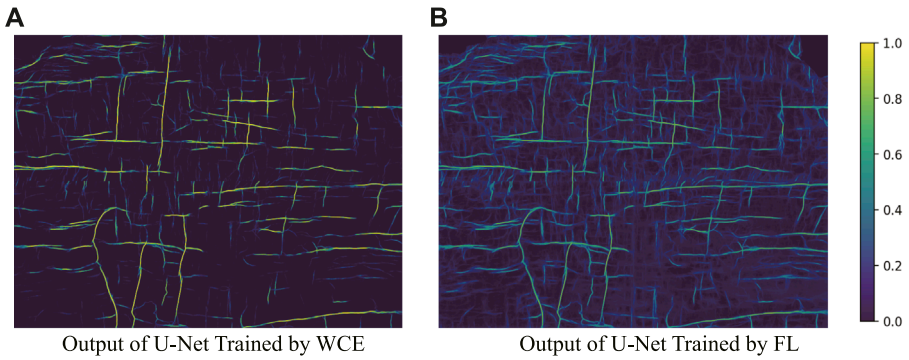


FIGURE 7
U-Net predictions of the image in Figure8. Here, **(A)** is the raw output from the U-Net with WCE as the loss function and **(B)** is the raw output from the U-Net with FL as the loss function. The two kinds of loss functions share the same $\lambda = 0.85$ value. We use the sigmoid as the activation function in the last layer so when a value of the U-Net output is closer to 1, it is more likely to be a crack.

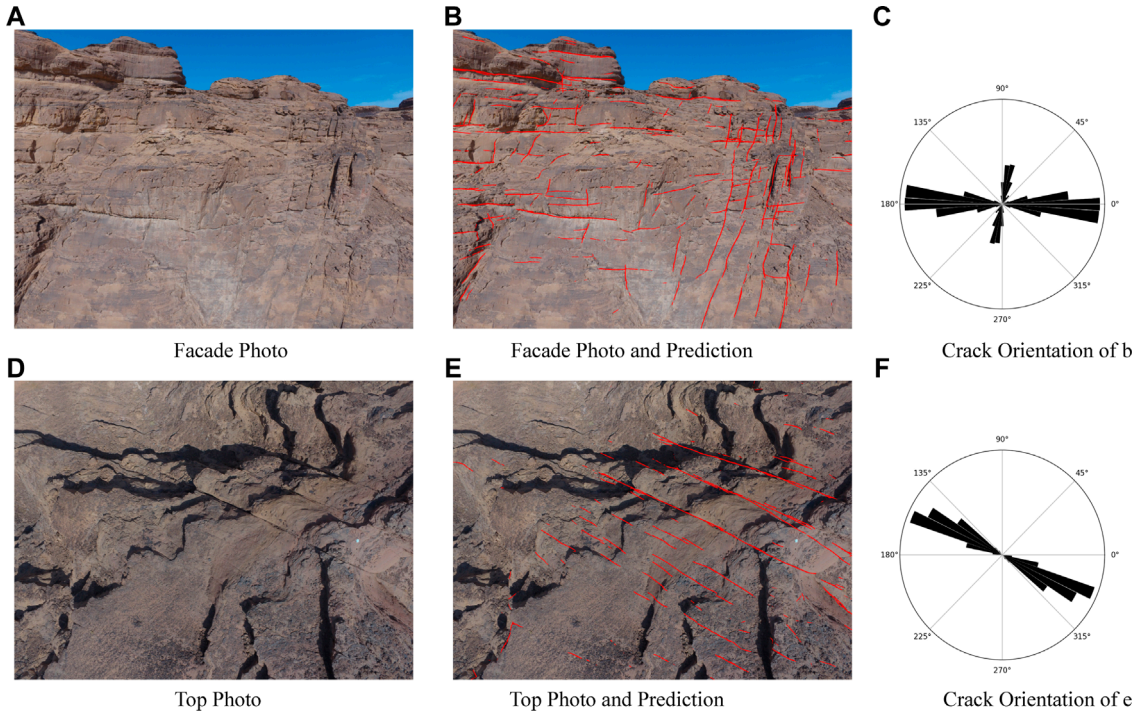


FIGURE 8
Photos labeled by U-Net from unlabeled photos. We picked two example images from the out-of-the training set. **(A,B)** depict a facade view; **(D,E)** are a top view. **(C,F)** are their distributions of crack orientations and densities.

the U-Net tends to be more aggressive in labelling more pixels as cracks, which increases the average width and recall values while the precision values goes down. Therefore, the choice of λ is important for balancing the precision and recall values. The underestimation of crack density can lead to safety problems, so that high FP counts are more acceptable than missing cracks of interest. We select $\lambda = 0.85$ to reduce the high number of FP counts associated with large values of λ.

Figure 6B displays the result from the AFD code. The result of AFD shows many false positives caused by sharp changes in intensity associated with non-crack features and it omitted some labeled cracks. Also, the P-R result in Figure 6D shows a quite low precision value and a modest recall compared to the U-Net. The result of the original U-Net is shown in Figure 6E. It correlates well with the human labelled result. However, the P-R values of the original U-Net are less than these of the

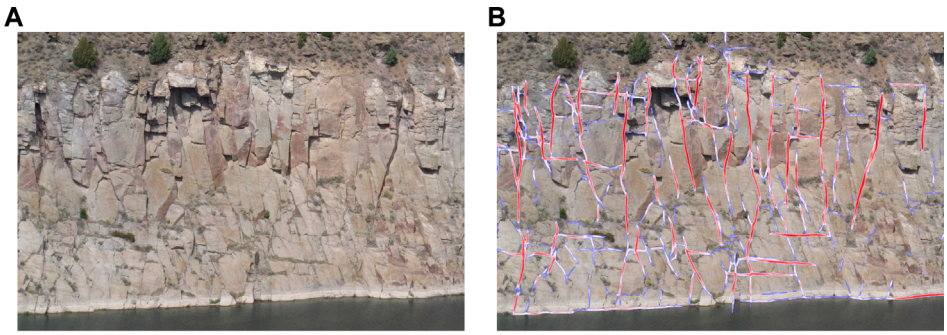


FIGURE 9
Crack detection of volcanic rocks at the Teton Dam Site, Idaho. **(A)**: near-vertical cooling cracks in volcanic rocks just downstream from the dam site; **(B)**: cracks labeled in red by U-Net. The red markings correspond to a probability between 0.7 and 1.0 in the softmax output, while the white-bluish markings correspond to a probability between 0.3 and slightly less than 0.7.

modified U-Net, which indicates the superiority of the modified U-Net.

The results with using the U-Net with WCE and FL are shown in Figures 6E, F. However, the P-R line of FL is always below that of WCE in Figure 5D. Figure 7 shows that the output of WCE more clearly highlights large cracks of interest compared to FL. Given the limited values used for comparison, WCE is a better choice for our work.

3 U-net crack detection in three geological bodies

The trained U-Net is now used to label cracks in the study area with the large sandstone massifs. Over 23,000 unlabeled photos of the sandstone massif are taken by a drone with a high-resolution camera. In addition, the trained U-Net is used to label cracks in photos of volcanic rocks in Idaho as well as those in Martian orbiter pictures. These last two examples were used to show that the U-Net trained on photos of sandstone cracks can also be used to label the cracks in photos of rocks with a different geological genesis. We also show that the accuracy of the U-Net crack detection can be improved by transfer learning.

3.1 Labeling of cracks in photos of sandstone massifs

A drone was flown over the sandstone massifs and recorded 23,845 photos of the tops and facades (sides) of the target rock masses. The paths of drone flights are optimized for achieving a uniform image resolution, where variations in camera-to-target distances created pixels approximately 0.8–10 cm wide. However, most have a spatial resolution of several centimeters, which ensures that the U-Net is not troubled by varying crack widths due to the drone being at significantly different distances from the rock face. The pixel dimensions of the photos range from are $4,000 \times 2000$ to $4,000 \times 3,000$ pixels with about 80% overlap of areas.

Our U-Net does not have any fully connected layers so its input size can be variable. We were limited by the amount of GPU memory, so we partitioned each photo into 4 small sub-photos, each with half the width and height of the original. Partitioning is shown by the green lines in Figure 1. The U-Net labeling of each photo takes about 3 s per GPU card.

Then the U-Net model trained was applied to the unlabeled facade images. As shown in Figures 8A, B, the U-Net results for unlabeled images are judged to be of acceptable accuracy. The orientation distribution in Figure 8C shows that most cracks in this photo are horizontal but there are some along 70°. We then applied the same method and trained a new U-Net for the top images, where we labeled 70 images which created a dataset with 16,800 subimages. Figures 8D, E shows a processed top image, showing results as accurate as those produced for the facade images. Figure 8F shows all cracks have the same orientation angle of about 150°, which is consistent with the cracks seen in Figure 8D. The U-Net labeling of the all images (facade and top) required approximately 22 GPU hours to finish. The labeled images were then used to assess the crack densities and orientations in the areas of engineering interest.

3.2 Labeling of cracks in photos of volcanic rocks and transfer learning

After achieving accurate detection of cracks in the sandstone photos, we apply the sandstone U-Net[1] to photos of volcanic rock faces at the Dam site in Idaho, United States. The Teton Dam in eastern Idaho failed catastrophically on the morning of 5 June 1976, causing a large flood. The dam is built on silica-rich volcanics (welded tuff) that were derived by the eruption of the Yellowstone super-volcano. Investigations suggest the failure of the dam was related to movement of water through large fractures (cracks) in the bedrock foundation.

1 The sandstone U-Net model is exclusively trained from the sandstone pictures on facade.

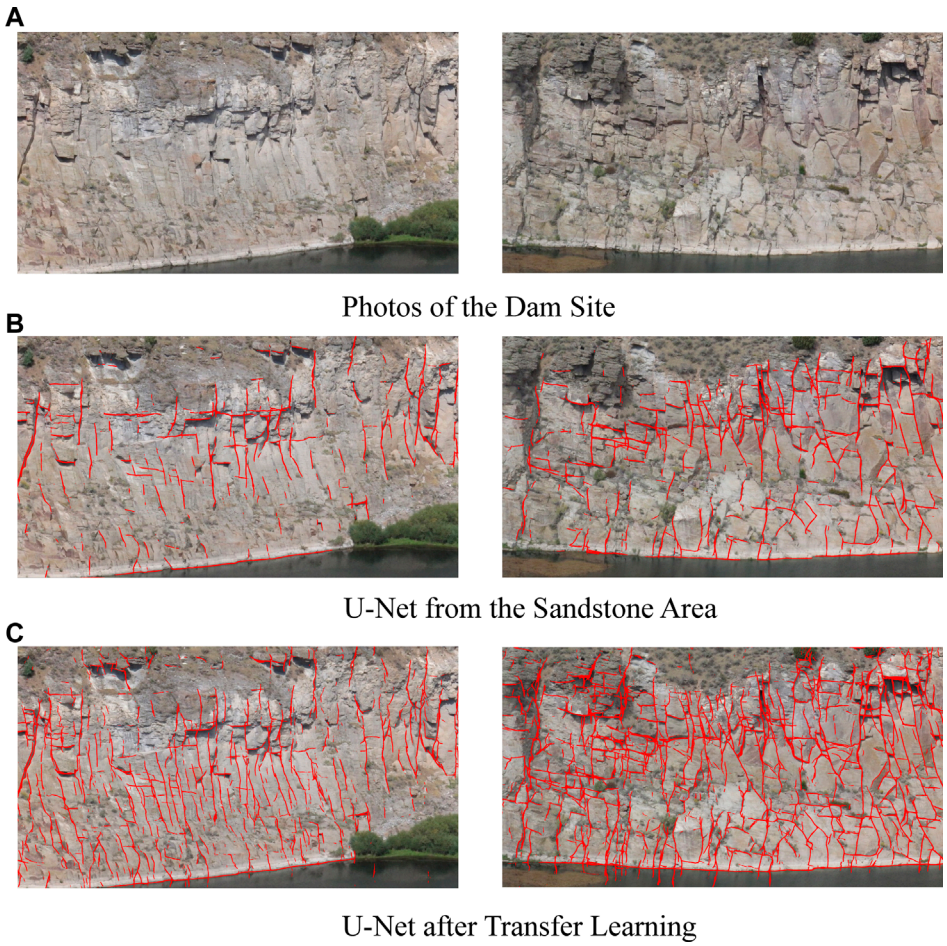


FIGURE 10
(A) Raw photos from the Teton dam site, and crack labels computed by the **(B)** sandstone U-Net and the **(C)** hybrid U-Net models. Red labels have a sigmoid probability between 0.5 and 1.0.

Figure 9A depicts a $4,000 \times 2000$ picture from http://gigapan.com/gigapans/163482, which is the cliffs at the Teton Dam site. Mapping the cracks in the Teton Dam photo tests the ability of the sandstone U-Net to accurately delineate cracks in photos of volcanic rocks. The nearly vertical cracks are cooling joints in welded volcanic tuff. The nearly vertical cracks mapped by our sandstone U-Net are labeled by the red lines in **Figure 9B**, where many cracks are correctly labeled. However, some cracks are not labeled even though they are visible to the eye and there are some false positives. This demonstrates that the sandstone U-Net algorithm is capable of mapping the trends in crack orientations of rocks that are different from those from which it was trained.

We can improve the accuracy of the sandstone U-Net model by using transfer learning. This method has been applied on pavement distress detection in (Gopalakrishnan et al., 2017), which uses a pre-trained VGG-16 to construct their CNN. In our work, transfer learning combines the weights learned from the sandstone images as well as those learned from a small number of labeled cracks from the Teton dam photos. It only requires a small number of labeled images from the photos of volcanic rock faces because it reuses common crack patterns from the sandstone U-Net. This can result in significantly less labeling and computation time compared to standard CNN training. For relabeling, we used another $4,000 \times 2000$ photo of the Gigapan image, manually labeled the cracks and broke it up into 120 256×256 sub-photos. The transfer learning required less than 60 min compared to an estimate of more than 20 h to train a new U-Net model.

For the transfer learning, we freeze the weights in blocks from 3 to 7 and allow four symmetric blocks (block 1,2 and 8,9 in **Figure 2**) to be trained on the newly labeled photos, which is called as in fine-tune (Shao et al., 2018). This decreases the trainable parameters from 20,000,000 to 730,000. An Adam optimizer is used and we set the learning rate to be 10^{-4} in order to fine-tune the trainable layers. The batch size is 5 and the total number of 256×256 sub-photos is 120. The transfer training is halted at epoch 30 which takes no more than 20 minutes of computation time on a Nvidia v100 GPU. We denote this U-Net as the hybrid U-Net model because the original weights were trained on images of sandstones and adjusted to those for volcanic rocks.

After transfer learning, the sandstone U-Net and hybrid U-Net were applied to the raw $4,000 \times 2000$ images in **Figure 10A** to give

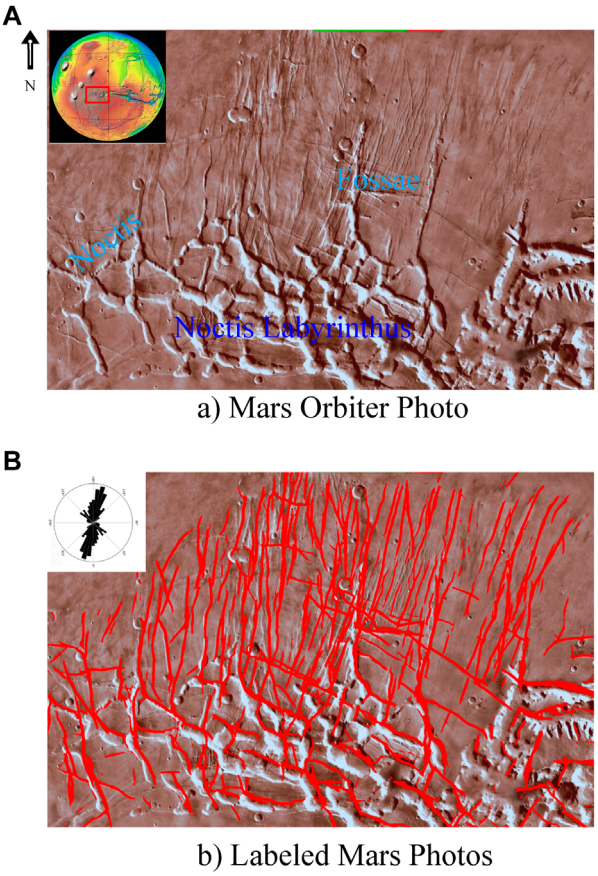


FIGURE 11
(A) Photo taken by a martian orbitor and **(B)** the labeled photo. The *crack* distribution is depicted in the inset Rose diagram of **(B)**, and the labels are computed by the sandstone U-Net. In this example, a *crack* is defined as a sharp linear change in the photo's intensity value.

the labeled cracks in **Figures 10B, C**, respectively. **Figure 10C** show a much denser and more accurate collection of labeled cracks than the ones in **Figure 10B**. The important accomplishment is that it required no more than an hour of manual labeling to achieve this goal.

3.3 Labeling of photos of martian Nocti Labyrinthus

We notice that the U-Net detector is effective at detecting lineaments in photos, where such lineaments might be associated with large-scale tectonics. Therefore we apply the U-Net crack detector to photos of the surface of Mars taken by the Viking orbiter. **Figure 11A** depicts the region known as Nocti Labyrinthus, an area of rift basins and grabens (Daniel and Cook, 2003).

Figure 11A depicts the original photo and **Figure 11B** shows the photo after labeling by the trained U-Net. Many of the lineaments seen in **Figure 11A** are clearly labeled in **Figure 11B**. Some of the subtle lineaments not easily detected in 13a are now delineated in the labeled image. Many of these lineaments are interpreted as grabens and fossae (Daniel and Cook, 2003). The Rose diagram here reveals the trends in their orientations, where the graben distribution is related to the ambient extensional stress field associated with local volcanic rifting.

4 Discussion

The sandstone U-Net Keras code and the trained weights will be available to the public by the public-domain site cited in the data availability statement. Part of this code will have the capability for transfer learning. We expect a number of practical applications for this crack detection code.

1. Semi-universal Cracks Detector for Cracks in Rock Faces. The U-Net trained on photos of a sandstone massif shows a precise detection of sandstone cracks and acceptable results for volcanic rocks. With a modest amount of extra labeling and transfer training we believe this U-Net can be a semi-universal detector of cracks in many types of rock faces. It can also be used to automatically detect large-scale lineaments in photos taken by planetary orbiters as a reconnaissance of a planet's or Moon's tectonics.
2. Real-Time Monitoring of Crack Development. The U-Net crack detector can be used as a real-time monitor of growing cracks if the photos are periodically taken and quickly analyzed for crack growth. One application with satellite photos of polar regions is to automatically detect the growth of crevasses and cracks in the ice, which would be a real-time monitor of climate changes. Other applications of real-time monitoring of crack growth include hazard assessment of failing dams, imminent landslides, and erupting volcanoes, which has some similarities with surface displacements monitoring by frequent InSAR recordings in (Sun et al., 2020).

There are some notable areas where our workflow can be improved. Foremost is to increase the efficiency of labeling. The pre-trained U-Net produces some FP and FN crack labels, which should be removed by manual labeling. Another complication is that close-up photos of the rock degrades the U-Net model in distinguishing erosional features from small cracks. The consequence is that it produces false positives. We believe this problem can be mitigated by also introducing distance information into the input of the U-Net. This distance information can be included with the use of lidar or radar instruments in the drone.

5 Conclusion

This work presents a successful use of an U-Net CNN to label cracks in rock faces. We compare U-Net with a ridge-detection method (AFD) and find that the U-Net trained with just a few images provides a more accurate detection of cracks. Our results shows that the U-Net approach provides a viable alternative to the conventional AFD method for detecting cracks in rock massifs. Besides, we find that the U-Net with residual shortcuts and additional convolutional branches shows a better accuracy than the original U-Net. And the performance of the U-Net is largely controlled by the value of the

λ weight in the loss function. We think the reason is that rock cracks often have a narrow linear shape containing only a limited number of pixels, which has to be balanced by weighted loss. Considering the location of cracks is the key feature for engineers, an improvement may be to identify the location of cracks by anchor points, which represents important future work of ours.

One disadvantage of U-Net is that it has millions of parameters and requires a large amount of manual labeling to train the network. However, U-Net has transferability capabilities so that a well-trained U-Net with transfer training can be used to detect cracks on other kinds of rocks without an extensive effort in relabeling. We proved this to be true by using transfer training to significantly improve the accuracy of labeling cracks in the Teton Dam photos. Less than 60 minutes were required for manual labeling of cracks in fewer Teton Dam photos. Moreover, the sandstone U-Net could delineate some line-like geology features in Mars photos.

In summary, our results suggest that our trained U-Net with transfer training is a semi-universal detector of cracks in images of almost any type of geological outcrop. The scale of cracks can range from centimeters to kilometers which depends on the resolution characteristics of the photographs. Practical applications of this method include the use of crack detection for real-time monitoring of crack growth, such as needed for safety assessment of dams, landslides, volcanoes and man-made structures.

Data availability statement

Our Python code, well-trained weights and all labelled images are be made available on Colab (colab.research.google.com/drive/1n7K1esyM_FmWb-6ZwT8eGWxylOk4xoOp) to the public to be used as a crack or lineament detector for images of many geological outcrops.

Author contributions

YS: propose, build, and train the networks, check result and improve accuracy, label data MB: train the networks, check result and improve accuracy KJ: collect and offer drone data, check result and improve accuracy, offer standard of cracks DT: label data, offer standard of cracks YH: label data, offer suggestions to the paper MM: offer drone data, offer suggestions to the paper RB: offer suggestions to cracks, offer cracks photos of volanic rocks GS: check result and improve accuracy, main proposer of the project.

Funding

The research reported in this publication was supported by the King Abdullah University of Science and Technology (KAUST) in Thuwal, Saudi Arabia. For computer time, this research used the resources of the Supercomputing Laboratory at KAUST and the IT Research Computing Group. We thank them for providing the computational resources required for carrying out this work.

Acknowledgments

We thank the great help in reviews from Prof. Ning Jieyuan in Peking University and valuable suggestions from reviewers.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Aydan, Ö., Ulusay, R., and Tokashiki, N. (2014). A new rock mass quality rating system: Rock mass quality rating (RMQR) and its application to the estimation of geomechanical characteristics of rock masses. *Rock Mech. rock Eng.* 47 (4), 1255–1276. doi:10.1007/s00603-013-0462-z

Behboodi, B., and Hassan, R. (2019). "Ultrasound segmentation using U-net: Learning from simulated data and testing on real data," in 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (IEEE), 6628–6631.

Bemis, S P., Micklethwaite, S., Turner, D., James, M R., Akciz, S., Thiele, S T., et al. (2014). Ground-based and uav-based photogrammetry: A multi-scale, high-resolution mapping tool for structural geology and paleoseismology. *J. Struct. Geol.* 69, 163–178. doi:10.1016/j.jsg.2014.10.007

Cao, V D., and Anh, L. D. (2019). Autonomous concrete crack detection using deep fully convolutional neural network. *Automation Constr.* 99, 52–58. doi:10.1016/j.autcon.2018.11.028

Cha, Y-J., Choi, W., and Büyüköztürk, O. (2017). Deep learning-based crack damage detection using convolutional neural networks. *Computer-Aided Civ. Infrastructure Eng.* 32 (5), 361–378. doi:10.1111/mice.12263

Cheng, J., Xiong, W., Chen, W., Gu, Y., and Li, Y. (2018). "Pixel-level crack detection using U-Net," in TENCON 2018-2018 IEEE Region 10 Conference (IEEE), 0462–0466.

Constantin, A., Ding, J-J., and Lee, Y-C. (2018). "Accurate road detection from satellite images using modified U-net," in IEEE Asia Pacific Conference on Circuits and Systems (APCCAS), 423–426.

Daniel, M., and Cook, A C. (2003). Erwan garel, yves lagabrielle, and marie-hélène cormier. Volcanic rifting at martian grabens. *J. Geophys. Res. Planets* 108 (E5).

Dorafshan, S., Thomas, R J., and Maguire, M. (2018). Comparison of deep convolutional neural networks and edge detectors for image-based crack detection in concrete. *Constr. Build. Mater.* 186, 1031–1045. doi:10.1016/j.conbuildmat.2018.08.011

Dyskin, A. V., and Germanovich, L. N. (1993). Model of rockburst caused by cracks growing near free surface. *Rockbursts seismicity mines* 93, 169–175.

Eckle, K., and Schmidt-Hieber, J. (2019). A comparison of deep networks with relu activation function and linear spline-type methods. *Neural Netw.* 110, 232–242. doi:10.1016/j.neunet.2018.11.005

Gopalakrishnan, K., Khaitan, S K., Choudhary, A., and Agrawal, A. (2017). Deep convolutional neural networks with transfer learning for computer vision-based data-driven pavement distress detection. *Constr. Build. Mater.* 157, 322–330. doi:10.1016/j.conbuildmat.2017.09.110

Guitton, A. (2018). 3d convolutional neural networks for fault interpretation. 80th EAGE Conference and Exhibition. Houten, Netherlands: European Association of Geoscientists and Engineers, 1–5.

Han, Y., and Ye, J. C. (2018). Framing U-Net via deep convolutional framelets: Application to sparse-view ct. *IEEE Trans. Med. Imaging* 37 (6), 1418–1429. doi:10.1109/tmi.2018.2823768

He, K., and Sun, J. (2015). "Convolutional neural networks at constrained time cost," in Proceedings of the IEEE conference on computer vision and pattern recognition, 5353–5360.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in The IEEE Conference on Computer Vision and Pattern Recognition (New York: CVPR).

Herbert, L. (2011). Hazard potential of zones of weakness in gravity dams under impact loading conditions. *Front. Archit. Civ. Eng. China* 5 (1), 90–97. doi:10.1007/s11709-010-0008-3

Hillier, J. K., Smith, M. J., Armugam, R., Barr, I., Boston, C. M., Clark, C. D., et al. (2015). Manual mapping of drumlins in synthetic landscapes to assess operator effectiveness. *J. Maps* 11 (5), 719–729. doi:10.1080/17445647.2014.957251

Ioffe, S., and Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167.

Kim, B., and Cho, S. (2018). Automated vision-based detection of cracks on concrete surfaces using a deep learning technique. *Sensors* 18 (10), 3452. doi:10.3390/s18103452

Li, G., Ma, B., He, S., Ren, X., and Liu, Q. (2020). Automatic tunnel crack detection based on U-Net and a convolutional neural network with alternately updated clique. *Sensors* 20 (3), 717. doi:10.3390/s20030717

Lin, T-Y, Goyal, P., Girshick, R., He, K., and Dollár, P. (2017). "Focal loss for dense object detection," in Proceedings of the IEEE international conference on computer vision, 2980–2988.

Liu, Z., Cao, Y., Wang, Y., and Wang, W. (2019). Computer vision-based concrete crack detection using U-net fully convolutional networks. *Automation Constr.* 104, 129–139. doi:10.1016/j.autcon.2019.04.005

Mikhail, E M., JamesBethel, S., and McGlone, J. C. (2001). *Introduction to modern photogrammetry*, 19. New York.

Mohan, A., and Poobal, S. (2018). Crack detection using image processing: A critical review and analysis. *Alexandria Eng. J.* 57 (2), 787–798. doi:10.1016/j.aej.2017.01.020

Pereira, F C., and Pereira, C E. (2015). "Embedded image processing systems for automatic recognition of cracks using uavs," in 2nd IFAC Conference on Embedded Systems, Computer Intelligence and Telematics CESCIT 2015, *IFAC-PapersOnLine*. 48. 10. 16–21. doi:10.1016/j.ifacol.2015.08.101

Pollard, D D., and Aydin, A. (1988). Progress in understanding jointing over the past century. *GSA Bull.* 100 (8), 1181–1204. doi:10.1130/0016-7606(1988)100<1181:piujot>2.3.co;2

Prabhakaran, R., Bruna, P.-O., Bertotti, G., and Smeulders, D. (2019). An automated fracture trace detection technique using the complex shearlet transform. *Solid earth.* 10 (6), 2137–2166. doi:10.5194/se-10-2137-2019

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-net: Convolutional networks for biomedical image segmentation," in International Conference on Medical image computing and computer-assisted intervention (Springer), 234–241.

Sander, P., Minor, T B., and Chesley, M M. (1997). Ground-water exploration based on lineament analysis and reproducibility tests. *Groundwater* 35 (5), 888–894. doi:10.1111/j.1745-6584.1997.tb00157.x

Sanderson, D J., Peacock, D C. P., Nixon, C W., Rotevatn, A., and Rotevatn, A. (2019). Graph theory and the analysis of fracture networks. *J. Struct. Geol.* 125, 155–165. doi:10.1016/j.jsg.2018.04.011

Shao, S., McAleer, S., Yan, R., and Baldi, P. (2018). Highly accurate machine fault diagnosis using deep transfer learning. *IEEE Trans. Industrial Inf.* 15 (4), 2446–2455. doi:10.1109/tii.2018.2864759

Shin, H., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., et al. (2016). Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. *IEEE Trans. Med. Imaging* 35 (5), 1285–1298. doi:10.1109/tmi.2016.2528162

Sobel, I. (1990). Machine vision for three-dimensional scenes. *Chapter Isotropic* 3 (3), 376–379.

Sousa Aurelio, Y., Matheus de Almeida, G., Leite de Castro, C., and Braga, A P. (2019). Learning from imbalanced data sets with weighted cross-entropy function. *Neural Process. Lett.* 50 (2), 1937–1949. doi:10.1007/s11063-018-09977-1

Sun, J., Wauthier, C., Stephens, K., Gervais, M., Guido, C., La Femina, P., et al. (2020). Automatic detection of volcanic surface deformation using deep learning. *J. Geophys. Res. Solid Earth* 125 (9), e2020JB019840. doi:10.1029/2020jb019840

Szegedy, C., Ioffe, S., and Vincent, V. (2016). Inception-v4, inception-resnet and the impact of residual connections on learning. *Corr. abs/1602*, 07261. doi:10.1609/aaai.v31i1.11231

Thrun, S. (2002). Probabilistic robotics. *Commun. ACM* 45 (3), 52–57. doi:10.1145/504729.504754

van der Walt, S., Schönberger, J L., Nunez-Iglesias, J., Boulogne, F., Warner, J D., Yager, N., et al. (2014). scikit-image: image processing in Python. *PeerJ* 2, e453. doi:10.7717/peerj.453

Vasuki, Y., Holden, E-J., Peter, K., and Micklethwaite, S. (2014). Semi-automatic mapping of geological structures using uav-based photogrammetric data: An image analysis approach. *Comput. Geosciences* 69, 22–32. doi:10.1016/j.cageo.2014.04.012

WilsonLeal da Silva, R., and de Lucena, D S. (2018). Concrete cracks detection based on deep learning image classification. *Multidiscip. Digit. Publ. Inst. Proc.* 2, 489.

Wu, X., Liang, L., Shi, Y., and Fomel, S. (2019). Faultseg3d: Using synthetic data sets to train an end-to-end convolutional neural network for 3d seismic fault segmentation. *Geophysics* 84, IM35–IM45. doi:10.1190/geo2018-0646.1

Wyllie, D C. (1999). *Foundations on rock: Engineering practice*. Second Edition2 edition. London: CRC Press.

Xiong, F., Zhou, J., Zhao, Q., Lu, J., and Qian, Y. (2022). "MAC-Net: Model-Aided nonlocal neural network for hyperspectral image denoising," in IEEE Transactions on Geoscience and Remote Sensing, 1–14. 60.

Yang, J., Faraji, M., and Basu, A. (2019). Robust segmentation of arterial walls in intravascular ultrasound images using dual path u-net. *Ultrasonics* 96, 24–33. doi:10.1016/j.ultras.2019.03.014

Yang, X., Li, X., Ye, Y., Lau, R Y. K., Zhang, X., and Huang, X. (2019). Road detection and centerline extraction via deep recurrent convolutional neural network U-NetConference name: IEEE transactions on geoscience and Remote sensing. *IEEE Trans. Geoscience Remote Sens.* 57 (9), 7209–7220. doi:10.1109/tgrs.2019.2912301

Yap, M H., Pons, G., Martí, J., Ganau, S., Sentís, M., Zwiggelaar, R., et al. (2017). Automated breast ultrasound lesions detection using convolutional neural networks. *IEEE J. Biomed. health Inf.* 22 (4), 1218–1226. doi:10.1109/jbhi.2017.2731873

Zhang, T. Y., and Suen, C Y. (1984). A fast parallel algorithm for thinning digital patterns. *Commun. ACM* 27 (3), 236–239. doi:10.1145/357994.358023

Zhang, Z., Liu, Q., and Wang, Y. (2018). Road extraction by deep residual U-NetConference name: IEEE geoscience and Remote sensing letters. *IEEE Geoscience Remote Sens. Lett.* 15 (5), 749–753. doi:10.1109/lgrs.2018.2802944

Zhou, Z., Rahman Siddiquee, M M., Tajbakhsh, N., and Liang, J. (2018). "UNet++: A nested U-net architecture for medical image segmentation," in Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support (Cham: Springer International Publishing), 3–11.

Zou, S., Xiong, F., Luo, H., Lu, J., and Qian Af-Net, Y. (2021). "All-scale feature fusion network for road extraction from Remote sensing images," in Digital Image Computing: Techniques and Applications (DICTA), 1–8.

frontiers | Frontiers in Earth Science

TYPE Original Research
PUBLISHED 13 March 2023
DOI 10.3389/feart.2023.1141220

OPEN ACCESS

EDITED BY
Wenjuan Zhang,
Aerospace Information Research Institute (CAS), China

REVIEWED BY
Yang Liu,
China University of Petroleum, China
Xiyan Zhou,
Chengdu University of Technology, China
He Qinglong,
Guizhou University, China

*CORRESPONDENCE
Chunlin Zhang,
mike_zcl@163.com
Guiting Chen,
gtchen2018@163.com

SPECIALTY SECTION
This article was submitted to Environmental Informatics and Remote Sensing, a section of the journal Frontiers in Earth Science

RECEIVED 10 January 2023
ACCEPTED 17 February 2023
PUBLISHED 13 March 2023

CITATION
Zhang C, Fan L, Chen G and Zeng X (2023), Efficient temporal high-order staggered-grid scheme with a dispersion-relation-preserving method for the scalar wave modeling.
Front. Earth Sci. 11:1141220.
doi: 10.3389/feart.2023.1141220

Efficient temporal high-order staggered-grid scheme with a dispersion-relation-preserving method for the scalar wave modeling

Chunlin Zhang[1]*, Liyong Fan[2], Guiting Chen[3]* and Xu Zeng[1]

[1]PetroChina Research Institute of Petroleum Exploration and Development, Beijing, China, [2]PetroChina Changqing Oilfield Company, Xian, China, [3]School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, China

Staggered-grid finite-difference (FD) method is widely used to solve the wave equation for the numerical seismic modeling, and it is a key step of the advanced seismic imaging and inversion problem. However, the conventional FD method is prone to instability and dispersion error due to the insufficient approximation accuracy. In this work, we propose an efficient temporal high-order finite-difference (FD) scheme with the cross-rhombus stencil. Compared with the standard cross-rhombus method, the new method has less computational cost due to we simplify the FD scheme. Moreover, the dispersion relation of the new method is easy to obtain the dispersion-relation-preserving (DRP) FD coefficients, which can significantly alleviate the spatial and temporal dispersion errors. Dispersion and stability analyses indicate that the new scheme has better performance in seismic modeling than the conventional method, and numerical experiments also indicate that the new scheme can effectively mitigate dispersion error and improve the numerical accuracy.

KEYWORDS

finite difference, staggered grid, simplified dispersion-relation-preserving scheme, cross-rhombus stencil, high-order approximation

1 Introduction

Staggered-grid finite-difference methods have been extensively applied in the seismic wave simulations due to their straightforward implementation and high computing efficiency (Kindelan et al., 1990; Moczo et al., 2000; Etgen and O¡¯Brien, 2007; Moczo et al., 2011; Moczo et al., 2014; Etemadsaeed et al., 2016; Liu et al., 2019; Zhang et al., 2022). High-order approximation for temporal derivatives in the staggered-grid FD scheme contributes to suppressing the temporal dispersion errors, and enhancing the stability with a large time step. However, the explicit high-order temporal derivative approximation in the FD scheme is always unstable (Chen, 2007; 2011). Generally, we use a second-order temporal approximation and an arbitrary even-order spatial approximation to solve the scalar wave equation.

To improve the temporal accuracy, Dablain (1986) proposed a new FD scheme based on the Lax-Wendroff approach, which can reach fourth-order accuracy in the temporal derivative approximation but has a limitation in the case of the large velocity contrast.

Chen (2007, 2011) further developed the fourth- and sixth-order schemes and analyzed the stability condition in the high-order cases. Alternatively, Liu and Sen (2009) proposed the time-space domain FD coefficients by incorporating the dispersion relation of the temporal and spatial terms. The time-space domain method can reach arbitrary even-order accuracy along with some specific propagation angles. However, it is still second-order accuracy along with other angles (Liu and Sen, 2009). To further improve the accuracy, Liu and Sen (2013) proposed a novel rhombus stencil. This new stencil with the time-space domain FD coefficients can reach arbitrary-order accuracy in both temporal and spatial approximations. However, the standard rhombus stencil is not a computationally friendly method for large-scale modeling, and it will increase the computational cost exponentially for the high-order cases. Afterwards, Tan and Huang (2014a),Tan and Huang (2014b) proposed an effective FD stencil with the sixth-order accuracy in the time approximation. Tan's stencil is similar to the rhombus stencil, but it involves fewer grid nodes outside the cross axis compared to the standard rhombus stencil, thus reducing the computational cost significantly. Wang E. et al. (2016) generally defined this stencil as the cross-rhombus stencil with arbitrary even-order temporal accuracy. The cross-rhombus stencil contains a large cross stencil and a small rhombus stencil. Among them, the small rhombus stencil increases the temporal accuracy and ensures computational efficiency, while the large cross stencil has a high-order spatial accuracy. Then, Ren et al. (2017) developed the cross-rhombus stencil in the staggered-grid FD scheme, and presented two methods for solving the FD coefficients. Wang et al. (2019) further developed the cross-rhombus stencil in the 3D case with the general cuboid grid.

To mitigate the dispersion error, the dispersion relation of the FD scheme should require many wavenumbers, because the spatial dispersion error usually comes from the high-wavenumber component. However, the conventional Taylor series expansion (TE) method for solving FD coefficient satisfies the dispersion relation near the zero wavenumbers, so it is prone to dispersion. The optimization method is a feasible way to obtain the FD coefficients (Liu, 2013; Zhang and Yao, 2013; Tan and Huang, 2014b; Chen et al., 2020) for mitigating dispersion, where the dispersion-relation-preserving method (Wang and Teixeira, 2003; Ye and Chu, 2005; Liang et al., 2015; Etemadsaeed et al., 2016; Chen et al., 2020) has been widely concerned because of its simplicity and easy implementation. The DRP-based method expands dispersion relation into an over-determined system associated with a series of wavenumbers and propagation angles, and then solves this over-determined system numerically to obtain the FD coefficients in the least square sense. The DRP-based FD coefficients satisfy a series of wavenumbers from low to high in the sense of the least square, thus the DRP-based method can effectively mitigate the temporal and spatial dispersion error.

The DRP-based FD coefficients have been successfully applied to the temporal high-order scheme (cross-rhombus stencil) in the regular grid Chen et al. (2020), in which the DRP-based coefficients can significantly mitigate dispersion error, while the cross-rhombus

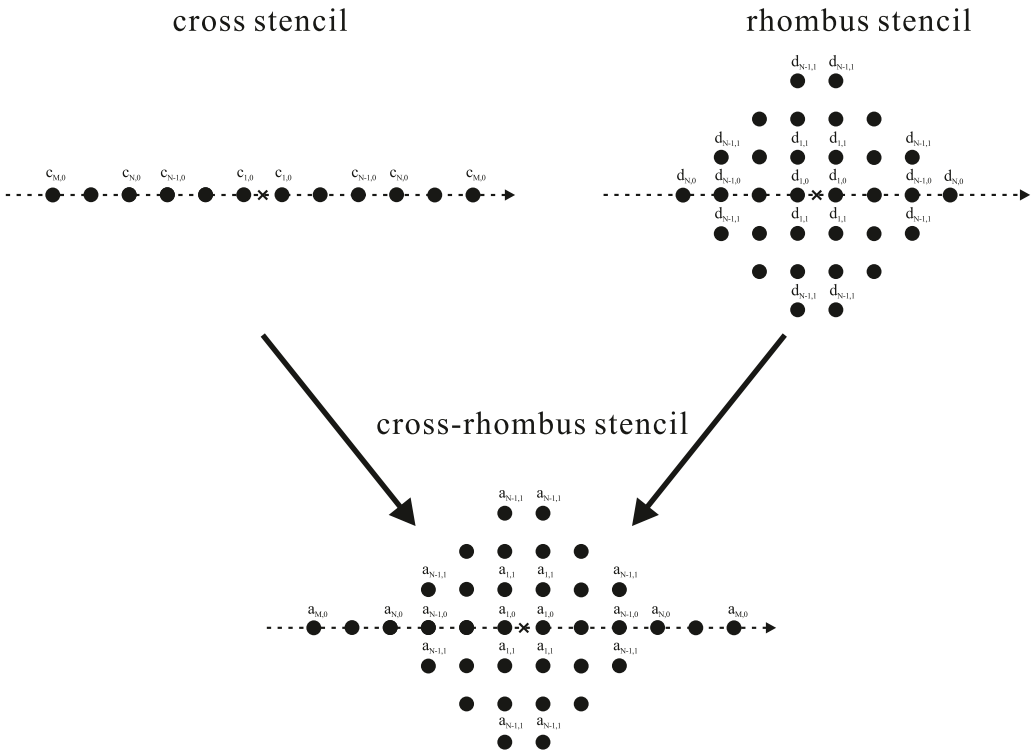


FIGURE 1
Illustrating the cross-rhombus stencil in the staggered grid.

TABLE 1 Abbreviation table of different FD methods used for dispersion analyses.

Abbreviations	FD coefficients	FD stencils
TE-C-S	TE-based space domain coefficients	Cross stencil
TE-C-TS	TE-based time-space domain coefficients	Cross stencil
TE-CR-TS	TE-based time-space domain coefficients	Cross-rhombus stencil (Standard)
DRP-CR-TS	DRP-based time-space domain coefficients	Cross-rhombus stencil (Proposed)

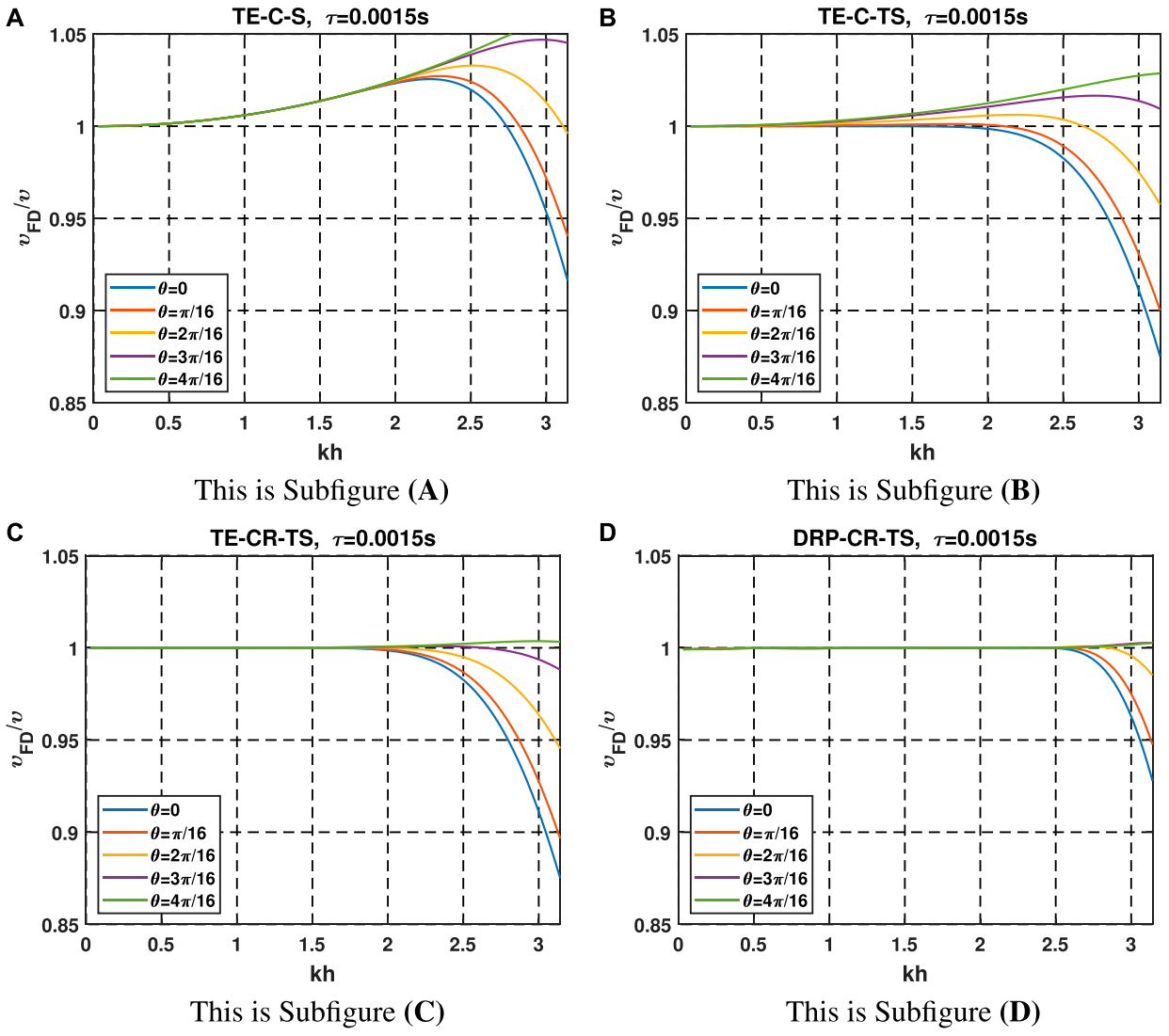


FIGURE 2
Dispersion curves of the four methods with a moderate time step $\tau = 0.0015$ s **(A)** TE-C-S method. **(B)** TE-C-TS method. **(C)** TE-CR-TS method. **(D)** DRP-CR-TS method.

stencil can effectively improve the temporal approximation accuracy. However, for the staggered-grid FD scheme, the DRP-based coefficients cannot be directly obtained because the dispersion relation is difficult to be extended into an over-determined system. Liang et al. (2018) has presented a special FD scheme with a high computational efficiency, in which the second-order FD operator is used to approximate some partial derivatives rather than the global high-order FD operator. And such replacement simplifies the dispersion relation into a form of the linear equation. Motivated by (Liang et al., 2018; Zhou et al., 2022), we propose a general simplified FD scheme for the temporal high-order modeling with a cross-rhombus stencil. The new scheme contains a cross stencil

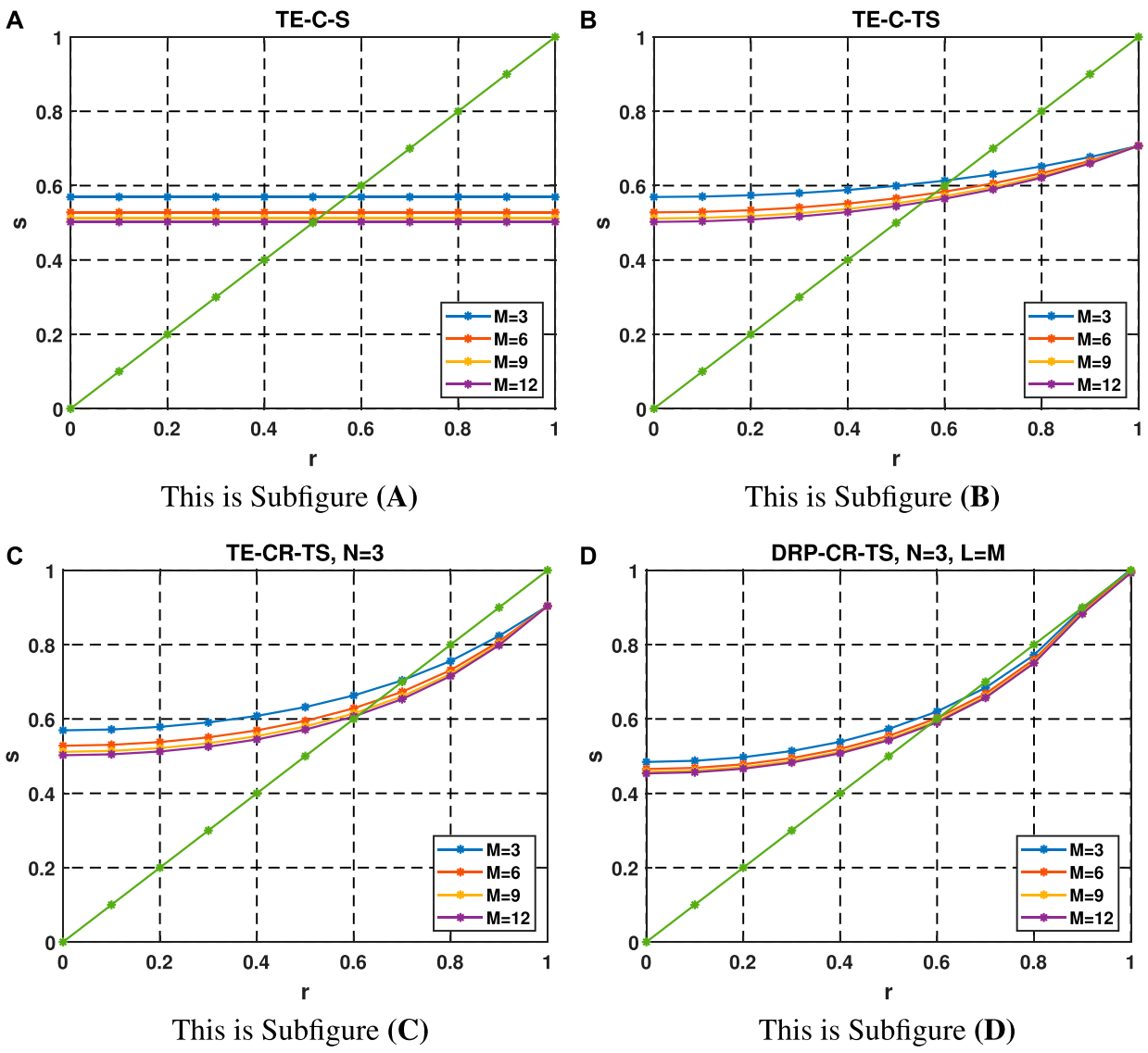


FIGURE 3
Stability factors of the four methods with different r and M, where $N = 3$ and $L = M$ for the cross-rhombus stencil. The green curve represents $s(r) = r$. **(A)** TE-C-S method. **(B)** TE-C-TS method. **(C)** TE-CR-TS method. **(D)** DRP-CR-TS method.

with the analytic FD coefficients and a cross-rhombus stencil with the DRP-based coefficients for different partial derivatives. The cross stencil can simplify the dispersion relation, which makes it easy to construct the over-determined system. The cross-rhombus stencil can make the FD scheme maintain a high-order temporal approximation. The dispersion relation of our new FD scheme can be expanded to an over-determined system with a series of wavenumbers and angles. Solving this over-determined system by the numerical methods (Wang et al., 2014; Wang et al., 2016 Y.; Chen et al., 2020; Wu et al., 2020; Li et al., 2022), we obtain the DRP-based FD coefficients. Therefore, the new FD scheme has three advantages: 1. The DRP-based FD coefficients can effectively mitigate the dispersion error; 2. It still has a temporal higher-order approximation accuracy; 3. The computational cost of the proposed method is significantly reduced compared with the standard cross-rhombus scheme.

2 Methods

2.1 Review of the standard cross-rhombus scheme

The 2D first-order acoustic wave equations with the constant density are

$$\frac{\partial p}{\partial t} + K\nabla \cdot \mathbf{v} = 0, \qquad \frac{\partial \mathbf{v}}{\partial t} + \frac{1}{\rho}\nabla p = 0, \tag{1}$$

where $K = \rho v^2$ is the bulk modulus, $\rho(x,z)$ is the density and $v(x,z)$ is the velocity. $p(x,z,t)$ is the pressure, and $\mathbf{v} = [v_x, v_z]^T$ is the particle velocity vector. The staggered-grid FD scheme with the

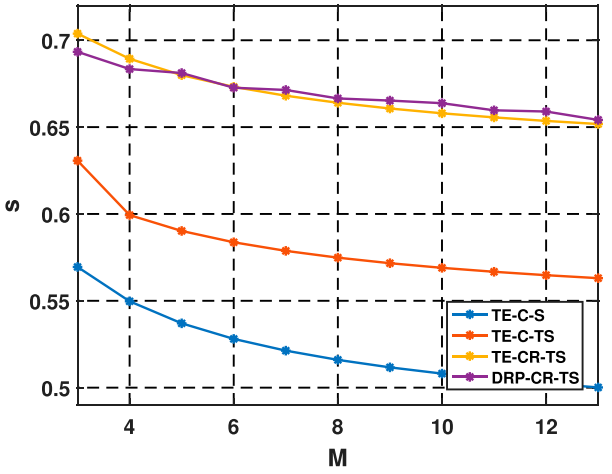


FIGURE 4
Maximum stability factors of the four methods satisfy $r \le s$ for different M, where $N = 3$ and $L = M$ for the cross-rhombus stencil.

cross-rhombus stencil for above equations is

$$\begin{aligned} p_{i,j}^{l} &= p_{i,j}^{l-1} - K\tau\left(D_x^{CR} v_{x\ i-1/2,j}^{l-1/2} + D_z^{CR} v_{z\ i,j-1/2}^{l-1/2}\right), \\ v_{x\ i+1/2,j}^{l+1/2} &= v_{x\ i+1/2,j}^{l-1/2} - \frac{1}{\rho}\tau H_x^{CR} p_{i,j}^{l}, \\ v_{z\ i,j+1/2}^{l+1/2} &= v_{z\ i,j+1/2}^{l-1/2} - \frac{1}{\rho}\tau H_z^{CR} p_{i,j}^{l}. \end{aligned} \tag{2}$$

Here, $p_{i,j}^{l} = p(x+ih, z+jh, t+l\tau)$, h is the grid spacing and τ is the time step. The FD operators H_x^{CR} and D_x^{CR} are

$$\begin{aligned} \frac{\partial p}{\partial x} \approx H_x^{CR} p_{0,0} = \frac{1}{h}\Bigg[&\sum_{m=1}^{M} a_{m,0}\left(p_{m,0} - p_{-m+1,0}\right) + \sum_{m=1}^{N-1} \\ &\times \sum_{n=1}^{N-m} a_{m,n}\left(p_{m,n} - p_{-m+1,n} + p_{m,-n} - p_{-m+1,-n}\right)\Bigg]. \end{aligned} \tag{3}$$

and

$$\begin{aligned} \frac{\partial v_x}{\partial x} \approx D_x^{CR} v_{-1/2,0} = \frac{1}{h}\Bigg[&\sum_{m=1}^{M} b_{m,0}\left(v_{x,m-1/2,0} - v_{x,-m+1,0}\right) \\ &+ \sum_{m=1}^{N-1}\sum_{n=1}^{N-m} b_{m,n}\left(v_{x,m-1/2,n} - v_{x,-m+1/2,n}\right. \\ &\left. + v_{x,m-1/2,-n} - v_{x,-m+1,-n}\right)\Bigg]. \end{aligned} \tag{4}$$

The superscript CR represents the cross-rhombus stencil composed of a standard cross stencil and a rhombus stencil (as shown in Figure 1). $a_{m,n}$ and $b_{m,n}$ represent the FD coefficients of the operators H_x^{CR} and D_x^{CR}, respectively. M and N are the spatial and temporal operator length parameters, respectively. Generally, when temporal operator length $N > 3$, the accuracy increases far less than the increase of the calculation cost. Thus, we recommend that $N = 3$ is enough. M and N represents the $(2M)$th-order accuracy in space and $(2N)$th-order accuracy in time respectively. And the FD operators along the z-axis (H_z^{CR} and D_z^{CR}) can be defined in the same way.

Assuming plane wave propagating in the grid, we let

$$\begin{aligned} p_{m,n}^{l} &= P_{0,0}^{0} e^{\mathrm{i}(k_x mh + k_z nh - \omega l\tau)}, \\ v_{x\ m+1/2,n}^{l+1/2} &= v_{x\ 1/2,0}^{1/2} e^{\mathrm{i}(k_x mh + k_z nh - \omega l\tau)}, \\ v_{z\ m,n+1/2}^{l+1/2} &= v_{z\ 0,1/2}^{1/2} e^{\mathrm{i}(k_x mh + k_z nh - \omega l\tau)}, \end{aligned} \tag{5}$$

where $k_x = k\cos(\theta)$ and $k_z = k\sin(\theta)$ are the wavenumbers in x- and z-axes, respectively. θ is a propagation angle of the plane wave, ω is the angular frequency and $\mathrm{i} = \sqrt{-1}$. Substituting Eq. 5 into Eq. 2, we obtain

$$\begin{aligned} &\left[\sum_{m=1}^{M} a_{m,0}\sin\left((m-0.5)k_x h\right) + 2\sum_{m=1}^{N} \right. \\ &\quad \left. \times \sum_{n=1}^{N-m} a_{m,n}\sin\left((m-0.5)k_x h\right)\cos\left(nk_z h\right)\right] \\ &\quad * \left[\sum_{m=1}^{M} b_{m,0}\sin\left((m-0.5)k_x h\right) + 2\sum_{m=1}^{N} \right. \\ &\quad \left. \times \sum_{n=1}^{N-m} b_{m,n}\sin\left((m-0.5)k_x h\right)\cos\left(nk_z h\right)\right] \\ &\quad + \left[\sum_{m=1}^{M} a_{m,0}\sin\left((m-0.5)k_z h\right) + 2\sum_{m=1}^{N} \right. \\ &\quad \left. \times \sum_{n=1}^{N-m} a_{m,n}\sin\left((m-0.5)k_z h\right)\cos\left(nk_x h\right)\right] \\ &\quad * \left[\sum_{m=1}^{M} b_{m,0}\sin\left((m-0.5)k_z h\right) + 2\sum_{m=1}^{N} \right. \\ &\quad \left. \times \sum_{n=1}^{N-m} b_{m,n}\sin\left((m-0.5)k_z h\right)\cos\left(nk_x h\right)\right] \\ &\quad = \frac{1-\cos(\omega\tau)}{2r^2}. \end{aligned} \tag{6}$$

Here, $r = v\tau/h$ is the Courant number. Generally, the operators H^{CR} and D^{CR} adopt the same FD coefficients, i.e.

$$a_{m,n} = b_{m,n}. \tag{7}$$

Then, Equation 6 can be rewritten as

$$\begin{aligned} &\left[\sum_{m=1}^{M} a_{m,0}\sin\left((m-0.5)k_x h\right) + 2\sum_{m=1}^{N} \right. \\ &\quad \left. \times \sum_{n=1}^{N-m} a_{m,n}\sin\left((m-0.5)k_x h\right)\cos\left(nk_z h\right)\right]^2 \\ &\quad + \left[\sum_{m=1}^{M} a_{m,0}\sin\left((m-0.5)k_z h\right) + 2\sum_{m=1}^{N} \right. \\ &\quad \left. \times \sum_{n=1}^{N-m} a_{m,n}\sin\left((m-0.5)k_z h\right)\cos\left(nk_x h\right)\right]^2 \\ &\quad = \frac{1-\cos(\omega\tau)}{2r^2}. \end{aligned} \tag{8}$$

Eq. 8 represents the time-space domain dispersion relation of the staggered-grid FD scheme with the cross-rhombus stencil. Using the Taylor series to expand the trigonometric functions with respect to the propagation angle θ, we can obtain the time-space domain TE-based FD coefficients (Ren et al., 2017). The cross-rhombus stencil with the TE-based FD coefficients can achieve arbitrary even-order temporal accuracy, thus mitigating the temporal dispersion error significantly.

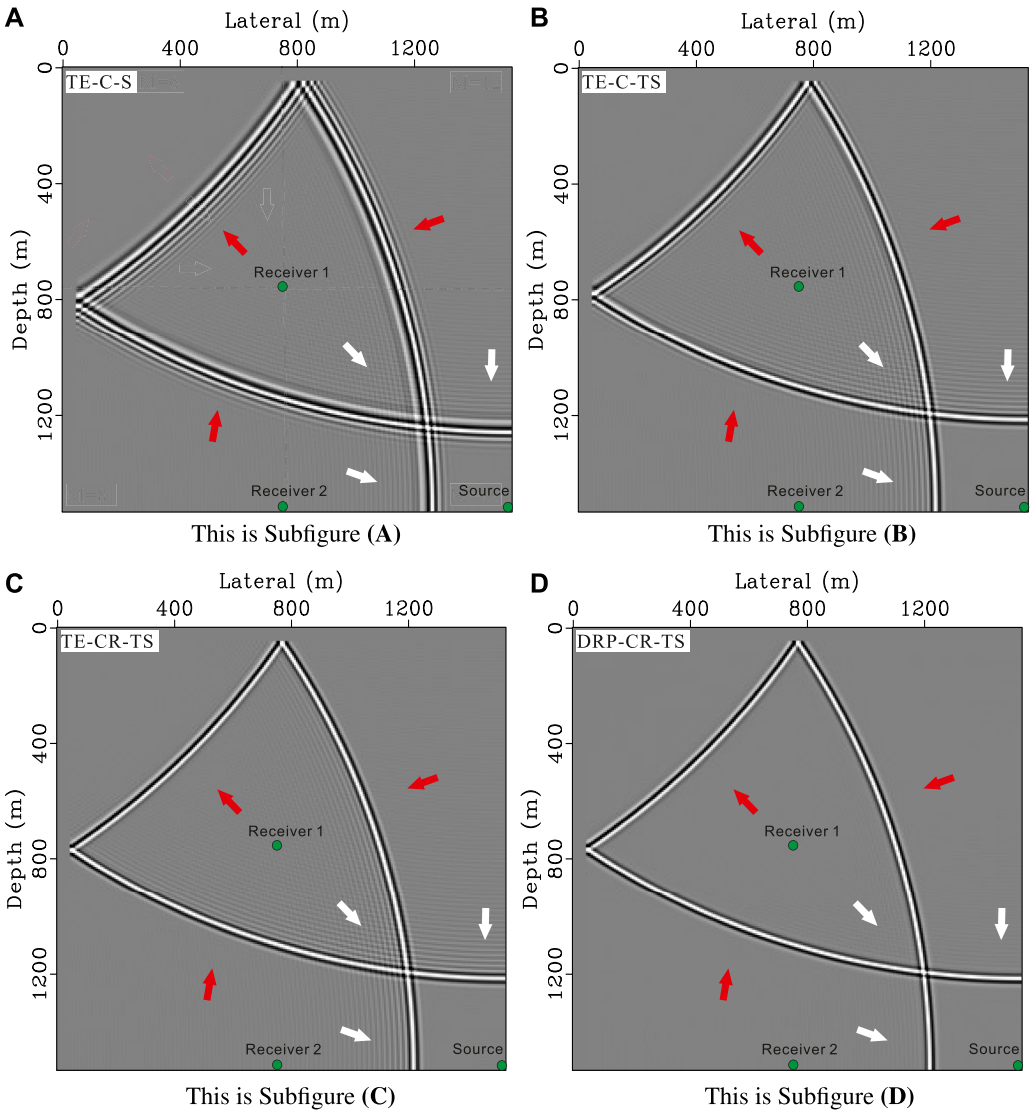


FIGURE 5
Snapshots of the four methods with the time step $\tau = 0.001s$, where $M = 8$, $N = 3$, $L = M$, $\upsilon = 1500\ m/s$ and $h = 6m$. The main frequency of the Ricker wavelet is $40Hz$. **(A)** TE-C-S method. **(B)** TE-C-TS method. **(C)** TE-CR-TS method. **(D)** DRP-CR-TS method.

2.2 A new simplified staggered-grid FD scheme with the cross-rhombus stencil

The TE-based FD coefficients satisfy the dispersion relation within a limited wavenumber bandwidth, resulting in the high-wavenumber components of seismic wavefield are prone to the spatial dispersion. However, the dispersion relation of the standard staggered-grid scheme is a quadratic equation, which is difficult to expand into the over-determined system for solving DRP-based FD coefficients. In this part, we develop a new simplified staggered-grid FD scheme. The new scheme can not only easily obtain the over-determined system for the DRP-based coefficients, but also greatly reduce the computational cost.

The FD operators D^{CR} and H^{CR} use same coefficients ($a_{m,n} = b_{m,n}$), which causes the dispersion relation to be a second-order non-linear equation. To obtain a simple dispersion relation, we propose a new simplified FD scheme as

$$\begin{aligned} p_{i,j}^{l} &= p_{i,j}^{l-1} - K\tau\left(D_x^C v_{x\ i-1/2,j}^{l-1/2} + D_z^C v_{z\ i,j-1/2}^{l-1/2}\right), \\ v_{x\ i+1/2,j}^{l-1/2} &= v_{x\ i+1/2,j}^{l-3/2} - \frac{1}{\rho}\tau H_x^{CR} p_{i,j}^{l-1}, \\ v_{z\ i,j+1/2}^{l-1/2} &= v_{z\ i,j+1/2}^{l-3/2} - \frac{1}{\rho}\tau H_z^{CR} p_{i,j}^{l-1}. \end{aligned} \tag{9}$$

Here, the superscript C represents the cross stencil. For example, the FD operator D_x^C can be defined as

$$D_x^C v_{-1/2,0} = \frac{1}{h}\sum_{l=1}^{L} b_{l,0}\left(v_{l-1/2,0} - v_{-l+1/2,0}\right). \tag{10}$$

We use the cross-stencil-based operators D^C to replace part of the operator D^{CR} in the new scheme. Thus, the new scheme contains a cross-rhombus stencil and a cross stencil for different partial derivatives.

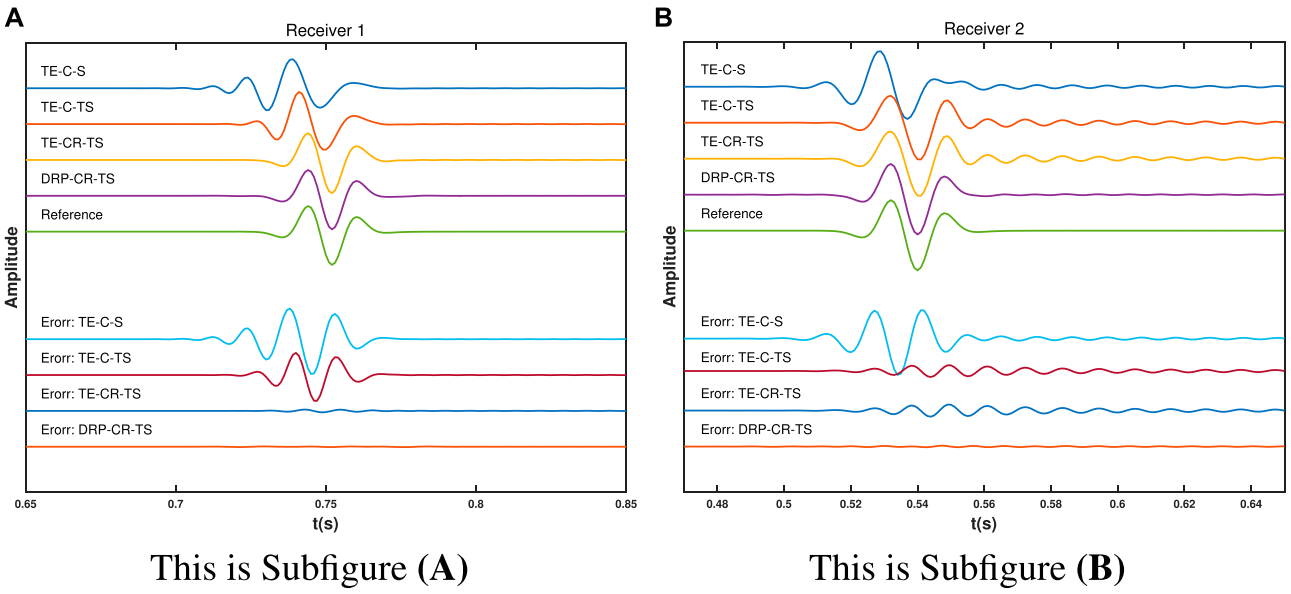


FIGURE 6
Seismic records of the four methods at Receiver 1 and Receiver 2. The reference traces represented by the green curves are obtained by the high-order FD scheme under the fine grid. **(A)** Seismic records at Receiver 1. **(B)** Seismic records at Receiver 2.

TABLE 2 Relative errors of the four methods at Receiver 1 and Receiver 2.

Methods	Relative Errors (Pa)	
	Receiver 1	Receiver 2
TE-C-S	0.3931	0.4323
TE-C-TS	0.2436	0.2092
TE-CR-TS	0.0366	0.2054
DRP-CR-TS	0.0113	0.0458

For convenience, D^C adopts the analytic time-space domain FD coefficients

$$b_{l,0} = \frac{(-1)^{l+1}}{2l-1} \prod_{1\le n\le L,\ n\neq l} \times \left| \frac{(2n-1)^2 - r^2}{(2n-1)^2 - (2l-1)^2} \right| \quad (l = 1,2,\ldots,L). \tag{11}$$

Here, L represents the length of the analytic FD operator. And the analytical cross stencil simplifies the original second-order dispersion relation to a linear form. The new dispersion relation can be easily extended to an over-determined linear system for solving the wide-bandwidth FD coefficients. Moreover, the analytical cross stencil has less computational cost compared with the standard cross-rhombus scheme, especially in the high-order cases.

2.3 Determining FD coefficients of the new stencil by the DRP-based method

In this part, we present the method for solving the DRP-based FD coefficients of the new scheme. We assume that the plane wave propagating in the grid, and then substitute Eq. 5 into Eq. 9, yield the new dispersion relation

$$\begin{aligned} &\left[\sum_{m=1}^{M} a_{m,0} \sin\left((m-0.5)k_x h\right) + 2\sum_{m=1}^{N}\sum_{n=1}^{N-m} a_{m,n} \sin \right. \\ &\left. \times \left((m-0.5)k_x h\right)\cos\left(nk_z h\right) \right] \sum_{l=1}^{L} b_{l,0} \sin\left((l-0.5)k_x h\right) \\ &+ \left[\sum_{m=1}^{M} a_{m,0} \sin\left((m-0.5)k_z h\right) + 2\sum_{m=1}^{N}\sum_{n=1}^{N-m} a_{m,n} \sin \right. \\ &\left. \times \left((m-0.5)k_z h\right)\cos\left(nk_x h\right) \right] \sum_{l=1}^{L} b_{l,0} \sin\left((l-0.5)k_z h\right) \\ &= \frac{1-\cos(\omega\tau)}{2r^2}. \end{aligned} \tag{12}$$

Clearly, the new dispersion relation can be easily extended to the linear system satisfying a series of wavenumbers and propagation angles. It is worth noting that if the FD operator H uses cross stencil, for example: H^C and D^{CR} are applied to Equation 9, the dispersion relation does not change. And the cross stencil is applied to the FD operator H or D, the corresponding FD schemes are equivalent.

Following our previous work (Chen et al., 2020), we define a new function $\psi_{m,\beta,\theta}$ to represent the weights of FD coefficients $a_{m,0}$ in

Eq. 12. Then $\psi_{m,\beta,\theta}$ can be denoted as

$$\begin{aligned}\psi_{m,\beta,\theta} = &\sin((m-0.5)\beta\cos(\theta))\sum_{l=1}^{L} b_{l,0}\sin((l-0.5)\beta\cos(\theta))\\ &+\sin((m-0.5)\beta\sin(\theta))\sum_{l=1}^{L} b_{l,0}\\ &\times\sin((l-0.5)\beta\sin(\theta)).\end{aligned} \tag{13}$$

Here, $\beta = kh$ and θ represents the propagation angle. Similarly, we define another function $\chi_{m,n,\beta,\theta}$ to represent the weights of $a_{m,n}$. The function $\chi_{m,n,\beta,\theta}$ is defined as

$$\begin{aligned}\chi_{m,n,\beta,\theta} = &2\sin((m-0.5)\beta\cos(\theta))\cos(n\beta\sin(\theta))\sum_{l=1}^{L} b_{l,0}\\ &\times\sin((l-0.5)\beta\cos(\theta)) + 2\sin((m-0.5)\beta\sin(\theta))\\ &\times\cos(n\beta\cos(\theta))\sum_{l=1}^{L} b_{l,0}\sin((l-0.5)\beta\sin(\theta)).\end{aligned} \tag{14}$$

Since $a_{m,n} = a_{n,m}$, we define a new function

$$\varphi_{m,n,\beta,\theta} = \begin{cases}\chi_{m,n,\beta,\theta} & (m=n)\\ \chi_{m,n,\beta,\theta} + \chi_{n,m,\beta,\theta} & (m\neq n).\end{cases} \tag{15}$$

Therefore, the dispersion relation (Eq. 12) of the new FD scheme can be rewritten as

$$\sum_{m=1}^{M} a_{m,0}\psi_{m,\beta,\theta} + \sum_{m=1}^{N}\sum_{n=m}^{N-m} a_{m,n}\varphi_{m,n,\beta,\theta} = \frac{1-\cos(\omega\tau)}{2r^2}. \tag{16}$$

Then we extend $\psi_{m,\beta,\theta}$ to a matrix involving a series of β and a fixed angle θ, and the matrix is

$$\mathbf{A}(\theta) = \begin{bmatrix}\psi_{1,\beta_1,\theta} & \psi_{2,\beta_1,\theta} & \cdots & \psi_{M,\beta_1,\theta}\\ \psi_{1,\beta_2,\theta} & \psi_{2,\beta_2,\theta} & \cdots & \psi_{M,\beta_2,\theta}\\ \vdots & \vdots & \vdots & \vdots\\ \psi_{1,\beta_\xi,\theta} & \psi_{2,\beta_\xi,\theta} & \cdots & \psi_{M,\beta_\xi,\theta}\end{bmatrix}. \tag{17}$$

Where $\beta_i = \beta_{max}/\xi^* i$, $\beta_{max} = 2\pi f_{max}/v$ with respect to the maximum frequency of the seismic wavefield Chen et al. (2022) and π is the circular constant.

Similarly, we extend the function $\varphi_{n,\beta,\theta}$ to the matrix.

$$\mathbf{B}(\theta) = \begin{bmatrix}\varphi_{1,1,\beta_1,\theta} & \cdots & \varphi_{1,N-1,\beta_1,\theta} & \varphi_{2,2,\beta_1,\theta} & \cdots & \varphi_{2,N-2,\beta_1,\theta} & \cdots & \varphi_{N/2,N/2,\beta_1,\theta}\\ \varphi_{1,1,\beta_2,\theta} & \cdots & \varphi_{1,N-1,\beta_2,\theta} & \varphi_{2,2,\beta_2,\theta} & \cdots & \varphi_{2,N-2,\beta_2,\theta} & \cdots & \varphi_{N/2,N/2,\beta_2,\theta}\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ \varphi_{1,1,\beta_\xi,\theta} & \cdots & \varphi_{1,N-1,\beta_\xi,\theta} & \varphi_{2,2,\beta_\xi,\theta} & \cdots & \varphi_{2,N-2,\beta_\xi,\theta} & \cdots & \varphi_{N/2,N/2,\beta_\xi,\theta}\end{bmatrix}. \tag{18}$$

Note that if N is an odd number, the matrix $\mathbf{B}(\theta)$ is.

$$\mathbf{B}(\theta) = \begin{bmatrix}\varphi_{1,1,\beta_1,\theta} & \cdots & \varphi_{1,N-1,\beta_1,\theta} & \varphi_{2,2,\beta_1,\theta} & \cdots & \varphi_{2,N-2,\beta_1,\theta} & \cdots & \varphi_{(N-1)/2,(N-1)/2,\beta_1,\theta} & \varphi_{(N-1)/2,(N+1)/2,\beta_1,\theta}\\ \varphi_{1,1,\beta_2,\theta} & \cdots & \varphi_{1,N-1,\beta_2,\theta} & \varphi_{2,2,\beta_2,\theta} & \cdots & \varphi_{2,N-2,\beta_2,\theta} & \cdots & \varphi_{(N-1)/2,(N-1)/2,\beta_2,\theta} & \varphi_{(N-1)/2,(N+1)/2,\beta_2,\theta}\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ \varphi_{1,1,\beta_\xi,\theta} & \cdots & \varphi_{1,N-1,\beta_\xi,\theta} & \varphi_{2,2,\beta_\xi,\theta} & \cdots & \varphi_{2,N-2,\beta_\xi,\theta} & \cdots & \varphi_{(N-1)/2,(N-1)/2,\beta_\xi,\theta} & \varphi_{(N-1)/2,(N+1)/2,\beta_\xi,\theta}\end{bmatrix}. \tag{19}$$

The right-hand side of Eq. 16 also be expanded to the matrix

$$\mathbf{D}(\theta) = \begin{bmatrix}r^{-2}\left[-2+2\cos(\beta_1 r)\right]\\ r^{-2}\left[-2+2\cos(\beta_2 r)\right]\\ \vdots\\ r^{-2}\left[-2+2\cos(\beta_\xi r)\right]\end{bmatrix}. \tag{20}$$

Then the dispersion relation satisfying ξ wavenumbers and ζ angles can be expressed as

$$\begin{bmatrix}\mathbf{A}(\theta_1) & \mathbf{B}(\theta_1)\\ \mathbf{A}(\theta_2) & \mathbf{B}(\theta_2)\\ \vdots & \vdots\\ \mathbf{A}(\theta_\zeta) & \mathbf{B}(\theta_\zeta)\end{bmatrix}\begin{bmatrix}a_{0,0}\\ a_{1,0}\\ \vdots\\ a_{N-1,1}\end{bmatrix} = \begin{bmatrix}\mathbf{D}(\theta_1)\\ \mathbf{D}(\theta_2)\\ \vdots\\ \mathbf{D}(\theta_\zeta)\end{bmatrix}. \tag{21}$$

This over-determined system has $\zeta\times\xi$ rows and $1+M+N^2/4$ columns with even number N and $1+M+(N-1)^2/4$ columns with odd number N. We can easily obtain the DRP-based FD coefficients by solving this over-determined system. We also introduce the simplified FD scheme in the 3D case, the details are shown in the **Supplementary Material S1**.

3 Numerical dispersion and stability analyses

3.1 Numerical dispersion analysis

In this part, we analyse the dispersion characteristics of our new scheme. The phase velocity of the new FD scheme can be expressed as

$$v_{\mathrm{FD}} = \frac{1}{k\tau}\arccos(1-2qr^2), \tag{22}$$

where q is

$$q = \sum_{m=1}^{M} a_{m,0}\psi_{m,\beta,\theta} + \sum_{m=1}^{N}\sum_{n=m}^{N-m} a_{m,n}\phi_{m,n,\beta,\theta}. \tag{23}$$

Thus, the dispersion parameter δ of the new FD scheme is defined as

$$\delta = \frac{v_{\mathrm{FD}}}{v} = \frac{1}{rkh}\arccos(1-2qr^2). \tag{24}$$

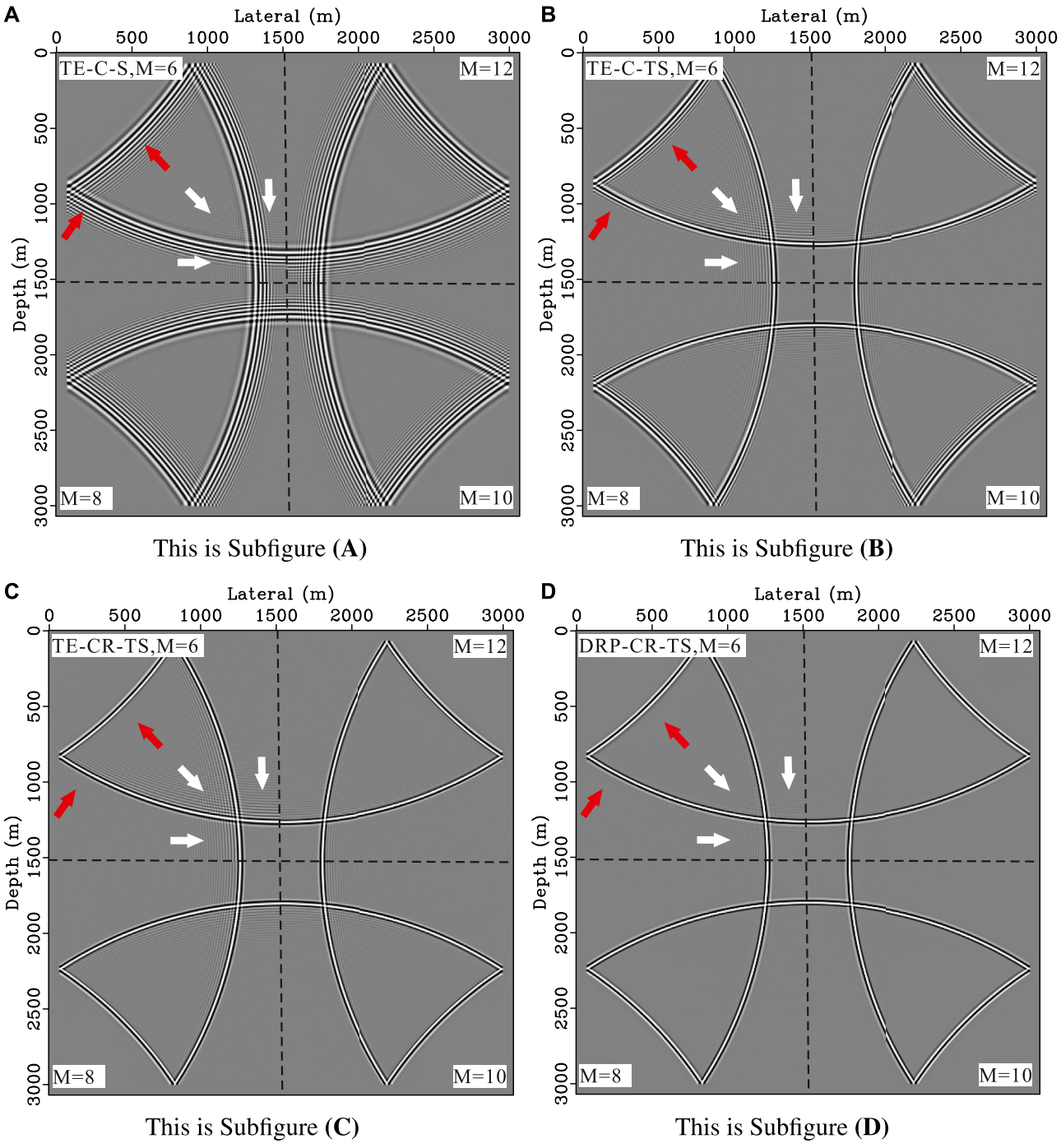


FIGURE 7
Snapshots of the four methods for different M, where $\tau = 0.0015s$, $\upsilon = 1500\ m/s$, $h = 6m$, and $N = 3$, $L = M$. The main frequency of the Ricker wavelet is $40Hz$. The four quadrants represent the snapshots of the same FD method for different $M = 6, 8, 10, 12$, respectively. **(A)** TE-C-S method. **(B)** TE-C-TS method. **(C)** TE-CR-TS method. **(D)** DRP-CR-TS method.

If δ is not equal to 1, the FD scheme suffers from the numerical dispersion, i.e., has the spatial dispersion error ($\delta < 1$) or temporal dispersion error ($\delta > 1$). We analyze and compare the dispersion parameter δ of the new FD scheme with the other three methods, and the abbreviations of these methods are listed in Table 1. The dispersion curves of δ varying with the kh are shown in Figure 2. It can be seen that the cross-stencil-based FD schemes (TE-C-S and TE-C-TS methods) have obviously temporal dispersion error ($\delta > 1$). The corresponding temporal dispersion of the cross-rhombus stencil (TE-CR-TS and DRP-CR-TS methods) is alleviated (Figures 2C, D) due to the temporal high-order approximation. It is worth noting that the proposed method (DRP-CR-TS method)

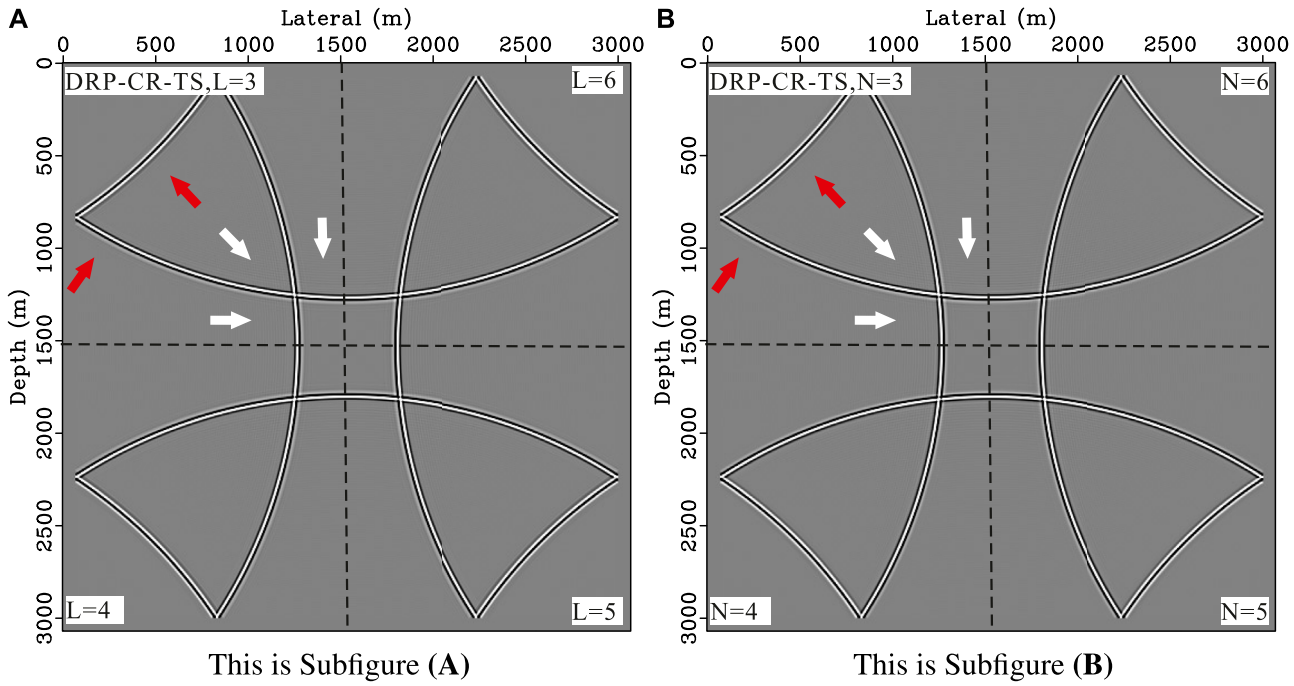


FIGURE 8
Snapshots of the simplified scheme for different L and N, where $M = 8$, $\tau = 0.0015s$, $\upsilon = 1500\ m/s$ and $h = 6m$. The main frequency of the Ricker wavelet is $40Hz$. **(A)** DRP-CR-TS method with different L, here $N = 3$. **(B)** DRP-CR-TS method with different N, here we set $L = M$.

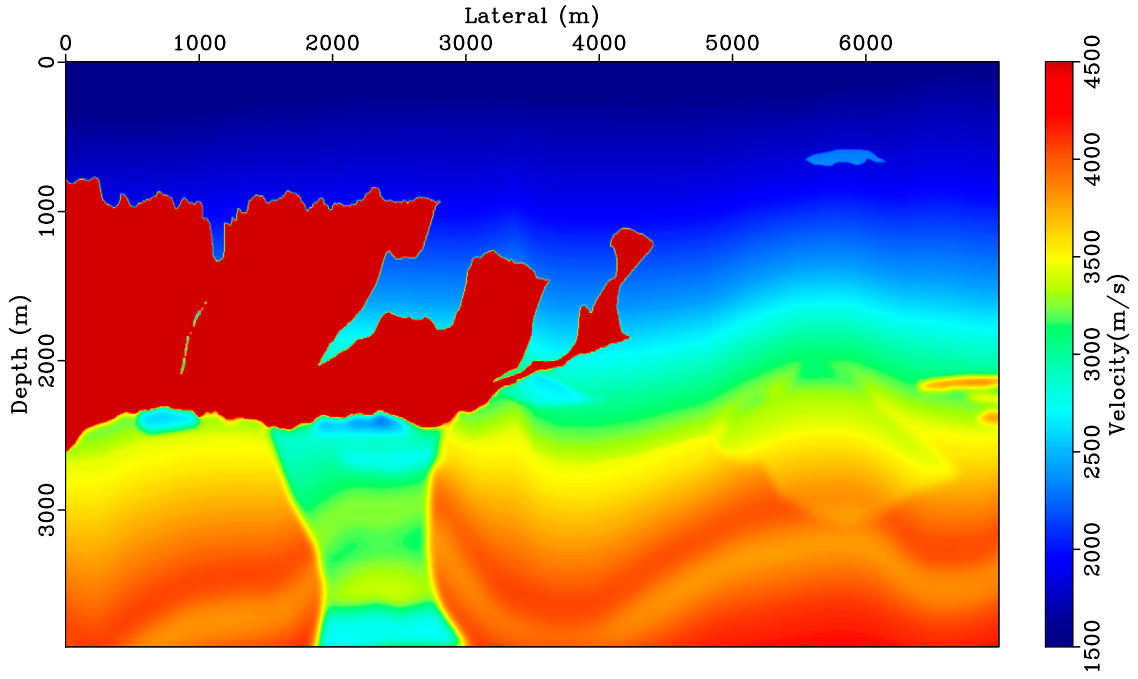


FIGURE 9
BP model with 560×1000 grid nodes and grid spacing $h = 7.5$ m, the variation of velocities from $1500\ m/s$ to $4500\ m/s$.

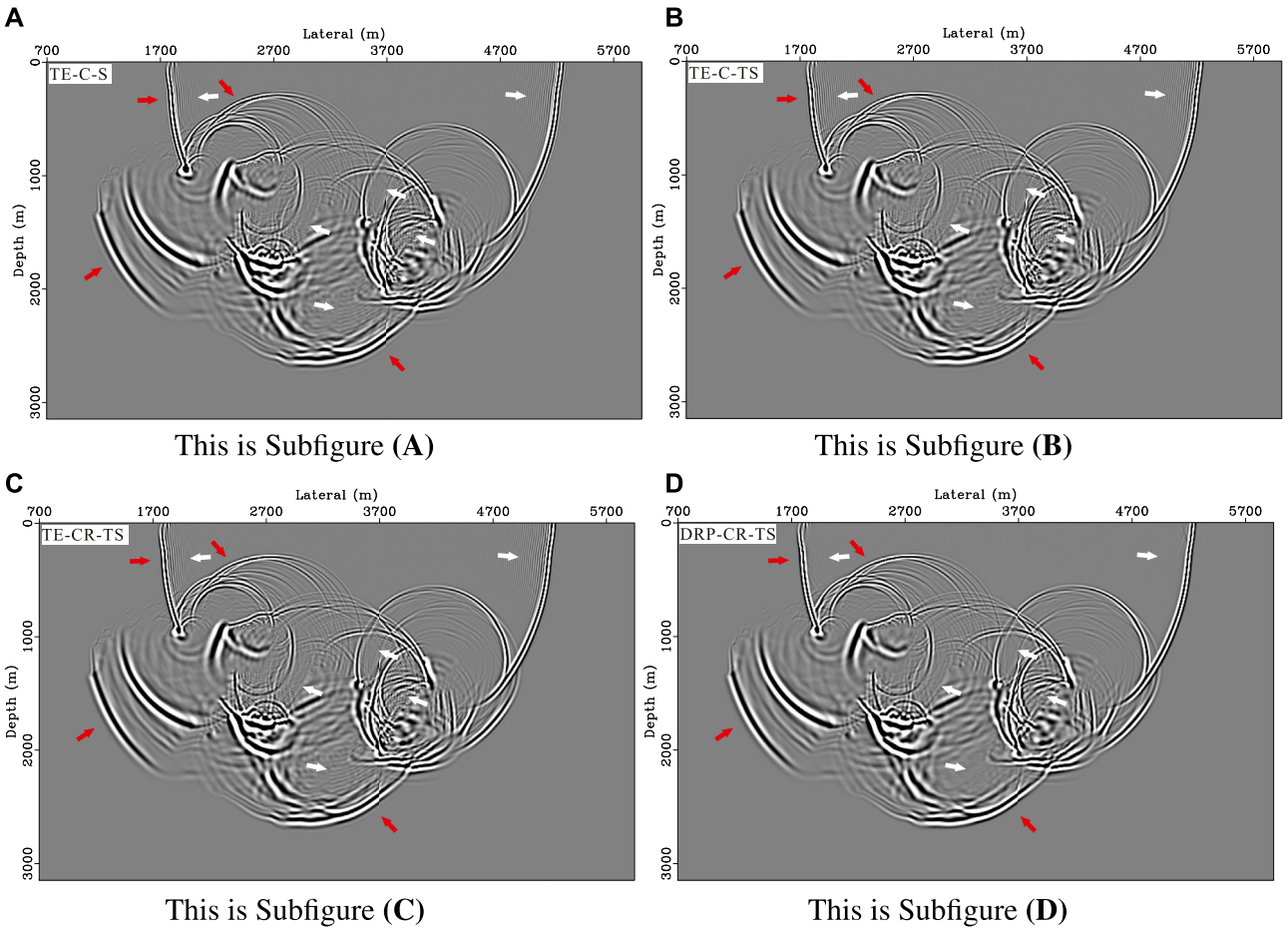


FIGURE 10
Snapshots of the four methods, here we set $M = 8$, $N = 3$, $L = M$ and $\tau = 0.007s$ **(A)** TE-C-S method. **(B)** TE-C-TS method. **(C)** TE-CR-TS method. **(D)** DRP-CR-TS method.

satisfies the widest range of the wavenumber (kh), which can mitigate the spatial dispersion error considerably.

3.2 Stability analysis

According to the dispersion relation of the new FD scheme, we obtain

$$\cos(\omega\tau) = 1 - 2qr^2. \tag{25}$$

It is clear that

$$-1 \le \cos(\omega\tau) \le 1. \tag{26}$$

Then, we obtain

$$q \ge 0 \tag{27}$$

and

$$r \le \sqrt{1/q}. \tag{28}$$

We consider the Nyquist wavenumber, that is

$$k_x h = k_z h = \pi. \tag{29}$$

Substituting Eq. 29 into (28), we obtain the stability condition

$$r \le \left[2\left(\sum_{m=1}^{M} (-1)^{m+1} a_{m,0} + 2\sum_{m=1}^{N-1} \times \sum_{n=1}^{N-m} (-1)^{m+n+1} a_{m,n} \right) \sum_{l=1}^{L} (-1)^{l+1} b_{l,0} \right]^{-\frac{1}{2}}. \tag{30}$$

We denote the right-hand side of inequation (30) as the stability factor

$$s = \left[2\left(\sum_{m=1}^{M} (-1)^{m+1} a_{m,0} + 2\sum_{m=1}^{N-1} \times \sum_{n=1}^{N-m} (-1)^{m+n+1} a_{m,n} \right) \sum_{l=1}^{L} (-1)^{l+1} b_{l,0} \right]^{-\frac{1}{2}}, \tag{31}$$

where the stability factor s is related to the FD coefficients $a_{m,n}$ and $b_{l,0}$, and these FD coefficients are determined by the Courant number r. In the following, we analyze the stability factors s varying

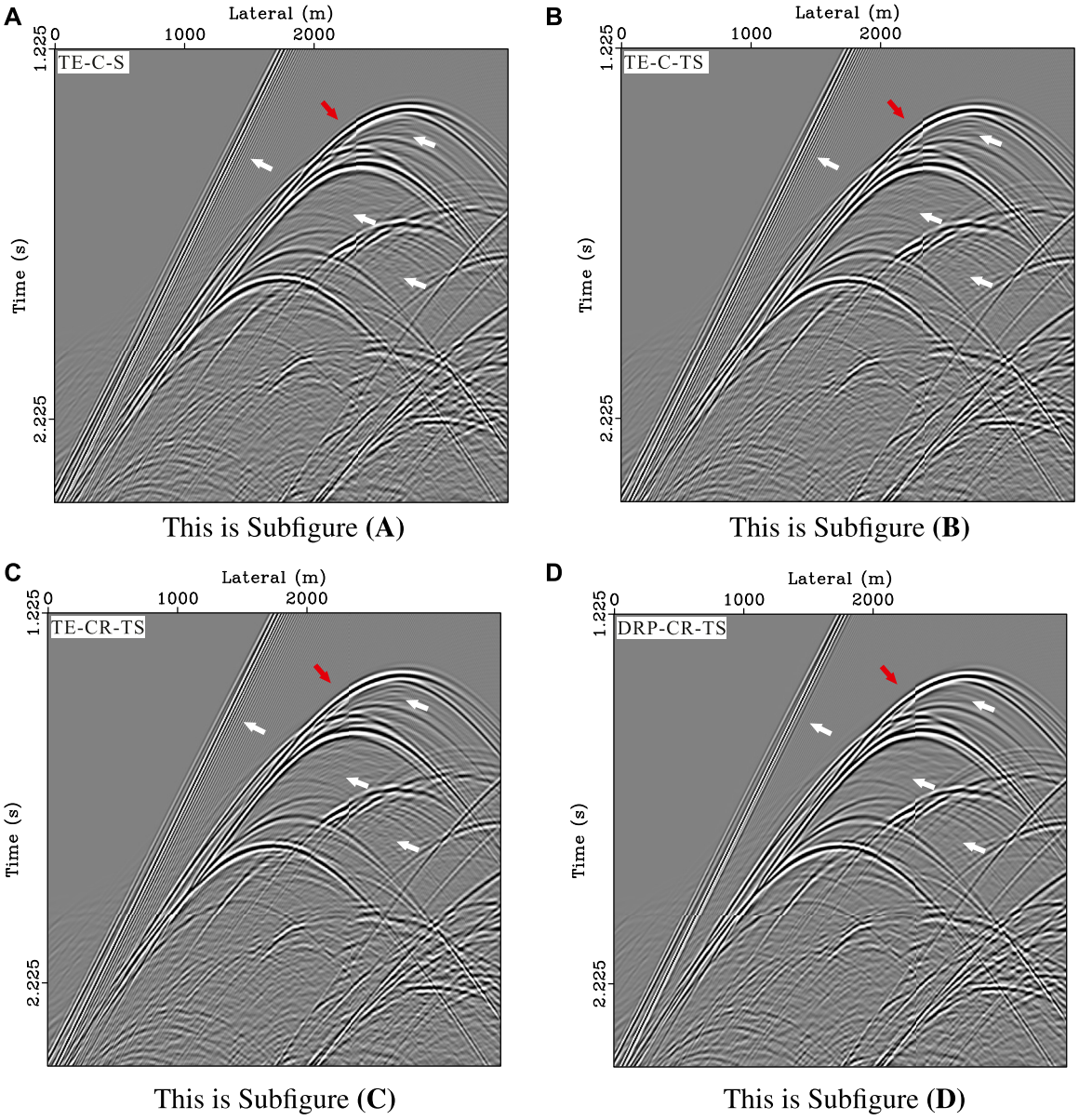


FIGURE 11
Seismic records of the four methods. **(A)** TE-C-S method. **(B)** TE-C-TS method. **(C)** TE-CR-TS method. **(D)** DRP-CR-TS method.

with the Courant number r, the stability curves of our new scheme and other three methods are shown in Figure 3. It can be seen that the stability factors of the proposed method are slightly less than that of the TE-CR-TS method. Although the DRP-CR-TS method adopts an analytical cross stencil, its stability does not sharp decrease, and it is much larger than that of the conventional methods (TE-C-S and TE-C-TS). Figure 4 shows the maximum value of stability factors satisfying $r \leq s$ for different orders M. The stability curve of DRP-CR-TS method is volatile due to the use of numerical method to solve the over-determined system, and it is sensitive to order M, but this does not affect the overall stability. It can be seen that the stability curve of the proposed method has the same level with the TE-CR-TS method. Stability analyses in Figures 3, 4 reveals that the proposed scheme has the same level stability as the standard temporal high-order scheme (TE-CR-TS method), and far better than the conventional scheme.

4 Numerical experiments

4.1 Homogeneous model

In this section, we use a 2-D homogeneous velocity model to examine our new scheme. The 2D homogeneous model has 512×512 grid nodes with the grid spacing $h = 6m$ and velocity $v = 1500\ m/s$. A Ricker-wavelet source with a main frequency of

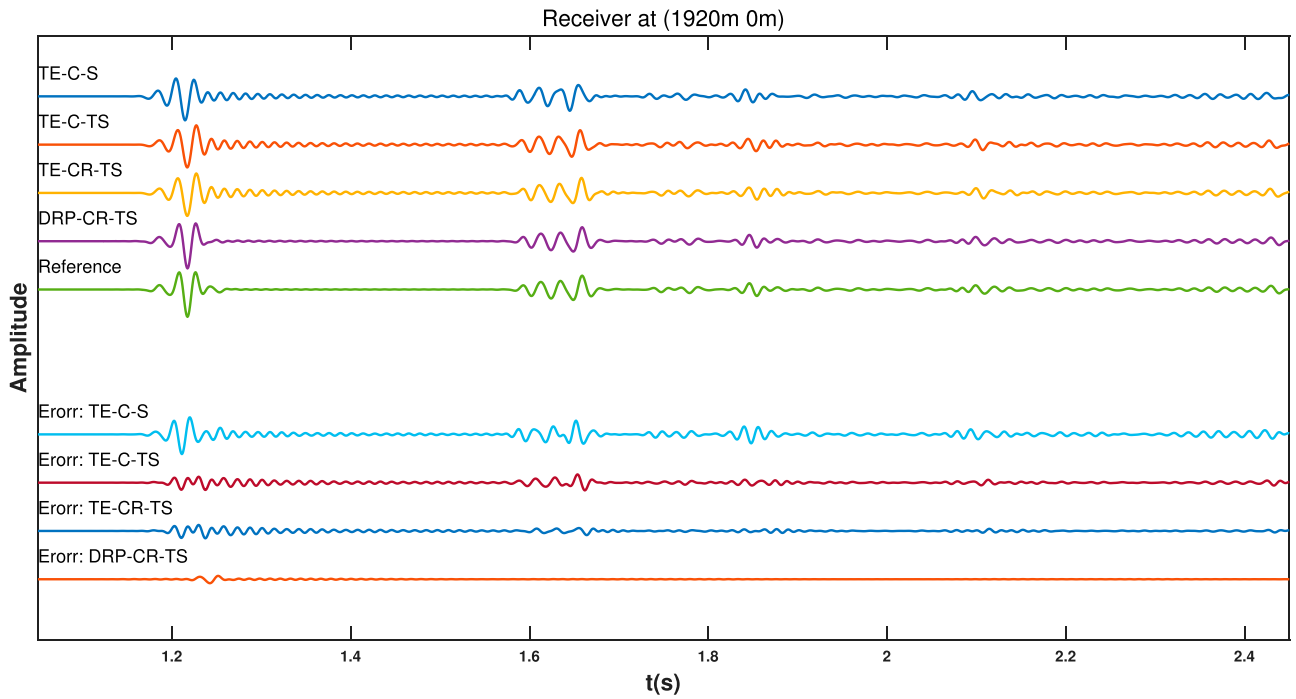


FIGURE 12
Seismic records of the four methods at (1920 m, 0 m).

$40Hz$ is located at the spatial point $(1536m, 1536m)$. Two receivers at the spatial points $(768m, 768m)$ and $(768m, 1536m)$ are used to record the waveforms.

Figure 5 shows the snapshots with a time step $\tau = 0.001s$. The TE-C-S and TE-C-TS methods have serious temporal and spatial dispersion errors (Figure 5A). The temporal dispersion error (red arrow) of the TE-CR-TS method is smaller than that of the TE-C-TS method. However, the TE-CR-TS method still has obvious spatial dispersion error (white arrow), because the TE-based FD coefficients preserve the dispersion relation in a limited range. Figure 5D shows that the corresponding spatial dispersion error is reduced considerably in the proposed method (DRP-CR-TS). Figure 6 shows the seismic records of the four methods at Receiver 1 and Receiver 2. The reference traces represented by the green curves are obtained by the high-order FD scheme under the fine grid. The Receiver 1 (Figure 6A) shows that the temporal dispersion errors of the TE-CR-TS and DRP-CR-TS methods are smaller than that of the TE-C-S and TE-C-TS methods, and the Receiver 2 shows that the spatial dispersion error are serious in the TE-based methods (TE-C-S, TE-C-TS, and TE-CR-TS). But the proposed method (DRP-CR-TS) still has a small level of the dispersion error. Table 2 lists the relative errors of the four methods compared to the reference trace at Receiver 1 and Receiver 2. The relative errors of the cross-rhombus stencil are smaller than that of the conventional cross stencil, especially the DRP-CR-TS method reduces the relative error significantly.

We also analyze the snapshots for different orders M, the snapshots are shown in Figure 7. In this case, simply increasing M can not effectively reduce the dispersion error in the TE-C-S method. The temporal dispersion errors of the TE-C-TS and TE-CR-TS methods are gradually reduced, but when $M = 12$, there are still obvious spatial dispersion error (white arrow). The corresponding spatial dispersion error is mitigated in the proposed method (DRP-CR-TS) even at low order ($M = 6$). Then, we study the snapshots of the proposed method for different N and L, the results are shown in Figure 8. The dispersion error of the proposed method is small for different L and N, thus we can select an appropriate low-order L or N to improve the computational efficiency.

4.2 Inhomogeneous model

4.2.1 2D BP model

We use a widely referred 2D BP velocity model (Figure 9) to test the four methods in the inhomogeneous model. The 2D BP model has 560×1000 grid nodes with the variation of velocities from 1500 m/s to 4500 m/s. In this case, we set time step $\tau = 0.0007$ s, $M = 8$, $N = 3$, $L = M$, grid spacing $h = 7.5m$ and main frequency $f_m = 30Hz$ for numerical simulation. A total of 1,000 receivers are located on the surface of the model.

Figure 10 shows the snapshots of the four methods. The TE-C-S method has obvious temporal dispersion error (red arrow), and the corresponding error in the TE-C-TS and TE-CR-TS methods is reduced (Figures 10B, C). However, the spatial dispersion error (white arrow) is still serious. Figure 10D shows the spatial dispersion error of the proposed method (DRP-CR-TS) is significantly reduced in the low- and high-velocity layers. Figure 11 shows the seismic records of the four methods, and Figure 12 shows the corresponding seismic records at $(1920m, 0m)$. It can be seen that the TE-based methods have serious spatial dispersion error

TABLE 3 CPU execution times and the relative errors of the four methods on the BP model.

Cases	Methods	M	N	L	Execution times s)	Relative errors (pa)
1	TE-C-S	8	\	\	649.0250	0.8941
2	TE-C-TS	8	\	\	649.3852	0.4469
3	TE-CR-TS	8	3	\	983.2374	0.3001
4	DRP-CR-TS	8	3	8	791.3253	0.1096
5	DRP-CR-TS	8	3	6	729.4710	0.1098
6	DRP-CR-TS	8	3	3	670.7974	0.1102

from the first arrivals, but the corresponding error is mitigated in the proposed method (DRP-CR-TS). The reflected wave from the high-velocity layers contain both the spatial and temporal dispersion errors, and the corresponding error in the proposed method (DRP-CR-TS) is smaller than that of the other three methods.

5 Discussion

In this section, we discuss the computational cost and accuracy simultaneously. Taking the above BP model as an example, we design a series of FD parameters for seismic modeling. The numerical experiments are executed on the same computer (Intel Core I7-700 with 3.6 GHz and 8 GB memory). **Table 3** shows the CPU execution times and the relative errors at spatial point (1920 m, 0 m) of the four methods. It is clear that the TE-C-S and TE-C-TS methods have fewer execution times in the numerical experiments (Cases 1 and 2 in **Table 3**), but their relative errors are larger than that of the TE-CR-TS and DRP-CR-TS methods. The relative error of the DRP-CR-TS method is significantly reduced, and their execution time is less than that of the TE-CR-TS method. It is worth noting that when we reduce the length of the analytic FD operator in the DRP-CR-TS method (Cases 5 and 6), the execution time is reduced considerably, and the relative error is almost unaffected. Besides, the DRP-CR-TS method can select a relatively larger time step to reduce the computational cost due to the temporal high-order approximation.

6 Conclusion

We propose a new staggered-grid DRP-based FD scheme with a cross-rhombus stencil for solving the scalar wave equation. The new scheme has a simplified dispersion relation, which is convenient for solving the dispersion-relation-preserving FD coefficients. Besides, the simplified scheme uses the cross stencil instead of the cross-rhombus stencil in some FD operators, thus reducing the computational cost considerably. Dispersion analyses reveals that the proposed FD scheme can effectively mitigate the dispersion error, and it still has a temporal higher-order approximation accuracy. The proposed scheme also has better stability compared with the conventional scheme. Numerical experiments show that the proposed scheme has smaller temporal and spatial dispersion errors while ensuring the computational efficiency, and it is an economical way for the large-scale seismic modeling.

Data availability statement

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

Author contributions

Conceptualization is contributed by CZ and GC; investigation is contributed by CZ, GC, LF, and XZ; data contributed by CZ, GC; formal analysis is contributed by CZ and GC; writing—original draft preparation is contributed by CZ, GC; writing—review and editing is contributed by CZ, LF, and XZ.

Funding

This research was funded by the Prospective and Basic Research Project of CNPC (2021DJ0506), Major Technical Project of CNPC (2022KT0302) and China National Science and Technology Major Project (2016ZX05007-002).

Acknowledgments

This study has benefitted from the reproducible codes and documentation provided in the Madagascar open-source software(Fomel et al., 2013). The authors appreciate the owner of the copyrights.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of

their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/feart.2023.1141220/full#supplementary-material

References

Chen, G., Peng, Z., and Li, Y. (2022). A framework for automatically choosing the optimal parameters of finite-difference scheme in the acoustic wave modeling. *Comput. Geosciences* 159, 104948. doi:10.1016/j.cageo.2021.104948

Chen, G., Wang, Y., Wang, Z., and Zhang, S. (2020). Dispersion-relationship-preserving seismic modelling using the cross-rhombus stencil with the finite-difference coefficients solved by an over-determined linear system. *Geophys. Prospect.* 68, 1771–1792. doi:10.1111/1365-2478.12953

Chen, J.-B. (2011). A stability formula for lax-wendroff methods with fourth-order in time and general-order in space for the scalar wave equation. *Geophysics* 76, T37–T42. doi:10.1190/1.3554626

Chen, J.-B. (2007). High-order time discretizations in seismic modeling. *Geophysics* 72, SM115–SM122. doi:10.1190/1.2750424

Dablain, M. (1986). The application of high-order differencing to the scalar wave equation. *Geophysics* 51, 54–66. doi:10.1190/1.1442040

Etemadsaeed, L., Moczo, P., Kristek, J., Ansari, A., and Kristekova, M. (2016). A no-cost improved velocity-stress staggered-grid finite-difference scheme for modelling seismic wave propagation. *Geophys. J. Int.* 207, 481–511. doi:10.1093/gji/ggw287

Etgen, J. T., and O¡¯Brien, M. J. (2007). Computational methods for large-scale 3d acoustic finite-difference modeling: A tutorial. *Geophysics* 72, SM223–SM230. doi:10.1190/1.2753753

Fomel, S., Sava, P., Vlad, I., Liu, Y., and Bashkardin, V. (2013). Madagascar: Open-source software project for multidimensional data analysis and reproducible computational experiments. *J. Open Res. Softw.* 1.

Kindelan, M., Kamel, A., and Sguazzero, P. (1990). On the construction and efficiency of staggered numerical differentiators for the wave equation. *Geophysics* 55, 107–110. doi:10.1190/1.1442763

Li, S., Yue, B., Chen, Y., Peng, Z., and Wu, R.-S. (2022). Multichannel impedance inversion in the frequency domain via anisotropic total variation with overlapping group sparsity regularization. *J. Inverse Ill-posed Problems* 30, 307–321. doi:10.1515/jiip-2018-0074

Liang, W., Wang, Y., and Yang, C. (2015). Determining finite difference weights for the acoustic wave equation by a new dispersion-relationship-preserving method. *Geophys. Prospect.* 63, 11–22. doi:10.1111/1365-2478.12160

Liang, W., Wu, X., Wang, Y., and Yang, C. (2018). A simplified staggered-grid finite-difference scheme and its linear solution for the first-order acoustic wave-equation modeling. *J. Comput. Phys.* 374, 863–872. doi:10.1016/j.jcp.2018.08.011

Liu, W., He, Y., Li, S., Wu, H., Yang, L., and Peng, Z. (2019). A generalized 17-point scheme based on the directional derivative method for highly accurate finite-difference simulation of the frequency-domain 2d scalar wave equation. *J. SEISMIC Explor.* 28, 41–71.

Liu, Y. (2013). Globally optimal finite-difference schemes based on least squares. *Geophysics* 78, T113–T132. doi:10.1190/geo2012-0480.1

Liu, Y., and Sen, M. K. (2009). A new time–space domain high-order finite-difference method for the acoustic wave equation. *J. Comput. Phys.* 228, 8779–8806. doi:10.1016/j.jcp.2009.08.027

Liu, Y., and Sen, M. K. (2013). Time-space domain dispersion-relation-based finite-difference method with arbitrary even-order accuracy for the 2d acoustic wave equation. *J. Comput. Phys.* 232, 327–345. doi:10.1016/j.jcp.2012.08.025

Moczo, P., Kristek, J., Galis, M., Chaljub, E., and Etienne, V. (2011). 3-d finite-difference, finite-element, discontinuous-galerkin and spectral-element schemes analysed for their accuracy with respect to p-wave to s-wave speed ratio. *Geophys. J. Int.* 187, 1645–1667. doi:10.1111/j.1365-246x.2011.05221.x

Moczo, P., Kristek, J., and Galis, M. (2014). *The finite-difference modelling of earthquake motions: Waves and ruptures*. Cambridge University Press.

Moczo, P., Kristek, J., and Halada, L. (2000). 3d fourth-order staggered-grid finite-difference schemes: Stability and grid dispersion. *Bull. Seismol. Soc. Am.* 90, 587–603. doi:10.1785/0119990119

Ren, Z., Li, Z., Liu, Y., and Sen, M. K. (2017). Modeling of the acoustic wave equation by staggered-grid finite-difference schemes with high-order temporal and spatial accuracy. *Bull. Seismol. Soc. Am.* 107, 2160–2182. doi:10.1785/0120170068

Tan, S., and Huang, L. (2014b). A staggered-grid finite-difference scheme optimized in the time–space domain for modeling scalar-wave propagation in geophysical problems. *J. Comput. Phys.* 276, 613–634. doi:10.1016/j.jcp.2014.07.044

Tan, S., and Huang, L. (2014a). An efficient finite-difference method with high-order accuracy in both time and space domains for modelling scalar-wave propagation. *Geophys. J. Int.* 197, 1250–1267. doi:10.1093/gji/ggu077

Wang, E., Ba, J., and Liu, Y. (2019). Temporal high-order time–space domain finite-difference methods for modeling 3D acoustic wave equations on general cuboid grids. *Pure Appl. Geophys.* 176, 5391–5414. doi:10.1007/s00024-019-02277-2

Wang, E., Liu, Y., and Sen, M. K. (2016a). Effective finite-difference modelling methods with 2-d acoustic wave equation using a combination of cross and rhombus stencils. *Geophys. J. Int.* 206, 1933–1958. doi:10.1093/gji/ggw250

Wang, S., and Teixeira, F. L. (2003). Dispersion-relation-preserving fdtd algorithms for large-scale three-dimensional problems. *IEEE Trans. Antennas Propag.* 51, 1818–1828. doi:10.1109/tap.2003.815435

Wang, Y., Liang, W., Nashed, Z., Li, X., Liang, G., and Yang, C. (2014). Seismic modeling by optimizing regularized staggered-grid finite-difference operators using a time-space-domain dispersion-relationship-preserving method. *Geophysics* 79, T277–T285. doi:10.1190/geo2014-0078.1

Wang, Y., Liang, W., Nashed, Z., and Yang, C. (2016b). Determination of finite difference coefficients for the acoustic wave equation using regularized least-squares inversion. *J. Inverse Ill-posed Problems* 24, 743–760. doi:10.1515/jiip-2015-0005

Wu, H., Li, S., Chen, Y., and Peng, Z. (2020). Seismic impedance inversion using second-order overlapping group sparsity with a-admm. *J. Geophys. Eng.* 17, 97–116. doi:10.1093/jge/gxz094

Ye, F., and Chu, C. (2005). "Dispersion-relation-preserving finite difference operators: Derivation and application," in *SEG technical program expanded abstracts 2005* (Houston, TX: Society of Exploration Geophysicists), 1783–1786.

Zhang, C., Fan, L., Chen, G., and Li, J. (2022). Avo-friendly velocity analysis based on the high-resolution pca-weighted semblance. *Appl. Sci.* 12, 6098. doi:10.3390/app12126098

Zhang, J.-H., and Yao, Z.-X. (2013). Optimized finite-difference operator for broadband seismic wave modeling. *Geophysics* 78, A13–A18. doi:10.1190/geo2012-0277.1

Zhou, H., Liu, Y., and Wang, J. (2022). Time–space domain scalar wave modeling by a novel hybrid staggered-grid finite-difference method with high temporal and spatial accuracies. *J. Comput. Phys.* 455, 111004. doi:10.1016/j.jcp.2022.111004

frontiers | Frontiers in Earth Science

TYPE Original Research
PUBLISHED 24 March 2023
DOI 10.3389/feart.2023.1047626

OPEN ACCESS

EDITED BY
Peng Zhenming,
University of Electronic Science and Technology of China, China

REVIEWED BY
Qiang Guo,
China University of Mining and Technology, China
Sanyi Yuan,
China University of Petroleum, Beijing, China
Shu Li,
Jishou University, China

*CORRESPONDENCE
Jiachun You,
youjiachun@cdut.edu.cn

SPECIALTY SECTION
This article was submitted to Environmental Informatics and Remote Sensing, a section of the journal Frontiers in Earth Science

RECEIVED 18 September 2022
ACCEPTED 13 March 2023
PUBLISHED 24 March 2023

CITATION
Wang Z, You J, Liu W and Wang X (2023), Transformer assisted dual U-net for seismic fault detection.
Front. Earth Sci. 11:1047626.
doi: 10.3389/feart.2023.1047626

Transformer assisted dual U-net for seismic fault detection

Zhiwei Wang[1], Jiachun You[2]*, Wei Liu[2] and Xingjian Wang[3]

[1]School of Petroleum Engineering, China University of Petroleum (East China), Qingdao, Shandong, China, [2]College of Geophysics, Chengdu University of Technology, Chengdu, Sichuan, China, [3]State Key Laboratory of Oil and Gas Reservoir Geology and Exploitation, Chengdu University of Technology, Chengdu, Sichuan, China

Automatic seismic fault identification for seismic data is essential for oil and gas resource exploration. The traditional manual method cannot accommodate the needs of processing massive seismic data. With the development of artificial intelligence technology, deep learning techniques based on pattern recognition have become a popular research area for seismic fault identification. Despite the progress made with U-shaped neural networks (Unet), they still fall short in meeting the stringent requirements of fault prediction in complex structures. We propose a novel approach by combining a standard Unet with a transformer Unet to create a parallel dual Unet model, called Dual Unet with Transformer. To improve the accuracy of fault prediction, we compare six loss functions (including Binary Cross Entropy loss, Dice coefficient loss, Tversky loss, Local Tversky loss, Multi-scale Structural Similarity and Intersection over Union loss) using synthetic data, based on three evolution metrics involving Dice coefficient, Sensitivity and Specificity, find that the binary cross entropy loss function is the most robust one. An example comparing the prediction performance of different Unet models on synthetic data demonstrates the superior performance of our Dual Unet model, verifying the practical application value. To further validate the practical feasibility of our proposed method, we use real seismic data with a complex fault system and find that our proposed model is more accurate in predicting the fault system compared to well-developed Unet models such as the classical Unet and classical coherence cube algorithm, without transfer learning. This confirms the potential for wide-scale application of our proposed model.

KEYWORDS

transformer, UNET, fault prediction, dual Unet, loss function

1 Introduction

Seismic fault detection is a crucial step of oil and gas reservoir exploration because faults often serve as pathways for hydrocarbon migration. Furthermore, faults have geological significance as they indicate changes in stress and provide valuable information for drilling. Fault identification technology is constantly developing with the development of seismic exploration technology. In the past, the discontinuity or edge of seismic images is considered as a sign of a fault. Therefore, many fault detection methods are proposed to enhance those discontinuities using some seismic attributes including the semblance, coherence and curvature (Marfurt et al., 1998; Marfurt et al., 1999; Roberts, 2001). To pursue better performance, more improved approaches are proposed including the ant tracking and attributes fused methods (Pedersen et al., 2002; Di et al., 2019; Yuan et al., 2020; Acuña-Uribe et al., 2021; Yuan et al., 2022), but the results still rely heavily on the experience of interpreters and the quality of the seismic attributes used. Moreover, the presence of noise in

seismic images can negatively impact the accuracy of fault detection. Therefore, it is imperative to develop an automatic fault identification method.

With the rapid development of deep learning, especially the deep convolution neural networks (CNN), more and more attention has been paid to processing and interpreting seismic data, such as velocity inversion, seismic salt interpretation and noise suppression (Shi et al., 2019; Wu and McMechan, 2019; You et al., 2020). The powerful capability of deep CNNs to establish non-linear relationships between inputs and targets has made automatic fault identification based on CNN models a popular area of application. Seismic fault detection is essentially a classification task, with labels of "fault" and "non-fault." Over the years, researchers have developed a variety of neural network architectures to tackle this task. In the early stages, support vector machine (SVM) and multi-layer perceptron (MLP) methods were applied to deal with this task (Di et al., 2017; 2018). In recent years, more end-to-end fault-detection deep CNN models have been developed (Xiong et al., 2018). The fault detection task is regarded as semantic segmentation of images, and the standard Unet architecture including encoder and decoder is introduced (Li et al., 2019; Wu et al., 2019). Because of the superiority of Unet models, its many variants have been successfully applied in seismic fault detection, such as a nested residual Unet, Unet 3plus and wavelet transform based CNN (Yang et al., 2020; Gao et al., 2022; Shen et al., 2022).

The main feature of a CNN model is that it shares receptive fields by using filters with limited size. Because of that, it is difficult for CNN-based methods to learn explicit global and long-term semantic information. In cases where the fault system is complex, the positive (fault) and negative (non-fault) labels in seismic images are highly unbalanced, and the CNN model may suffer from an unsatisfactory result, which seems to be unable to fully meet the strict requirements of seismic fault detection. Inspired by the significant success of the transformer with attention mechanism in the field of Natural Language Processing (NLP), a vision transformer (ViT) module with an attention mechanism was introduced (Dosovitskiy et al., 2021). However, transformers were originally designed to process one-dimensional sequences and focus on building global relationships between inputs and targets, which results in a lack of localization information, which coincidentally is the advantage of a CNN model. Integrating the strengths of both models is becoming a new trend, leading to the development of combined CNN and transformer architectures, such as the Transformer-based Unet (TransUnet) and Shifted Windows Transformer-based Unet (Swin TransUnet) (Cao et al., 2021; Chen et al., 2021). Although these two hybrid models have been successfully applied to medical image segmentation, there are few reports on their use in seismic fault prediction. Using the combined CNN-Transformer model to develop a new end-to-end hybrid structure for seismic fault prediction is both promising and significant.

In our manuscript, we begin by presenting our newly developed hybrid CNN-Transformer architecture. Next, we investigate the loss function used in the image segmentation and compare their performances. Afterwards, we detailed compare several well-established CNN architectures using synthetic data and evaluate their metrics. Lastly, we apply the developed CNN models to perform seismic fault prediction on real data and summarize our work.

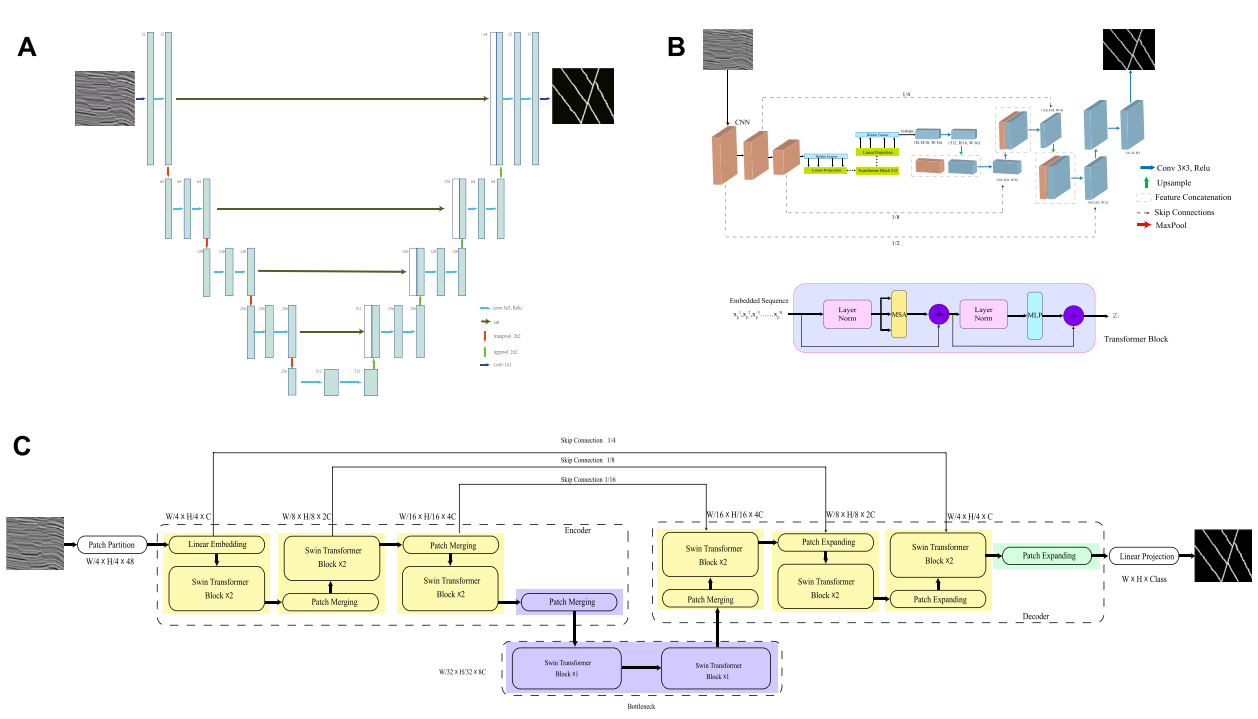


FIGURE 1
Architectures of **(A)** classical Unet, **(B)** TransUnet, **(C)** SwinUnet.

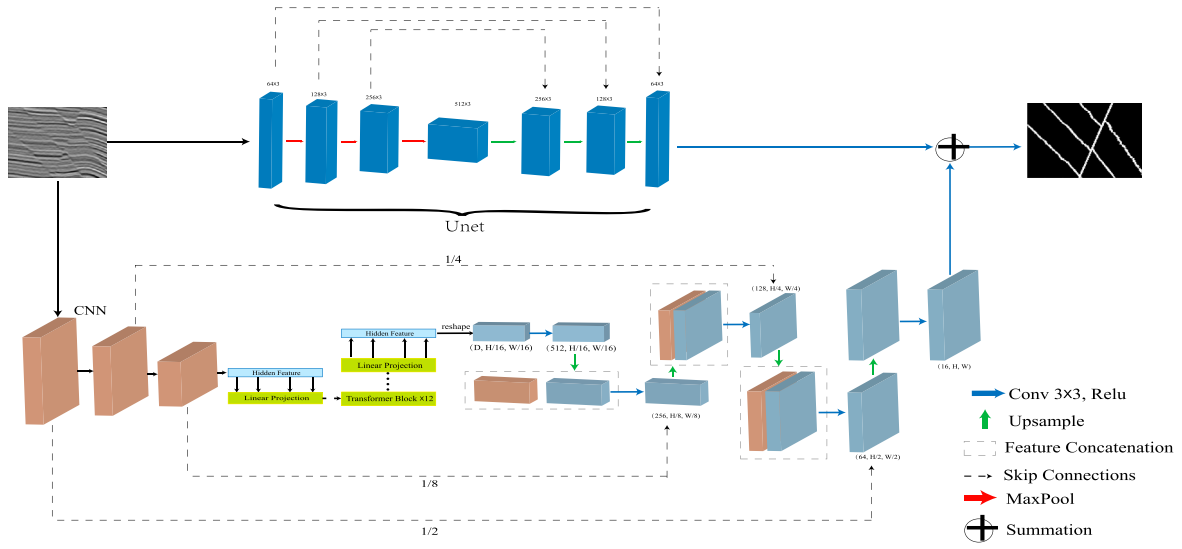


FIGURE 2
Architecture of dual Unet with transformer.

2 Methodology

2.1 Architecture of Unet model

In our manuscript, for the task of semantic segmentation, various variants of the standard Unet model that incorporate transformers are gaining increasing attention. Two of these well-developed models are the TransUnet and Swin TransUnet. The TransUnet model integrates multiple transformer blocks into the bottom layer of the standard Unet model, while the Swin TransUnet replaces the convolutional blocks in the encoder-decoder components with transformer blocks. TransUnet combines the convolution blocks with transformers, showing more fused features; Swin TransUnet illustrates a purely U-shaped Transformer architecture. Further research is needed to determine which architecture produces better results in seismic fault prediction. The architectures of Unet models are shown in Figure 1.

In order to extend the applications of CNN models, a transformer with attention mechanism embedded within a CNN is proposed and serves as a powerful tool in computer vision. In the later examples of synthetical data, we can observe that predicted results of the traditional Unet model are more continuity but lack detailed information whereas transformer assistant Unet models are short of continuity in seismic fault prediction. Due to the use of shared convolution kernels, conventional convolutional neural network models such as Unet are more suitable for learning local features of input images but have limited ability to capture global features. The Transformer models show a good performance of global learning, but its description of local features of images is not ideal. To take advantage of the strengths of both models, we propose a new hybrid architecture called the Dual Unet with Transformer, as illustrated in Figure 2.

2.2 Loss function

In deep learning, the loss function plays a crucial role. By minimizing the loss function, the model converges and reduces the predictive error of the CNN model. Therefore, different loss functions have a significant impact on the model. In the case where the parameters of the deep neural network architecture have been determined, there is a need for a deeper comparative study on how to select the loss function so that the deep neural network converges to an optimal solution. In the seismic fault detection, the positive (faults) and negative (no-faults) labels are extremely unbalanced. The selection of the loss function is crucial for prediction accuracy as it has advantages in handling label imbalances. In our manuscript, we discuss a loss function that is introduced. As seismic fault detection is a binary classification, the loss function we discuss belongs to the binary segmentation problem.

Binary Cross Entropy loss: Binary cross entropy is a classic and widely used loss function in binary classification, but for image segmentation, it is defined to predict a binary label at a pixel level. Its function is defined as

$$loss^{BCE} = -[y log \tilde{y} + (1 - \tilde{y}) log (1 - \tilde{y})] \quad (1)$$

Where y and $\tilde{y}$ are the ground truth and predicted labels, respectively.

Dice coefficient loss: Dice coefficient is a widely used measurement in computer vision, which is applied to calculate the similarity between two images. It has also been suggested for use as a loss function (Milletari et al., 2016). The dice coefficient loss between labels and outputs can be written as

$$loss^{DC} = 1 - \frac{2y\tilde{y}}{y + \tilde{y}} \quad (2)$$

Tversky loss: The loss function of dice coefficient keeps an equal weigh between precision and recall. However, it is difficult to train a network for highly imbalanced data by using the dice coefficient loss function, in which predicting small scaled seismic faults is crucial. To improve performance, the Tversky loss function (Salehi et al., 2017) based on the Tversky index is defined as

$$loss^{T} = 1 - \frac{y\tilde{y}}{y\tilde{y} + \alpha(1-y)\tilde{y} + \beta y(1-\tilde{y})} \tag{3}$$

Where α and β are coefficients. Noted that when $\alpha = \beta = 0.5$, the Tversky loss function is degenerated into the dice coefficient loss function.

Local Tversky loss: Based on the Tversky index, to balance precision and recall ratios in the small regions-of-interest and make the loss function more sensitive to the small regions of interest, a local Tversky loss (Abraham and Khan, 2019) is proposed with a parameter γ, and its loss function is defined as

$$loss^{local\_T} = \left(loss^{T}\right)^{1/\gamma} \tag{4}$$

Where γ is in the range of 1–3.

Multi-scale Structural Similarity (MS-SSIM) loss: The structural similarity index (SSIM) is used to measure image quality evaluation between a processed image and a reference image. However, the SSIM index is a single-scale assessment. To calculate image quality assessment more flexibly, a multi-scale structural similarity (MS-SSIM) index is proposed (Wang et al., 2003), it can be computed by combining the evaluation at different scales using

$$loss^{ms-ssim} = 1 - [l(y,\tilde{y})]^{\alpha_M} \prod_{j=1}^{M} [c(y,\tilde{y})]^{\beta_j} [s(y,\tilde{y})]^{\gamma_j} \tag{5}$$

Where $l(y,\tilde{y}) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}$, $c(y,\tilde{y}) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}$ and $s(y,\tilde{y}) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}$, $C_1 = (K_1 L)^2$, $C_2 = (K_2 L)^2$ and $C_3 = \frac{C_2}{2}$. In generally, L =255 and $K_1 \ll 1$, $K_2 \ll 1$.

Intersection over Union (IoU) loss: The IoU index (Rahman and Wang, 2016) is performed to measure a standard similarity between the predicted and ground truth images for a segment issue, this loss function is generally used in object detection and its definition is written as

$$loss^{IoU} = 1 - \frac{|y \cap \tilde{y}|}{|y \cup \tilde{y}|} \tag{6}$$

3 Numerical experiments

3.1 Performance of loss functions on synthetic data

The selection of loss functions in seismic fault detection is a less concerned topic. In this example, we compare the performance of different loss functions using synthetic data, which lays a solid foundation for the following works. We use synthetic 2D seismic images with faults and their corresponding fault labels as training samples to train a standard Unet (Wu et al., 2019). The synthetic 2D seismic images and their corresponding labels are shown in Figure 3. In the stage of training a neural network, we employ a total of 5,120 samples, in which 80% of them are used as training samples and the remaining 20% are used as validation datasets while an additional 256 samples are applied to test the accuracy of neural networks. In order to quantitatively evaluate the performance of different loss functions, we use three.

Evaluation indexes, including dice coefficient, sensitivity and specificity, in evaluating the prediction results of CNN models. Dice coefficient is used to account for overlapping pixels between the predicted and ground-truth images while sensitivity and specificity mirror the ratios of true positive and true negative, respectively. These metrics are calculated by using the following equations

$$\begin{aligned} DC &= \frac{2TP}{2TP + FP + FN}, \\ sensitivity &= \frac{TP}{TP + FN} \\ specificity &= \frac{TN}{TN + FP} \end{aligned} \tag{7}$$

Where TP, FP FN and TN represent the number of true positive, false positive, false negative and true negative, respectively.

During the training of the neural network, we set the number of epochs to 30 and use the same optimization parameters, including Adam algorithm and learning rate of 0.0001. The loss and accuracy curves using different loss functions are drawn in Figure 4. Additionally, we also compile statistics for these three metrics using different loss functions, which are listed in Table 1. Examining the loss and accuracy curves, it can be seen that the binary cross entropy loss function achieves the lowest error and the highest accuracy. In the prediction results, we obtain a best dice coefficient of 0.9101 by using the binary cross entropy loss, and IoU loss also gets a very close value, a dice coefficient of 0.9050. As for sensitivity, binary cross entropy loss also surpasses other loss functions and IoU loss follows it closely. Except for MS-SSIM loss function, the specificity of most loss functions is almost equal. The comparison of metrics mutually confirms the accuracy curves using different loss functions (Figure 4B).

In this test, we can conclude that it seems that a single loss function is very difficult to get the best scores in all indexes. In the seismic fault detection task, based on the seismic fault labels, it can be observed that the positive and negative labels seem to be imbalanced, but the binary cross entropy function outperforms other loss functions in most metrics and it is probably the most robust one. According to manuscript of Jadon (2020), other loss functions may work better in the case of highly imbalanced data sets. Therefore, our further work is based on the binary-cross entropy loss function.

3.2 Fault prediction on synthetic data using different architectures

After determining the performances of different loss functions, we carry out an example to compare the performance of different

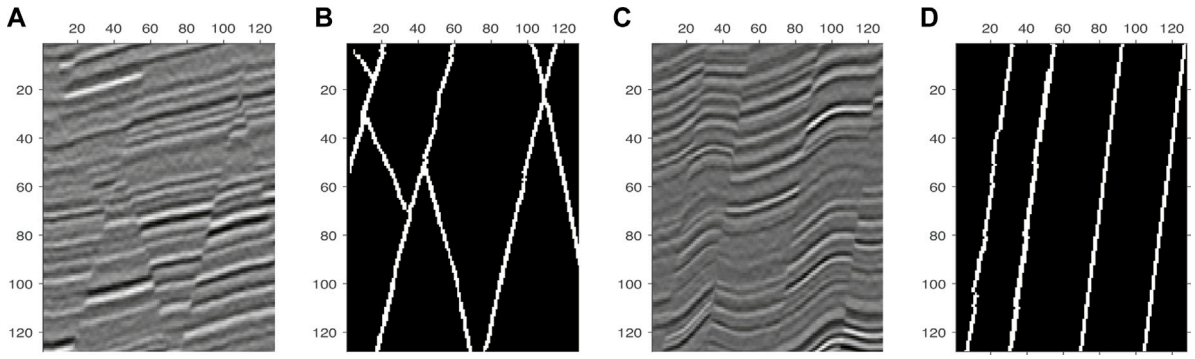


FIGURE 3
Synthetic 2D seismic images and their corresponding labels, **(A,C)** are the seismic patch, **(B,D)** are their corresponding fault labels.

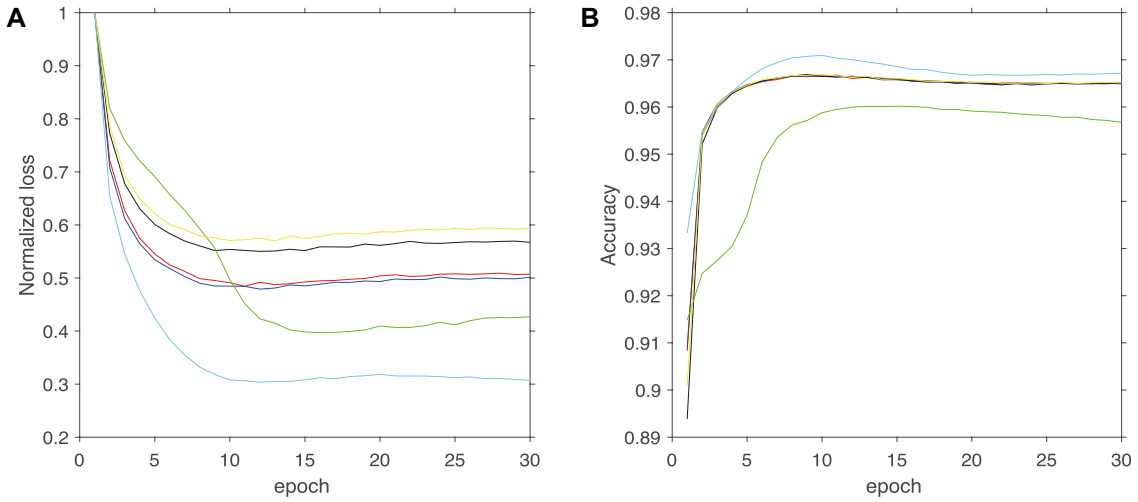


FIGURE 4
Loss **(A)** and accuracy **(B)** curves using different loss functions, the red, blue, back, green, yellow and cyan lines present dice loss, Tversky, local Tverskry, MS-SSIM, IoU and binary cross entropy loss functions.

TABLE 1 Evaluation metrics by using different loss functions.

Loss function	Metrics		
	DC	Sensitivity	Specificity
Binary Cross-Entropy	0.9101	0.8592	0.9616
Dice	0.8688	0.7820	0.9696
Tversky	0.8887	0.8161	0.9687
Local Tversky	0.8730	0.7901	0.9673
MS SSIM	0.8953	0.8375	0.9532
IoU	0.9050	0.8473	0.9652

Unet architectures, including the standard Unet, transUnet, swinTrans Unet and dual Unet with transformer. In the neural network training stage, we use the same training datasets in the examples of comparing loss function. Figure 5 shows the accuracy of the validated data sets using different Unet models. By observing the accuracy curves, it can be seen that the predicted accuracy of the proposed dual Unet with transformer model is superior to the other Unet models. For a fair comparison, we pick up some predicted fault images from the test data set by using different Unet models, which are shown in Figure 6. By comparing the results, we notice that the predicted faults from our proposed models exhibit more accurate information than that of other Unet models. In the experiment, the faults predicted by the traditional Unet model have greater discreteness and less continuity, and the transUnet seems to produce more artifacts. Our proposed dual Unet model combines characteristics and properties of the traditional Unet model and the transUnet model. This example illustrates the superiority of our proposed method and provides a foundation for its application in practical data.

It is interesting to note that the purely swin transformer U-type model seems to produce an imperfect prediction. The predicted

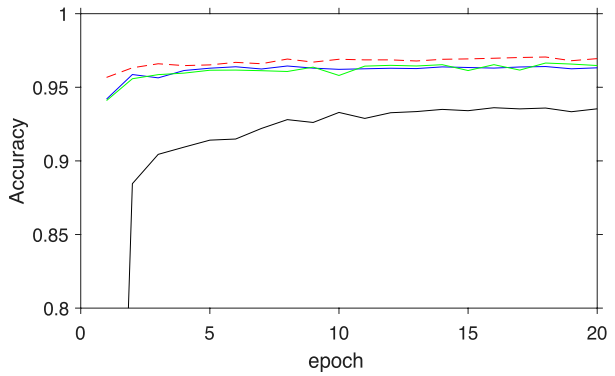


FIGURE 5
Accuracy curves of the validation date sets recorded by using different Unet models, the black, blue and green lines indicate accuracy of swinUnet, standard Unet and transUnet while the red dashed line is the accuracy of our proposed dual Unet with transformer.

results of Swin TramsUnet model have an obvious gap from those of other models. The emphasis on global feature extraction makes it difficult to consider the local continuity of seismic events in the linear mapping of swin transformer blocks, and the precision curve of Swin TransUnet in processing validation data sets also proves this view. At present, we doubt that whether swinTrans Unet is able to achieve a better performance than other methods as described in the medical image segmentation, for the seismic fault detection task (Cao et al., 2021). Fortunately, TransUnet seems to hold a good accuracy compared with the standard Unet. Because of that, we prefer to merge TransUnet and the standard Unet, to build a merged Unet architecture, the predicted accuracy and fault images verify our judgement.

4 Application of real data

In the actual seismic data fault prediction, our work selects a shallow sea area in the southwest of Bohai Bay where the faults are relatively well developed. In terms of regional structure, the study area is located in the east of the low uplift in the Cheng Bei, at the junction of the Bohai Depression and Jiyang Depression. To the south is the Zhendong Depression, to the north is the Bohai Depression, and to the east and west are the Chengbei low uplift and the Bonan low uplift, respectively. The study area is rich in hidden mountains, which have experienced the evolution stages of ancient platform development, Triassic platform disintegration, Yanshan rapid deformation, ancient Quaternary faulting, and recent Quaternary depression. The

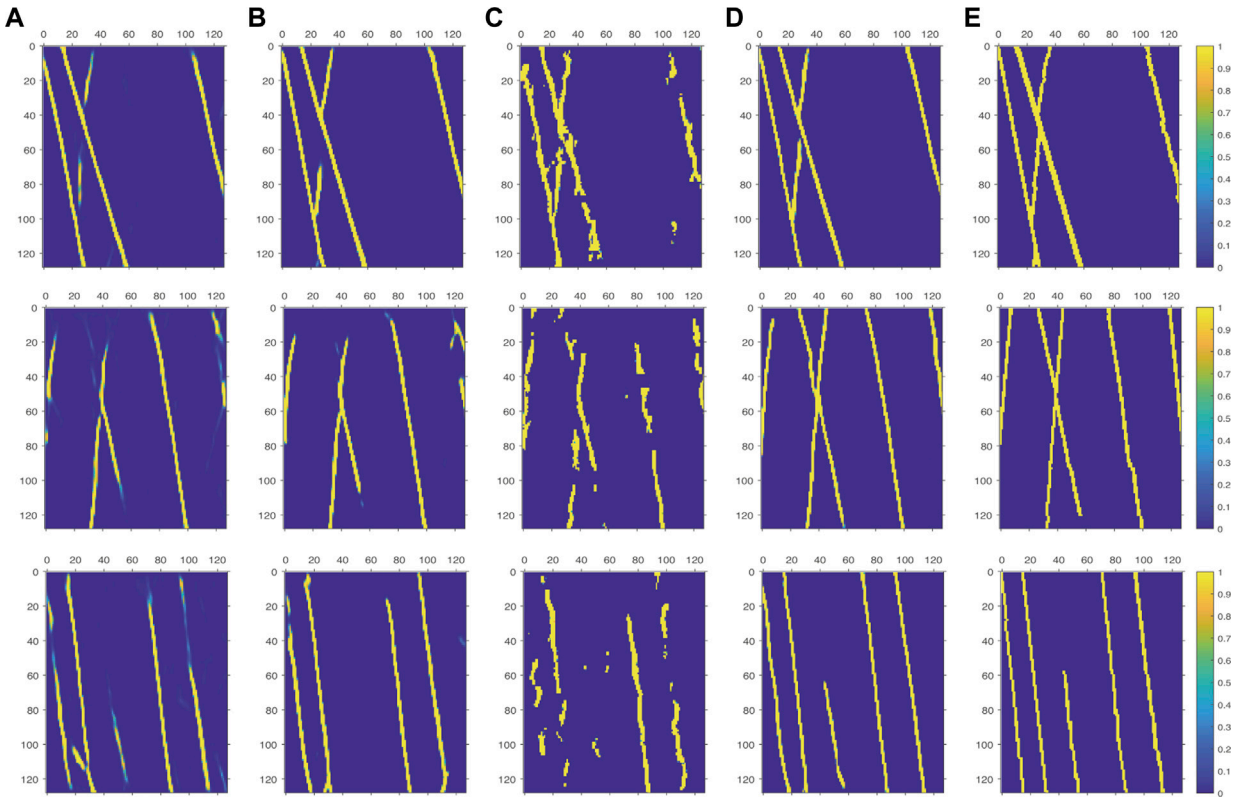


FIGURE 6
Comparison of predicted faults by using different Unet models in the test data set: **(A)** standard Unet; **(B)** TransUnet; **(C)** Swin TransUnet; **(D)** our proposed dual Unet with Transformer; **(E)** ground truth label.

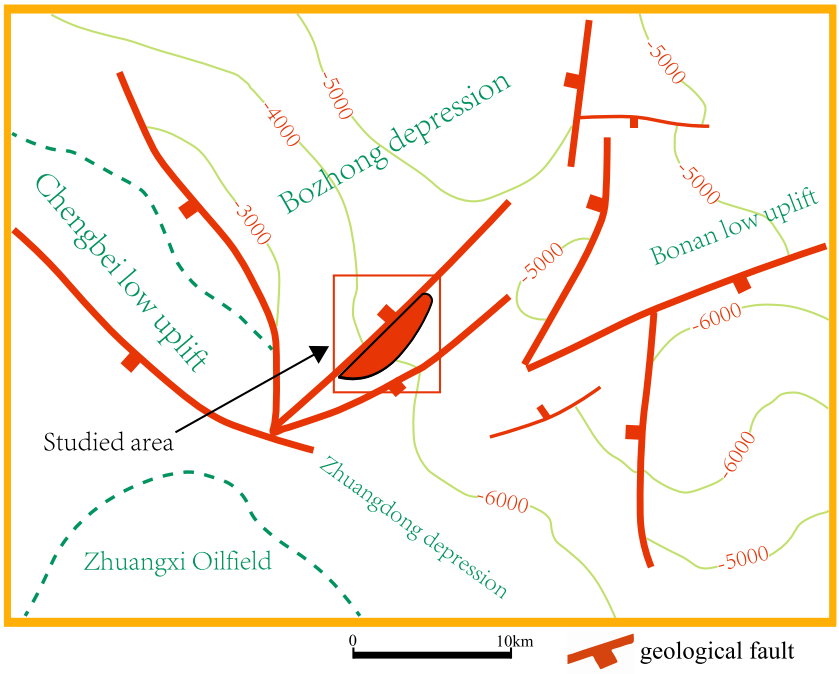


FIGURE 7
Geological background of the research area.

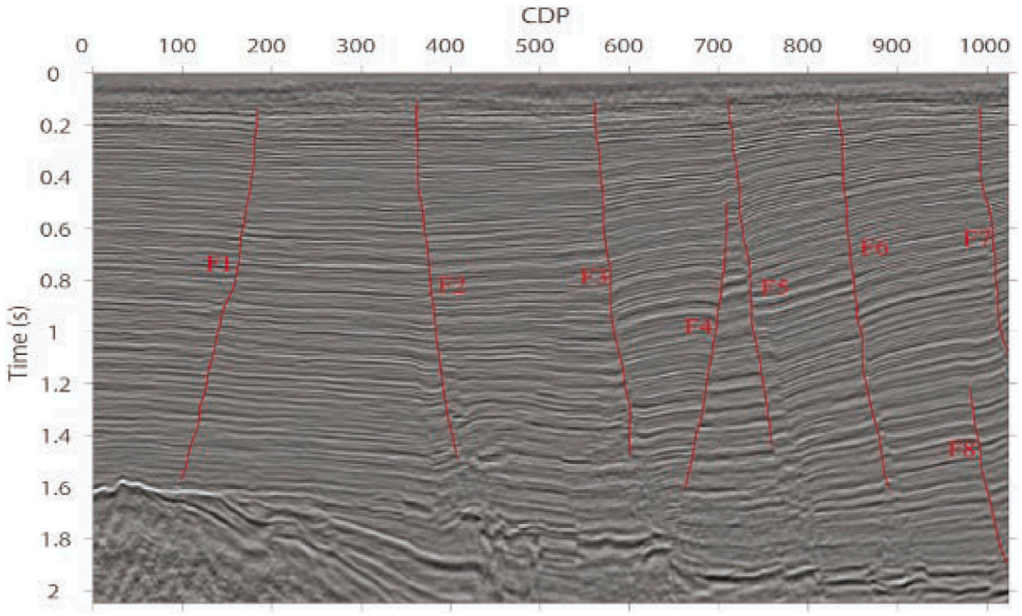


FIGURE 8
Original seismic section with manual interpreted faults.

internal structure of the hidden mountain belt is quite complex, with a large number of folds and fault structures, as shown in Figure 7. Therefore, carrying out the characterization and description of faults in this study area is of great significance for understanding the evolution of the hidden mountains and predicting oil and gas resources.

After the neural network training of synthetic data is completed, we try to use our pretrained Unet models to perform seismic fault

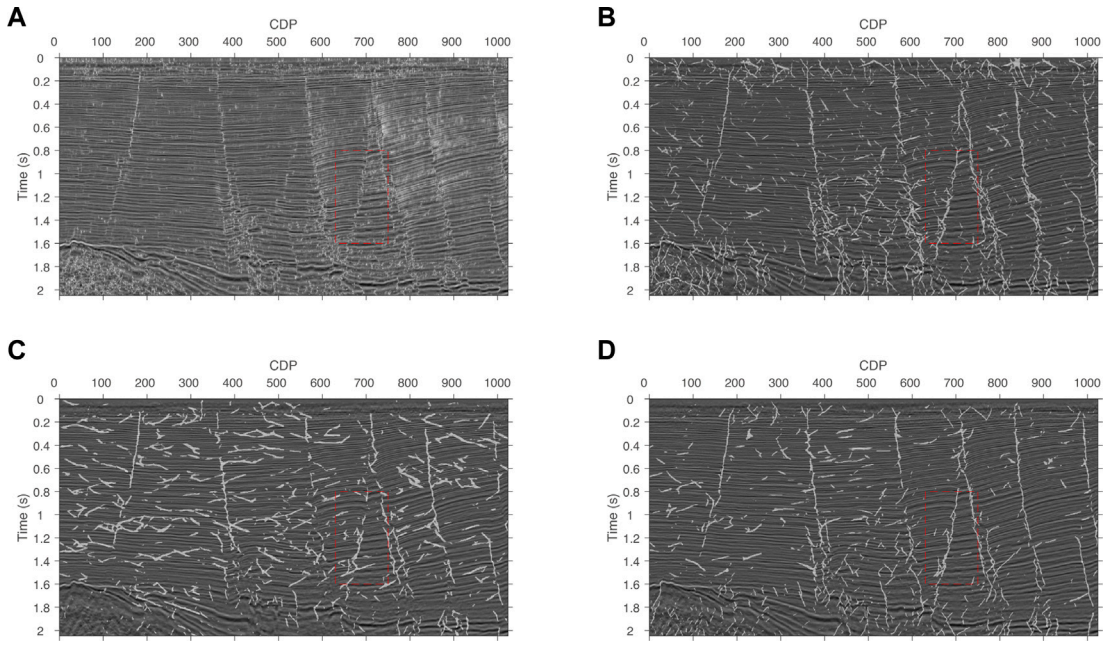


FIGURE 9
Predicted faults by using different neural networks: **(A)** classical coherence cube algorithm; **(B)** conventional Unet; **(C)** TransUnet; **(D)** our proposed dual Unet with transformer.

prediction on the real data. The real seismic section is painted in Figure 8. In Figure 8, some faults are easy to notice directly, which have been marked by red lines. For a seismic fault detection task, we prefer applying the pretrained Unet models to predict the seismic faults straightforwardly without transfer learning, which is a tough challenge. The predicted probability of faults overlapping with the seismic section is shown in Figure 8. In the predicted faults, for some large-scale faults such as fault F1, three Unet models generate similar results. For the case of fault F3, the TransUnet model can only predict it intermittently or hardly. Maybe inherited the ability of standard Unet model, our proposed dual Unet with transformer can produce clearer fault lines than the standard Unet and TransUnet models, especially fault F4 at 1.2–1.4 s. To furtherly compare the performance of fault prediction, we enlarge on the red dashed box (F4) in Figure 9 and display it in Figure 10, it is obvious to see that our proposed model yields a better quality of fault prediction than other two methods. Note that because the Swin TransUnet model has not obtained ideal results in the synthetic example, hence we do not include it in the practical application. In order to compare the application effects of neural network methods and traditional fault identification methods in practical examples, this article used the classical coherence cube algorithm to process the actual example (Bahorich and Farmer, 1995). As shown in Figures 8, 9, the neural network method provides a clearer and more continuous characterization of the fault compared to conventional methods. This also demonstrates the necessity and superiority of conducting deep research on neural network methods.

It is worth noting that the seismic fault prediction of actual seismic data using our proposed model is not performed using transfer learning. The predicted results supply hard evidence to prove that our proposed model has a better generalization than the standard Unet and TransUnet models, and it is of great significance for seismic fault prediction of practical data.

5 Discussion

Our target is to emphasize and raise the significance of loss function in deep learning. Loss functions are crucial in determining the performance of a model. However, for complex objectives like segmentation, it's not feasible to choose a single, universal loss function. The optimal loss function depends mostly on the dataset properties used for training, such as distribution, skewness, and boundaries. It's worth noting that none of the existing loss functions are universally superior in all use cases. Specifically, the binary-cross entropy function performs well in our cases, and we do not think this is a conclusion that applies to all deep learning problems. It may perform well for fault detection task, but for different deep learning tasks, other loss functions may be more effective. In our opinion, specific deep learning tasks need to be analyzed in detail. For example, for deep learning tasks that involve noise suppression in seismic data, which loss function enables better performance of the deep neural network model, and relevant numerical experiments and comparative studies need to be conducted. For medical image semantic segmentation tasks, Jadon (2020) has made a detailed comparison of the performance of different loss functions, which has a different conclusion with fault detection task.

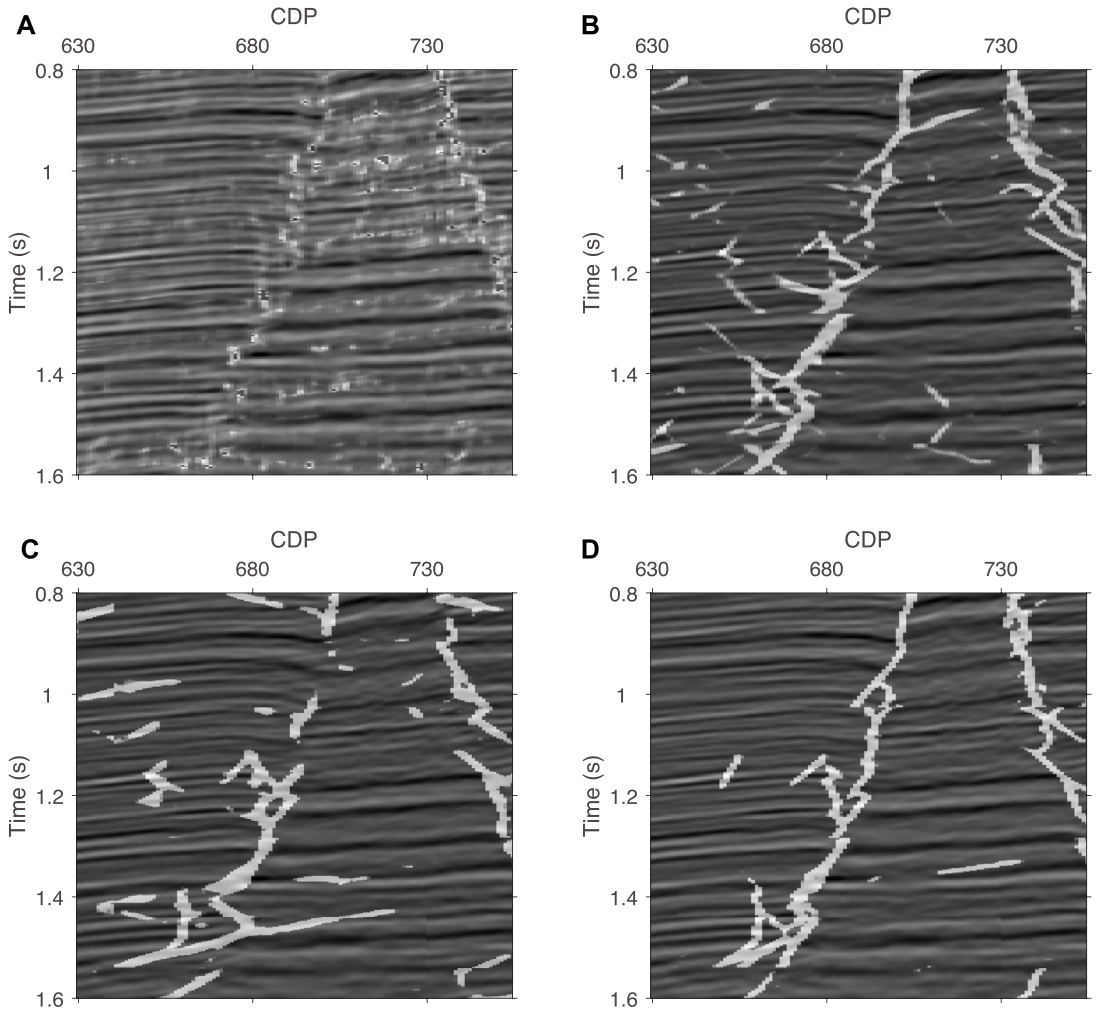


FIGURE 10
Enlarged images of the red dashed box in Figure 8: **(A)** classical coherence cube algorithm; **(B)** conventional Unet; **(C)** TransUnet; **(D)** our proposed dual Unet with transformer.

For the interpretation of 3D seismic data, due to the large amount of data, manual interpretation is difficult to efficiently and quickly complete the relevant interpretation tasks. Fully automatic or semi-automatic computer interpretation solutions have received increasing attention and research. In theory, 3D seismic data can be regarded as an unfolding form of multiple 2D data. For seismic data interpretation, we believe that the successful application of 2D seismic data is the basis for the application of 3D seismic data. Therefore, for fault recognition work, the feasibility and effectiveness of the proposed method in this article were first verified on 2D seismic data. Of course, the development of 3D Transformer-based fracture recognition technology is also one of our future research directions. Currently, in the research of 3D medical image semantic segmentation, some scholars have developed 3D Transformer models, which can provide references for our future research on 3D Transformer-based fault detection (Hatamizadeh et al., 2022; Liang et al., 2022). However, we need to develop a 3D Transformer model suitable for fault recognition in seismic data according to the characteristics of seismic data.

6 Conclusion

Aiming at the problem of seismic fault identification, after analyzing the shortcomings of the convolution block and the transformer block, we attempt to integrate a standard Unet model with a TransUnet model, and develop a dual Unet with transformer. In order to discuss which kinds of loss function can make CNN models converges quickly and produce a best performance, we carried out a numerical example to compare the performance of six loss functions and find that the binary cross entropy loss function has a superior performance in the task of seismic fault detection. In addition, a synthetic data is employed to compare performaces of different Architectures, the predicted fault sections show that our proposed transformer assisted dual Unet

depicts the fault system clearer than that of the standard Unet, TransUnet, Swin TransUnet and classical coherence cube algorithm. Based on that, through seismic fault prediction and qualitative comparison, predicted results demonstrate that our proposed dual Unet with transformer model obtains a more accurate and convergent fault prediction than that of the standard Unet, TransUnet and Swin transUnet models in a synthetical case. In the application of real data, our proposed model generates a higher quality fault predicted image, compared with other Unet models, proving its practical application value.

Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

Author contributions

JY: The conception and design of the study, manuscript editing; ZW: Manuscript writing and revising, processing of data; WL: manuscript reviewing and editing; XW: Provide the real seismic data and some suggestion.

Funding

This research is supported by National Natural Science Foundation of China (Grant Nos 42050104, 42030812 and 42004103).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer SY declared a shared affiliation with the author ZW to the handling editor at time of review.

Publisher's note

References

Abraham, N., and Khan, N. M. (2019). "A novel focal Tversky loss function with improved attention U-net for lesion segmentation," in *2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019)*. 683–687.

Acuña-Uribe, M., Pico-Forero, M. C., Goyes-Peñafiel, P., and Mateus, D. (2021). Enhanced ant tracking: Using a multispectral seismic attribute workflow to improve 3D fault detection. *Lead. Edge* 40 (7), 502–512. doi:10.1190/tle40070502.1

Bahorich, M., and Farmer, S. (1995). 3-D seismic discontinuity for faults and stratigraphic features: The coherence cube. *Lead. edge* 14 (10), 1053–1058. doi:10.1190/1.1437077

Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., et al. (2021). *Swin-Unet: Unet-like pure transformer for medical image segmentation*. ArXiv, abs/2105.05537.

Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., et al. (2021). *TransUNet: Transformers make strong encoders for medical image segmentation*. ArXiv, abs/2102.04306.

Di, H., Shafiq, M. A., Wang, Z., and Alregib, G. (2019). Improving seismic fault detection by super-attribute-based classification. *Interpretation* 7 (3), SE251–SE267.

Di, H., Shafiq, M., and Alregib, G. (2018). "Patch-level MLP classification for improved fault detection," in *SEG technical program expanded abstracts 2018*.

Di, H., Shafiq, M., and Alregib, G. (2017). "Seismic-fault detection based on multiattribute support vector machine analysis," in *Seg technical program expanded abstracts*.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., et al. (2021). *An image is worth 16x16 words: Transformers for image recognition at scale*. ArXiv, abs/2010.11929.

Gao, K., Huang, L., and Zheng, Y. (2022). fault detection on seismic structural images using a nested residual U-net. *IEEE Trans. Geoscience Remote Sens.* 60, 1–15. doi:10.1109/tgrs.2021.3073840

Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., et al. (2022). "Unetr: Transformers for 3d medical image segmentation," in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 574–584.

Jadon, S. (2020). "A survey of loss functions for semantic segmentation," in *2020 IEEE conference on computational intelligence in bioinformatics and computational biology (CIBCB)*, 1–7.

Li, S., Yang, C., Sun, H., and Zhang, H. (2019). Seismic fault detection using an encoder–decoder convolutional neural network with a small training set. *J. Geophys. Eng.* 16 (1), 175–189. doi:10.1093/jge/gxy015

Liang, J., Yang, C., Zhong, J., and Ye, X. (2022). "BTSwin-unet: 3D U-shaped symmetrical swin transformer-based network for brain tumor segmentation with self-supervised pre-training," in *Neural processing letters*, 1–19. doi:10.1007/s11063-022-10919-1

Marfurt, K. J., Kirlin, R. L., Farmer, S. L., and Bahorich, M. S. (1998). 3-D seismic attributes using a semblance-based coherency algorithm. *Geophysics* 63 (4), 1150–1165. doi:10.1190/1.1444415

Marfurt, K. J., Sudhaker, V., Gersztenkorn, A., Crawford, K. D., and Nissen, S. E. (1999). Coherency calculations in the presence of structural dip. *Geophysics* 64 (1), 104–111. doi:10.1190/1.1444508

Milletari, F., Navab, N., and Ahmadi, S-A. (2016). "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 fourth international conference on 3D vision (3DV)*, 565–571.

Pedersen, S. I., Skov, T., Randen, T., and Sønneland, L. (2002). "Automatic Fault extraction using artificial ants," in *Seg technical program expanded abstracts*, 107–116.

Rahman, M. A., and Wang, Y. (2016). *Optimizing intersection-over-union in deep neural networks for image segmentation*, 234–244.

Roberts, A. (2001). Curvature attributes and their application to 3D interpreted horizons. *First Break* 19, 85–100. doi:10.1046/j.0263-5046.2001.00142.x

Salehi, S. S. M., Erdoğmuş, D., and Gholipour, A. (2017). *Tversky loss function for image segmentation using 3D fully convolutional deep networks*. ArXiv, abs/1706.05721.

Shen, S., Li, H., Chen, W., Wang, X., and Huang, B. (2022). Seismic Fault interpretation using 3-D scattering wavelet transform CNN. *IEEE Geoscience Remote Sens. Lett.* 19, 1–5. doi:10.1109/lgrs.2022.3183495

Shi, Y., Wu, X., and Fomel, S. (2019). SaltSeg: Automatic 3D salt segmentation using a deep convolutional neural network. *Interpretation* 7 (3), SE113–SE122. doi:10.1190/int-2018-0235.1

Wang, Z., Simoncelli, E. P., and Bovik, A. C. (2003). Multiscale structural similarity for image quality assessment. The Thrity-Seventh Asilomar Conference on Signals. *Syst. Comput.* 2, 1398–1402.

Wu, G. Y., and Mcmechan, G. A. (2019). Parametric convolutional neural network-domain full-waveform inversion. Geophysics.

Wu, X., Liang, L., Shi, Y., and Fomel, S. (2019). FaultSeg3D: Using synthetic data sets to train an end-to-end convolutional neural network for 3D seismic fault segmentation.

Xiong, W., Ji, X., Ma, Y., Wang, Y., Albinhassan, N. M., Ali, M. N., et al. (2018). Seismic fault detection with convolutional neural network. *Geophysics* 83 (5), O97–O103. doi:10.1190/geo2017-0666.1

Yang, D., Cai, Y., Hu, G., Yao, X., and Zou, W. (2020). "Seismic fault detection based on 3D Unet++ model," in *Seg technical program expanded abstracts*.

You, J., Xue, Y-J., Cao, J., and Li, C. (2020). "Attenuation of seismic swell noise using convolutional neural networks in frequency domain and transfer learning," in *Interpretation*, 1–77.

Yuan, S., Jiao, X., Luo, Y., Sang, W., and Wang, S. (2022). Double-scale supervised inversion with a data-driven forward model for low-frequency impedance recovery. *Geophysics* 87 (2), R165–R181. doi:10.1190/geo2020-0421.1

Yuan, S., Wang, J., Liu, T., Xie, T., and Wang, S. (2020). 6D phase-difference attributes for wide-azimuth seismic data interpretation. *Geophysics* 85 (6), IM37–IM49. doi:10.1190/geo2019-0431.1

frontiers | Frontiers in Earth Science

TYPE Original Research
PUBLISHED 07 July 2023
DOI 10.3389/feart.2023.1153619

OPEN ACCESS

EDITED BY
Peng Zhenming,
University of Electronic Science and Technology of China, China

REVIEWED BY
Jiachun You,
Chengdu University of Technology, China
Wenchao CHEN,
Xi'an Jiaotong University, China
Yatong Zhou,
Hebei University of Technology, China

*CORRESPONDENCE
Yungui Xu,
yungui.xu@outlook.com

RECEIVED 29 January 2023
ACCEPTED 11 April 2023
PUBLISHED 07 July 2023

CITATION
Wang H, Xu Y, Tang S, Wu L, Cao W and Huang X (2023), Well log prediction while drilling using seismic impedance with an improved LSTM artificial neural networks. *Front. Earth Sci.* 11:1153619. doi: 10.3389/feart.2023.1153619

Well log prediction while drilling using seismic impedance with an improved LSTM artificial neural networks

Heng Wang[1,2], Yungui Xu[1,2]*, Shuhang Tang[1,2], Lei Wu[3], Weiping Cao[1,2] and Xuri Huang[1,2]

[1]State Key Laboratory of Oil and Gas Reservoir Geology and Exploitation, Southwest Petroleum University, Chengdu, China, [2]School of Geosciences and Technology, Southwest Petroleum University, Chengdu, China, [3]Shale Gas Project Management Department of CNPC Chuanqing Drilling Engineering Co., Ltd., Chengdu, China

Well log prediction while drilling estimates the rock properties ahead of drilling bits. A reliable well log prediction is able to assist reservoir engineers in updating the geological models and adjusting the drilling strategy if necessary. This is of great significance in reducing the drilling risk and saving costs. Conventional interactive integration of geophysical data and geological understanding is the primary approach to realize well log prediction while drilling. In this paper, we propose a new artificial intelligence approach to make the well log prediction while drilling by integrating seismic impedance with three neural networks: LSTM, Bidirectional LSTM (Bi-LSTM), and Double Chain LSTM (DC-LSTM). The DC-LSTM is a new LSTM network proposed in this study while the other two are existing ones. These three networks are thoroughly adapted, compared, and tested to fit the artificial intelligent prediction process. The prediction approach can integrate not only seismic information of the sedimentary formation around the drilling bit but also the rock property changing trend through the upper and lower formations. The Bi-LSTM and the DC-LSTM networks achieve higher prediction accuracy than the LSTM network. Additionally, the DC-LSTM approach significantly promotes prediction efficiency by reducing the number of training parameters and saving computational time without compromising prediction accuracy. The field data application of the three networks, LSTM, Bi-LSTM, and DC-LSTM, demonstrates that prediction accuracy based on the Bi-LSTM and DC-LSTM is higher than that of the LSTM, and DC-LSTM has the highest efficiency overall.

KEYWORDS

machine learning, long short-term memory, well log prediction, drilling bit, seismic impedance, neural networks

1 Introduction

During the drilling process, reliable prediction of the rock properties of the geological formation below the drill bit is of great significance. If well logs can be effectively obtained within a certain depth range below the drill bit, it can surely help to improve the drilling process with lower risk and optimal strategy. Researchers have proposed various methods for well log prediction while drilling (Wang, 2017; Tamim et al., 2019). Wang (2017) established a 3D model by analyzing the drilling core, rock

chip, and seismic data. The well log curve values below the drilling bit along the borehole trajectory are predicted based on the 3D model. Tamim et al. (2019) constructed a Bayesian classifier using well logs as the classification target. With the posterior probability calculated with prior distribution and conditional probability, the spatial distribution of well logs can be predicted. However, the actual geological formations are non-homogeneous and complex. The logging curve values show great fluctuations with depth even when the sampling interval of the well log is small. The mapping relationships between different data points are strongly non-linear, and the traditional methods are not able to effectively predict the actual formation changes.

Predicting rock property with geophysical data is often a non-linear problem in many cases. Describing the intrinsic relationship between properties and their geophysical responses using explicit mathematical or physical equations can be challenging. However, machine learning has shown great potential in recent years, as it can describe these relationships with network parameters by using a large number of training datasets. Examples of relevant machine learning applications include geological parameter estimation (Ahmed Ali Zerrouki and Baddari, 2014; Iturrarán-Viveros and Parra, 2014), lithology discrimination (Wang et al., 2014; Silva et al., 2015), and stratigraphic boundary determination (Singh, 2011; Silversides et al., 2015). In recent years, there has been a surge of research on well log prediction, resulting in significant improvements in performance (Rolon et al., 2009; Alizadeh et al., 2012; Mo et al., 2015; Long et al., 2016; Salehi et al., 2017).

The well log prediction methods discussed above use fully connected neural networks (FCNN) to construct a point-to-point mapping. However, FCNN shows less effectiveness in characterizing the trend of the data because it cannot capture the relations between the point data of well log at different depths. This means that the correlation between the rock properties at a shallower depth and those at a deeper depth is ignored, potentially contradicting the sedimentary principles in the geological sense. To fix this problem with FCNN in well log prediction, many researchers have improved FCNN by coupling non-machine learning methods, such as wavelet transform (Adamowski and Chan, 2011) and singular spectrum analysis (Sahoo et al., 2017). However, these improvements are often complex and cumbersome to implement. An alternative way to utilize recurrent neural networks (RNNs) is able to deal the problems (Schuster and Paliwal, 1997). In the RNN structure, there is a self-looping structure within each neural unit, which allows previous information to be retained and used later. Since the information can flow freely in the RNN, well log prediction with this method integrates the intrinsic connection between different logs and the overall trend with depth, which follows geological principles.

Long Short-Term Memory Neural Networks (LSTM), an advanced form of RNN, have become widely utilized in the deep learning community for various tasks. LSTM incorporates gate structures within each automatic cycle structure to mimic biological neurons' information conduction

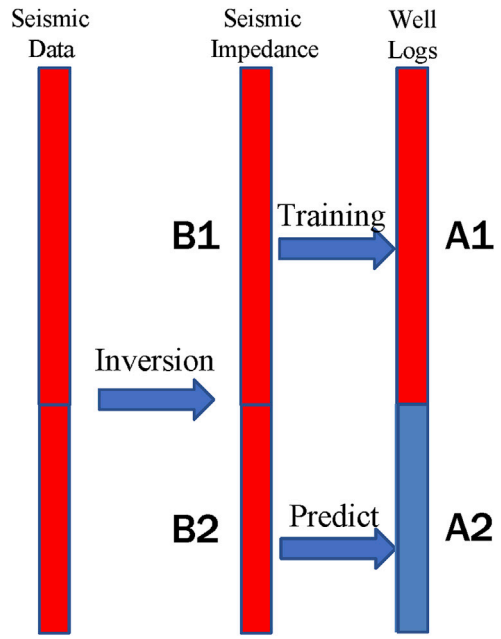


FIGURE 1
Workflow of data procession. (Red means known value and blue means unknown value).

patterns, thereby storing more long-term sequential information without additional tuning. This attribute has facilitated LSTM's extensive use in natural language processing (Deng et al., 2019), machine translation (Lokeshkumar et al., 2020), and speech recognition (Graves and Jaitly, 2014). Moreover, RNNs and LSTMs have also found applications in the field of hydrology to deal with time series data-related problems (da Silva and Saggioro, 2013), and some researchers have used LSTM to generate logs (Jin, 2018).

In the case of well log prediction while drilling, Wang et al. (2020) employed the LSTM model, whereas Shan et al. (2021) utilized the CNN-LSTM hybrid model to achieve the goal. Both methods employed neighboring well logs to predict the logs at the undrilled segment, resulting in successful outcomes. However, when neighboring wells are too distant (as in the case of sparse neighboring wells) and well correlation is weak, the prediction results can become biased. In certain scenarios, prediction may not be possible due to the unavailability of well logs from neighboring wells at the prediction depth.

This study proposes a novel approach for well log prediction while drilling, utilizing seismic impedance and three different artificial intelligence networks: the LSTM network and its two derivative networks, namely, the Bi-LSTM and the newly developed Double Chain LSTM (DC-LSTM). The proposed methodology effectively integrates seismic impedance constraints for well log prediction. The results demonstrate that the two derivative networks outperform the LSTM network in terms of prediction accuracy. Additionally, the DC-LSTM network exhibits superior computational efficiency as compared to the Bi-LSTM network, owing to the reduction in

TABLE 1 The steps of the work.

Steps of well log prediction while drilling
Step 1: Obtaining seismic impedance by the inversion of seismic data.
Step 2: Constructing a dataset with B1 and A1 as the label library.
Step 3: Training the neural network with the label library and establishing the mapping from seismic impedance to well log.
Step 4: Performing the prediction from B2 to A2 based on the trained network.

the number of network parameters and consequently the computing time, while maintaining prediction accuracy. A practical case study is conducted to verify the effectiveness of the proposed methodology.

2 Methodology

2.1 Well log prediction while drilling principle: training and prediction

The acquisition of well logs is traditionally carried out after drilling, while well log prediction while drilling refers to the real-time estimation of well logs at a certain depth below the drill bit during the drilling process. This paper presents a workflow for such prediction, as depicted in Figure 1. The known observed logs are shown in red, whereas the logs yet to be predicted are shown in blue. The seismic impedance obtained through seismic inversion is utilized as input data for the training and prediction of well logs using three different networks, as outlined in Table 1.

2.2 Three networks: LSTM, Bi-LSTM, and DC-LSTM

2.2.1 The LSTM adaption

The LSTM neural network (Hochreiter and Schmidhuber, 1997) is a specialized type of recurrent neural network (RNN) designed to capture long-term dependencies within sequential data. This makes it particularly suitable for processing well logs, which are a type of sequential data with a relatively small sampling interval. A significant characteristic of well logs is the representation of long-term depth trends within a large depth interval, which can play an essential role in prediction tasks. Unfortunately, these trends are often overlooked in existing models. However, LSTM's ability to retain information with long-term dependencies from previous, more distant steps allows it to capture and incorporate these trends into its predictions. As a result, LSTM represents an effective tool for well log prediction.

Figure 2 illustrates the architecture of the LSTM network. The inputs to the network are denoted by x_1 to x_n, while h_1 to h_n represent the hidden state vectors. The computation of the LSTM network's outputs is performed iteratively in a stepwise fashion, from Step 1 to Step n, with the hidden state vectors h_1 to h_n and outputs y_1 to y_n calculated sequentially by the LSTM unit. In the context of well log prediction, only the final output, y_n, is utilized for prediction, and is subsequently fed into a Fully-Connected Neural Network (FCNN).

2.2.2 Bidirectional LSTM network adaption

The Bi-LSTM neural network (Graves and Schmidhuber, 2005) represents a variant of the LSTM architecture that incorporates two parallel LSTMs: one running forward along the input sequence, and another running backward in reverse order. By exploiting both the past and future context of the input data, the Bi-LSTM model can capture more comprehensive and accurate representations of sequential data. Specifically, the forward and backward LSTMs compute hidden state vectors in opposite directions, and the final output of the network is produced by combining the two LSTM' outputs with appropriate weights in the output layer. A diagram of the Bi-LSTM structure is presented in Figure 3.

In Figure 3, the input is denoted by x, and the hidden state vectors of the forward and backward hidden layers are represented by h^f and h^b, respectively. The forward layer computes its outputs in a sequential manner from Step 1 to n, while the backward layer calculates its outputs from Step n to 1. The LSTM unit calculates the hidden state vectors of the forward and backward layers, i.e., h^f and h^b, respectively. The network generates its outputs based on the following equation,

$$y_n = h^f \otimes h^b \tag{1}$$

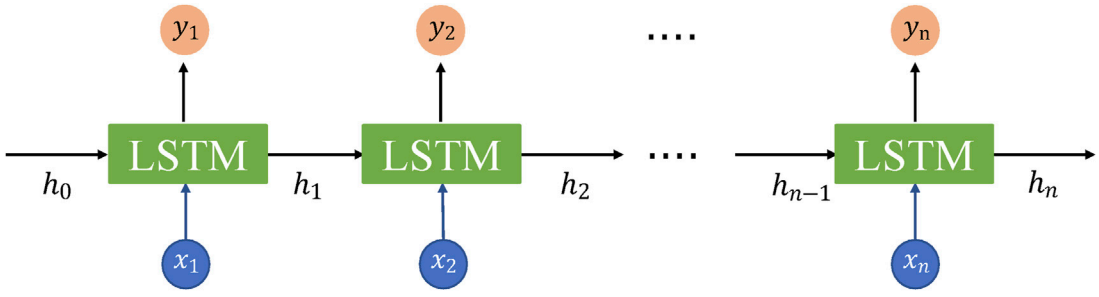


FIGURE 2
The architecture of the LSTM. (Each green box stands for one LSTM unit).

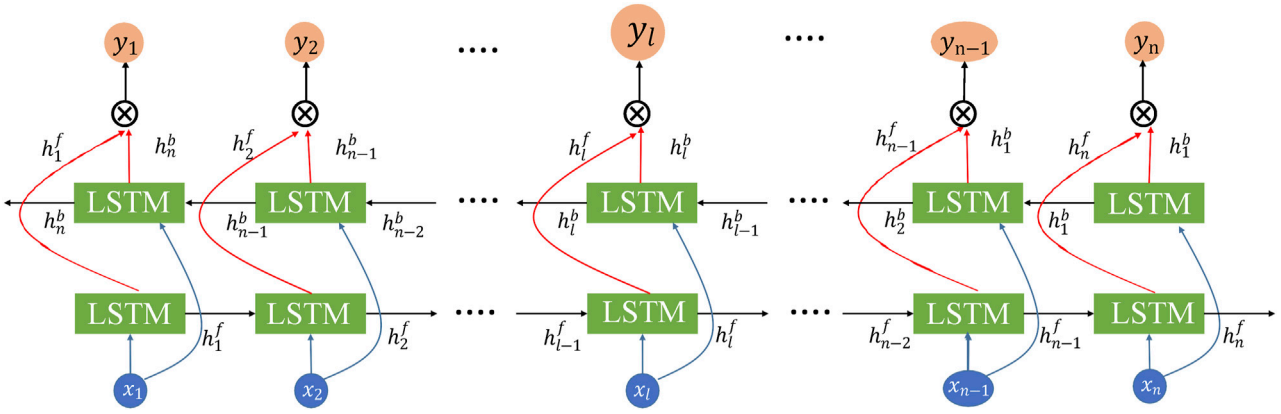


FIGURE 3
The architecture of the BiLSTM.

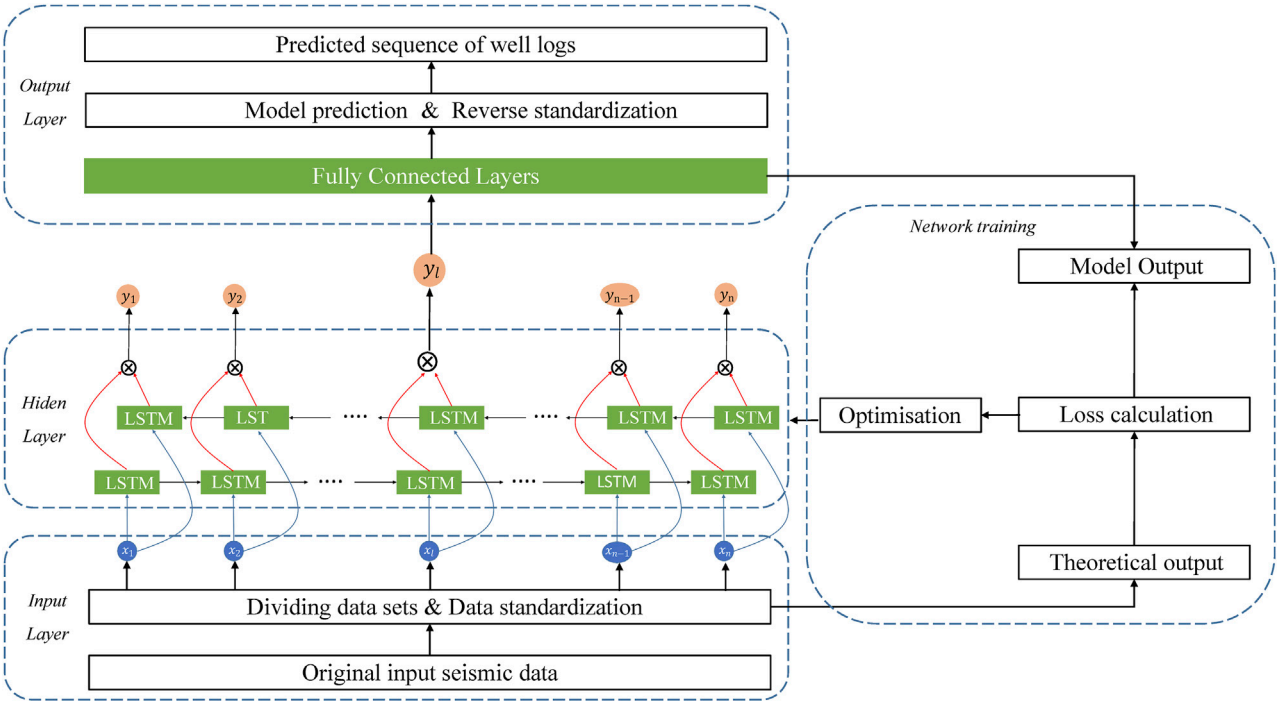


FIGURE 4
Prediction framework of Bi-LSTM network.

where a ⊗ is a matrix operator, which is utilized to couple h^f and h^b sequences. The operator ⊗ could be a summation, multiplication, concatenating, or average function. In the experiment of this paper, the operator ⊗ is concatenating the horizontal matrix. The final output of a Bi-LSTM network is expressed as vectors, $[y_1, y_2, \ldots, y_l \ldots, y_{n-1}, y_n]$. The element in the middle position, y_l, is used to predict the log values.

Figures 4, 5 present the prediction framework using a Bi-LSTM network for mapping the relationship between well logs and seismic impedance. The prediction framework is comprised of three layers, namely, the input layer, the output layer, and the hidden layer. The network is trained using the training set to obtain the mapping relationship between the log segment of the well and seismic

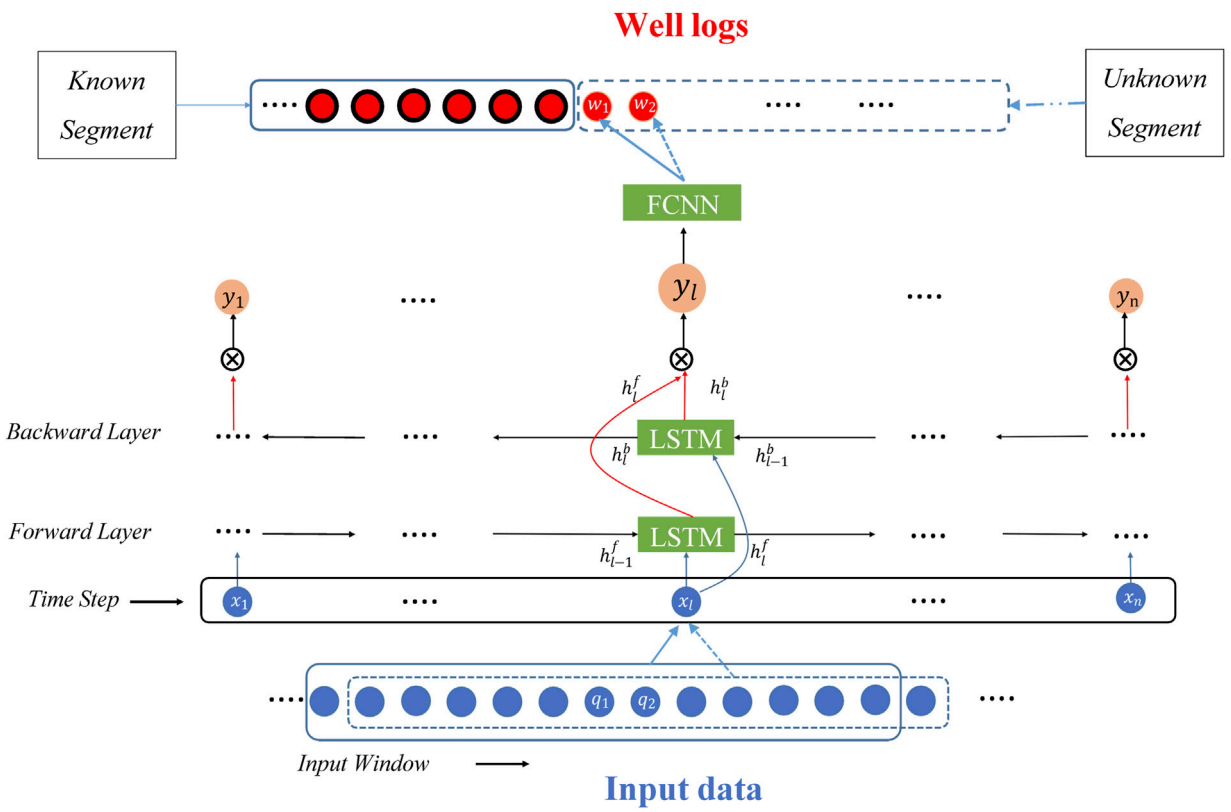


FIGURE 5
Flow chart for prediction of Bi-LSTM network.

TABLE 2 The specific steps of the prediction phase.

Steps of Bi-LSTM model for well log prediction while drilling
Step 1: Determine the location of the input window
Step 2: As shown in Figure 5, when the data is input to the network (the data in the solid box), q1 is located in the midpoint of the input window, i.e., q_l is the input of time step x_l, then the LSTM cell at the position of time step= l can obtain as well as the original seismic data q_1 and the input of two cell states h_{l-1}^f and h_{l-1}^b, outputs h_l^f and h_l^b, After the ⊗ operation y_l is obtained and from this, the logging data w_1 for the unknown segment is predicted, w_1 and q_1 being the corresponding data for the same depth.
Step 3: The Input Window is shifted one place to the right, so that when the data is entered into the network (the data inside the dashed box), q_2 is located right in the middle of the Input Window, and it can be predicted that w_2, w_2, and q_2 are the corresponding data of the same depth.
Step 4: Repeat the step3 to predict all unknown segment data

impedance. The seismic impedance within the undrilled well segment is then used to predict well logs.

In Figure 5, the model is trained using well logs observed from the drilled segment and the corresponding seismic impedance to optimize the network. The trained model is subsequently used to predict log values of the well segment below the training data along the well trajectory. The input window specifies the length of the input seismic impedance sequence to the network, and the input window length should be consistent between the training and prediction phases. In the current example, q_1 represents the value at the position x_l, while h_{l-1}^f and h_{l-1}^b denote the hidden state vectors before and after processing q_1, respectively. Table 2 outlines the prediction steps in detail.

2.2.3 The double-chain LSTM adaption

The Bi-LSTM model utilized for well log prediction while drilling entails numerous prediction parameters. Specifically, the final output of a Bi-LSTM layer is expressed as a set of vectors, denoted as $[y_1, y_2, \ldots, y_l \ldots, y_{n-1}, y_n]$. However, only the vector positioned in the middle, y_l, is employed to predict the log values. As a result, other vectors computed during both training and prediction are not utilized, resulting in a strain on computational efficiency. To address this issue, this paper proposes a DC-LSTM network.

The DC-LSTM network is an optimized network that aims to enhance the computational efficiency in well log prediction while drilling in comparison to the Bi-LSTM model. As shown in Figure 6, DC-LSTM divides the input sequence, represented by $[x_1, x_2, \ldots, x_l \ldots, x_{n-1}, x_n]$, into two sequences: $[x_1, \ldots, x_l]$ and $[x_l, \ldots, x_n]$, where x_l is positioned at the middle of the sequence. The two sequences $[x_1, \ldots, x_l]$ and $[x_l, \ldots, x_n]$ are then separately inputted into two LSTM. Following the ⊗ operation, y_l, which is identical to the y_l mentioned in the Bi-LSTM, is obtained. Because

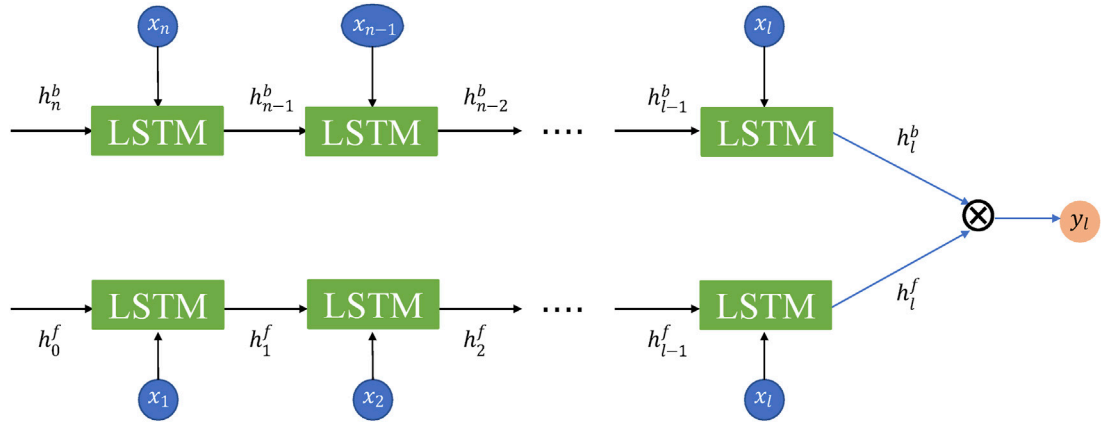


FIGURE 6
The architecture of the Double-chain network.

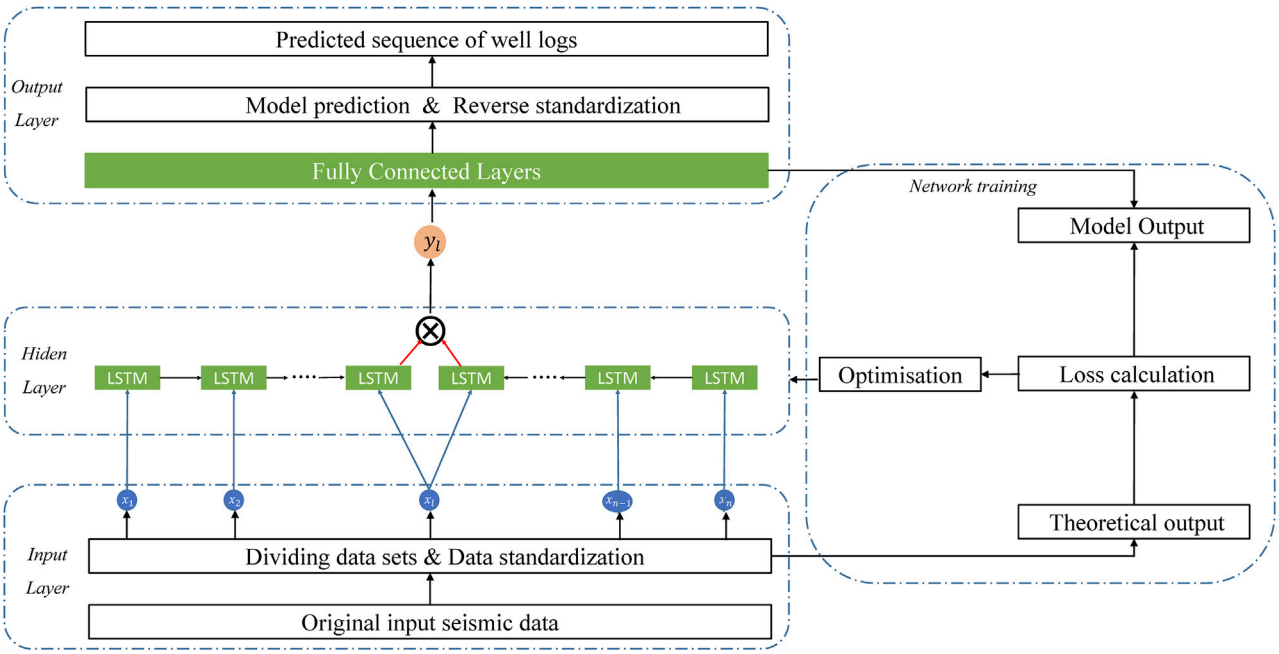


FIGURE 7
Prediction framework of DC-LSTM network.

only one vector, y_l, is involved in the prediction process, DC-LSTM significantly reduces the number of network parameters and enhances calculation efficiency relative to Bi-LSTM.

Figures 7, 8 depict the prediction process with DC-LSTM, which is a simplified and more efficient version of Bi-LSTM in well log prediction while drilling. As illustrated in Figure 7, DC-LSTM replaces Bi-LSTM in the hidden layer, resulting in a simpler overall structure. Furthermore, Figure 8 shows that DC-LSTM requires only n+1 LSTM units, whereas Bi-LSTM utilizes 2n LSTM units. This reduces the number of parameters and improves computational efficiency. Importantly, y_l calculated in DC-LSTM is identical to that in Bi-LSTM, thereby ensuring the same level of prediction accuracy.

The steps of prediction with DC-LSTM are largely analogous to those outlined in Table 2 for Bi-LSTM. The only difference is that DC-LSTM divides the input sequence $[x_1, x_2, \ldots, x_l \ldots, x_{n-1}, x_n]$

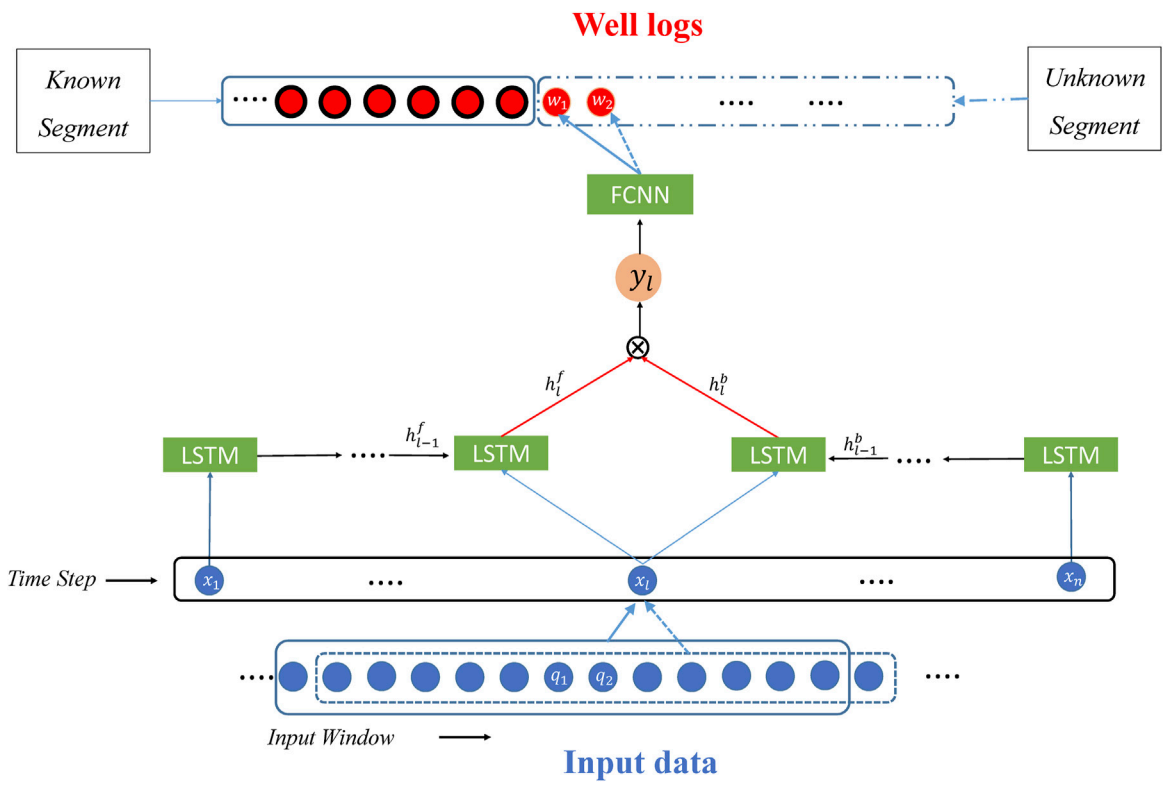


FIGURE 8
Flow chart for prediction of DC-LSTM.

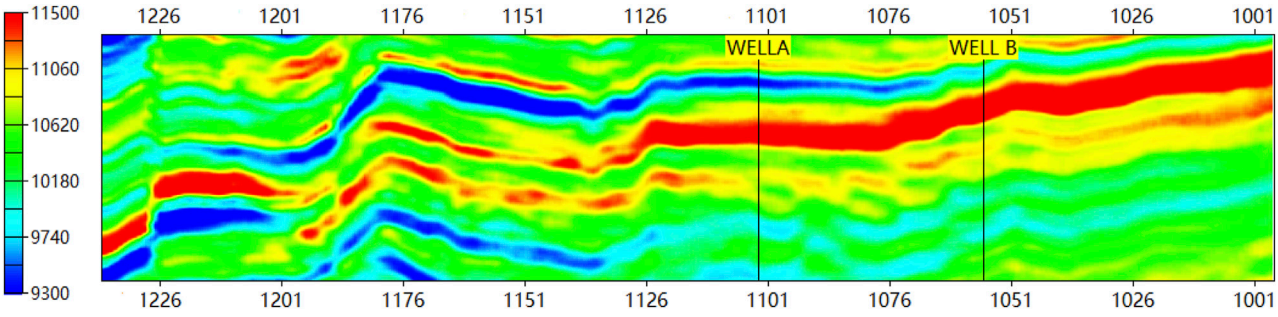


FIGURE 9
Seismic impedance profile of input data.

into two shorter sequences, namely, $[x_1, \ldots, x_l]$ and $[x_l, \ldots, x_n]$, which are then fed into two separate LSTMs.

3 Experiment

In this part, we present the results of well log prediction while drilling using seismic impedance with the LSTM, Bi-LSTM, and DC-LSTM.

3.1 Dataset

In this experiment, a pilot area of an oil field in East China is selected for well log prediction. Two wells and a seismic impedance cube in the area are used in this experiment (see Figure 9). The two wells have four logs for testing, density (DEN), natural gamma ray (GR), compensated neutron log (CNL), and induction (CON).

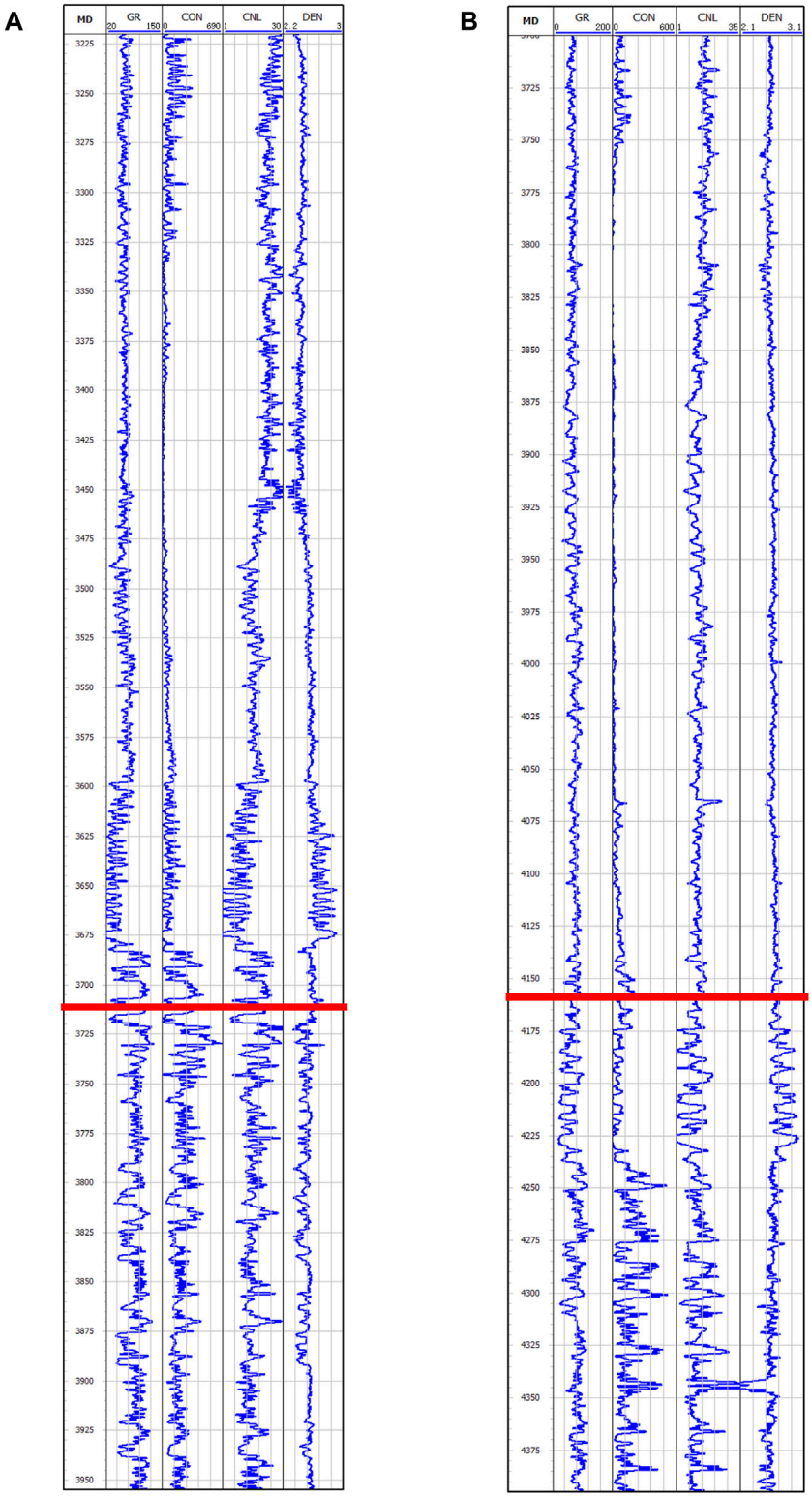


FIGURE 10
Logging curves for well A and well B used in training and validation: **(A)** well A, **(B)** well B

Before the prediction with the three networks, two data preprocessing tasks, data cleaning, and standardization are required.

3.1.1 Data cleaning

Data cleaning (Zhang et al., 2015) aims to improve the data quality of input data. Common problems with logs, such as

invalid and missing values, are cleared. Check on the data value range are necessary as well. Any data point, falling outside the normal range, logically unreasonable, or contradictory, is also cleared.

3.1.2 Data standardization

The purpose of data standardization is to centralize data values so that the characteristics of each type of data are balanced. Data standardization also reduces the disturbance of anomalous values that still exist after data cleaning. This paper uses the z-score normalization method (Hrynaszkiewicz, 2010) to standardize input data standardization, expressed in the equations below,

$$x^* = \frac{x_i - \bar{x}}{s} \tag{2}$$

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{3}$$

$$s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2} \tag{4}$$

where x^* is the standardized data, x_i is the input data, $\bar{x}_i$ is the mean of the input data, and s is the standard deviation of the input data. The standardized data not only retains the correlations present in the original data but also removes the effects of different magnitudes and ranges of data values.

3.2 Evaluation metrics

The problem discussed in this paper is called the 'regression task' in the field of the neural network, and the common loss functions of the regression task are Mean Square Error (MSE), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE). If there are some unusual values (large or small) in the data, these values are more likely to indicate the important geological contrast in rock properties, such as sudden change of lithology, measured acoustic velocity, etc. Therefore, it is necessary to attach more weight on the unusual values. This is achieved by using MSE as the loss function.

MAE and RMSE are used as criteria to evaluate prediction accuracy. MAE is the average of the absolute error between the predicted value and the true value, indicating the true situation of the error, while RMSE is generally used to indicate the deviation between the predicted value and the true value. MAE, MSE, and RMSE are defined as,

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - \hat{y}_i\right| \tag{5}$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{6}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{7}$$

TABLE 3 Training parameter setting.

Parameter	Setting
Optimization algorithm	ADAM
Hidden layers	25
Batch size	32
Learning rate	0.002
Dropout	15%
Epochs	150
Input size	1
Output size	1
Layer number	2

where y_i is the predicted value and $\hat{y}_i$ is the true value. The smaller the MAE and RMSE values, the more accurate the model's prediction of the well logs will be.

In this study, the MAE and RMSE of four logs are calculated and averaged to evaluate the performance of different models. However, the value ranges of these four curves are very different, and the value ranges of the same curve in different wells are also different. For the fair evaluation of the results, the true and predicted values of each curve are put together and normalized before calculating the MAE and RMSE, according to the following equation,

$$y_i = \frac{(y_i - \bar{y})}{(y_{max} - y_{min})} \tag{8}$$

where y_i are the normalized values of a well log, $\bar{y}$ is the mean of the well log, y_{min} refers to the minimum value, and y_{max} denotes the maximum value.

3.3 Experiment results and analyze

In this experiment, two wells, namely, Well A (with measured depth ranging from 3220m to 3960 m) and Well B (ranging from 3700 to 4400 m), along with a seismic impedance segment traversing both wells, were utilized as input data (as depicted in Figure 9). The training data set was constructed by incorporating the first 70% of the well log data (the section above the red line shown in Figure 10), along with the corresponding seismic impedance. The remaining 30% of the logs (the section below the red line in Figure 10) and their corresponding seismic impedance were used for blind test validation. The sampling interval for both wells was set to 0.076 m. The input window length for the networks was fixed at 51 neighboring data points, which corresponds to an actual sampling length of 3.876 m. By using seismic impedance, each input window predicts a single well log point at the depth of the

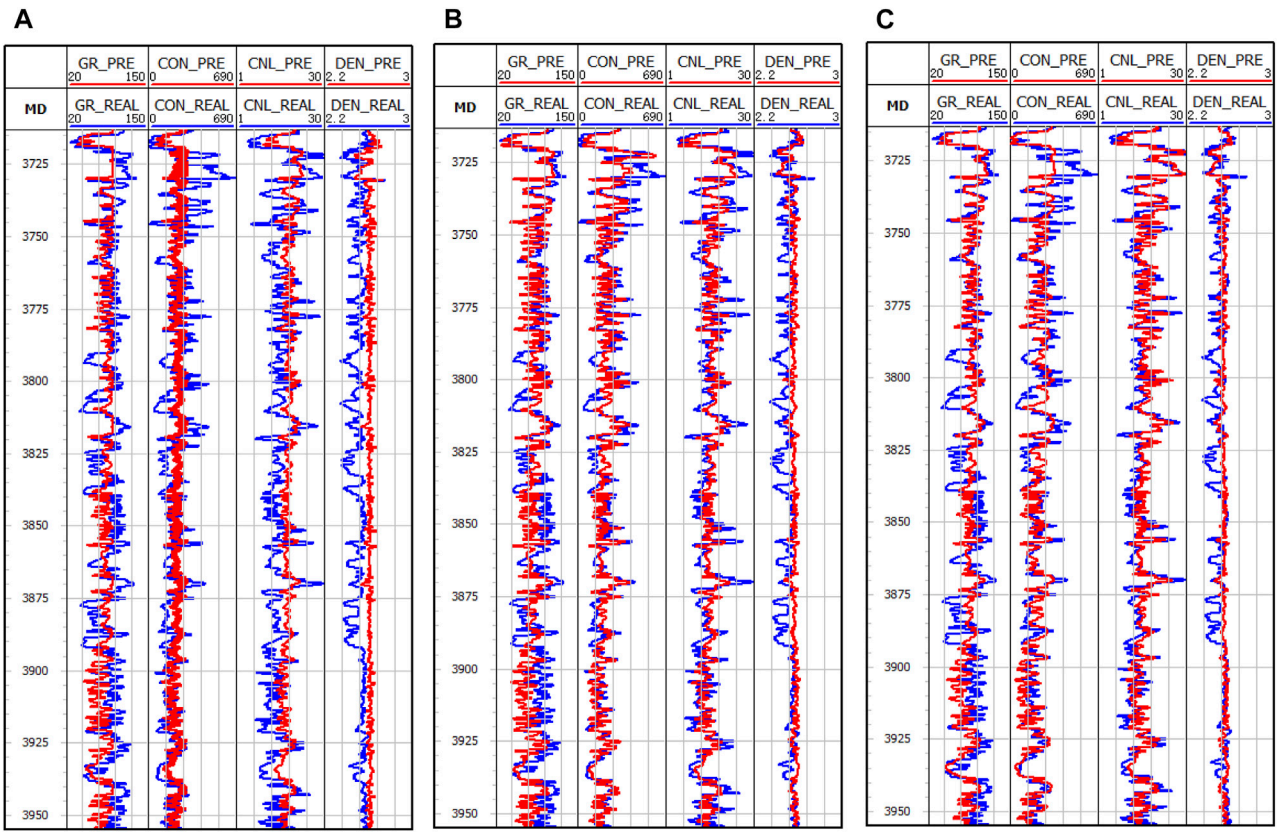


FIGURE 11
Results of the three networks applied to well A: **(A)** LSTM, **(B)** Bi-LSTM, **(C)** DC-LSTM.

TABLE 4 MAE and RMSE of the three methods applied to well A.

	CNL		CON		DEN		GR		Average	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
LSTM	0.154	0.182	0.119	0.153	0.145	0.198	0.176	0.214	0.149	0.187
Bi-LSTM	0.088	0.111	0.067	0.090	0.146	0.183	0.154	0.182	0.114	0.142
DC-LSTM	0.078	0.099	0.076	0.102	0.115	0.156	0.129	0.164	0.099	0.130

TABLE 5 MAE and RMSE of the three methods applied to well B.

	CNL		CON		DEN		GR		Average	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
LSTM	0.062	0.088	0.143	0.207	0.085	0.115	0.167	0.207	0.114	0.154
Bi-LSTM	0.056	0.075	0.119	0.169	0.069	0.110	0.113	0.141	0.089	0.124
DC-LSTM	0.057	0.076	0.110	0.139	0.062	0.089	0.121	0.147	0.086	0.112

middle point of the input window. The window length determines the memory range of the network model during both the training and prediction processes.

In this study, the training data set is obtained by dividing the impedance and well logs into segments with a length of 51. The impedance is utilized as the input, while four well logs, including

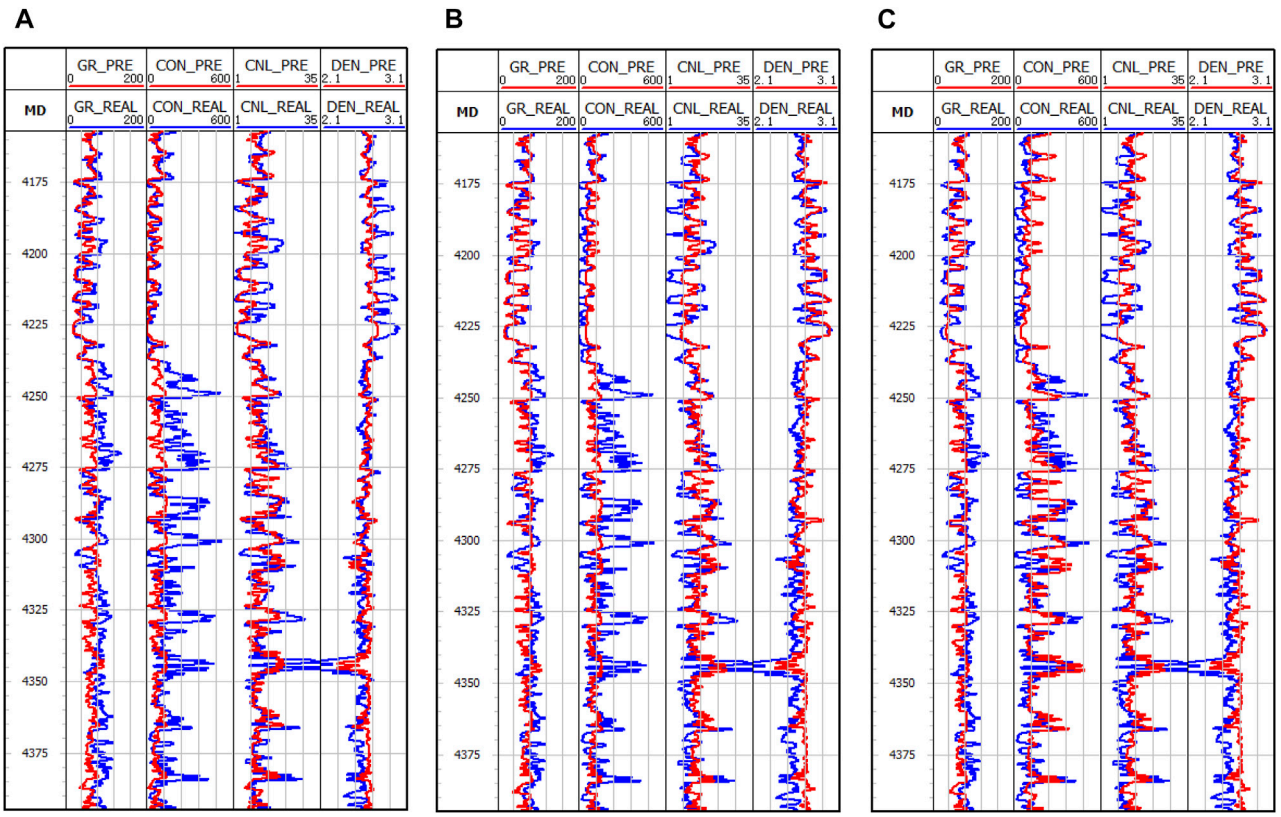


FIGURE 12
Results of the three networks applied to well B: **(A)** LSTM, **(B)** Bi-LSTM, **(C)** DC-LSTM.

TABLE 6 Time spent on three methods (Unit: seconds).

	WELL A		WELL B		Over All		
	Train	Predict	Train	Predict	Train	Predict	Total
LSTM	17.6 ± 1.1	3.4 ± 0.3	18.9 ± 1.2	3.2 ± 0.3	36.5	6.5	43.0
Bi-LSTM	43.5 ± 3.4	8.8 ± 1.1	41.3 ± 3.2	8.1 ± 1.1	84.8	16.9	101.7
DC-LSTM	25.4 ± 2.1	5.1 ± 0.7	24.1 ± 2.1	4.7 ± 0.8	49.5	9.8	59.3

density (DEN), natural gamma ray (GR), compensated neutron log (CNL), and induction (CON) curves, serve as the output. The optimal training parameters, as demonstrated in Table 3, are determined through several rounds of parameter tests.

The training sequences (impedance and well logs) are divided into segments with a length of 51 to produce the training and prediction data set. The impedance is used as the input. Four well logs, including density (DEN), natural gamma ray (GR), compensated neutron log (CNL), and induction (CON) curves, are used as the output. The best results are obtained by using the training parameters shown in Table 3 after a few times of parameter tests.

Figures 11, 12 depict the outcomes of the three methods employed in the analysis of the two wells, where the red and blue curves indicate the predicted and actual values, respectively. As illustrated in these figures, the performance of the LSTM-based method is suboptimal, as it fails to capture the geological changes adequately, leading to inaccurate overall predictions.

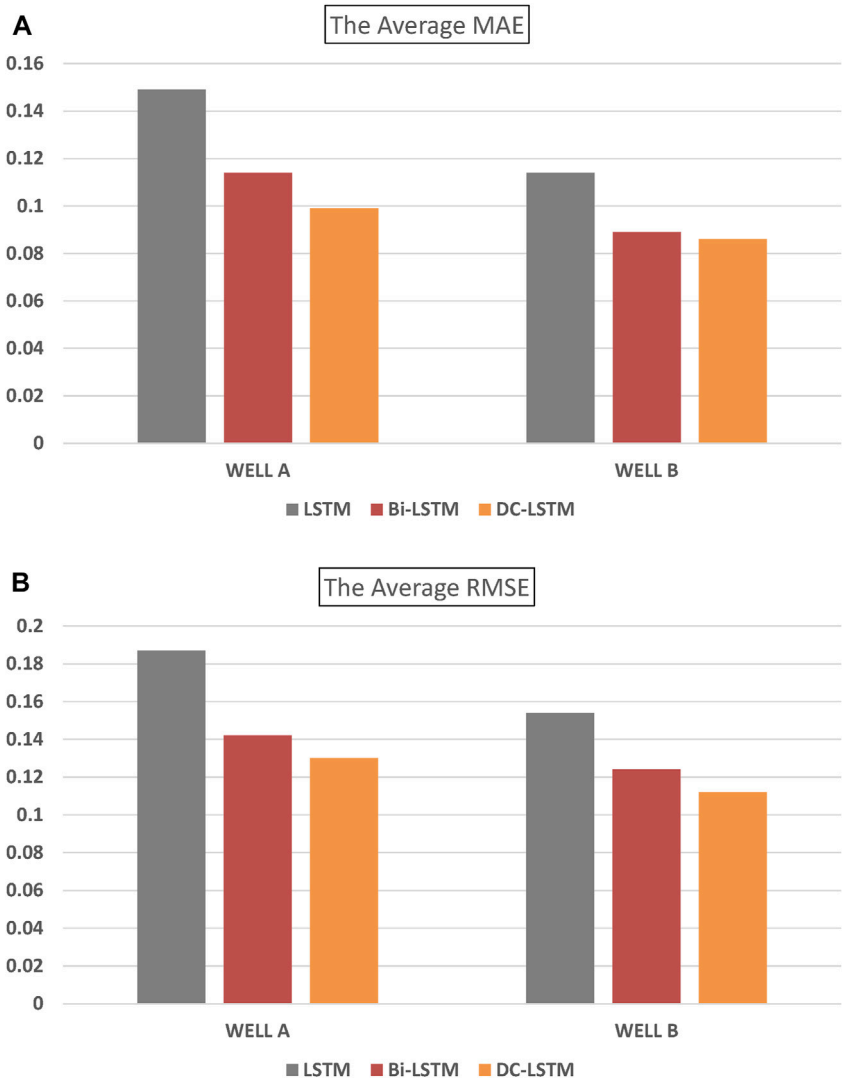


FIGURE 13
The average MAE and RMSE of three methods on three wells: **(A)** MAE, **(B)** RMSE.

Conversely, the Bi-LSTM and DC-LSTM methods yield favorable results in both detecting the geological variations in detail and predicting the overall trend.

Table 4, 5 present the Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) values of three methods applied to two wells. Figure 13 is a bar chart illustrating the performance of these methods. The results indicate that the MAE and RMSE values of the Bi-LSTM and DC-LSTM models are significantly lower than those of the LSTM model in terms of average value. This finding implies that the predicted values of Bi-LSTM and DC-LSTM are closer to the true values than those of the LSTM model. The MAE and RMSE values of DC-LSTM are similar to those of Bi-LSTM. However, Table 6 shows that using DC-LSTM significantly reduces the time required for model training and prediction.

In well log prediction while drilling, accuracy is a crucial reference factor. In this study, the models are evaluated based on the time spent, using the same personal computer with NVIDIA GPU (GTX 1650 with 4 GB video storage) and the same set of parameters for training and prediction. Table 6 and Figure 14 show the time spent on the models. Results reveal that the running time of the LSTM model is the shortest among the three methods, but its accuracy is the lowest. Bi-LSTM and DC-LSTM models exhibit comparable higher accuracy, with DC-LSTM being more efficient in terms of computational time.

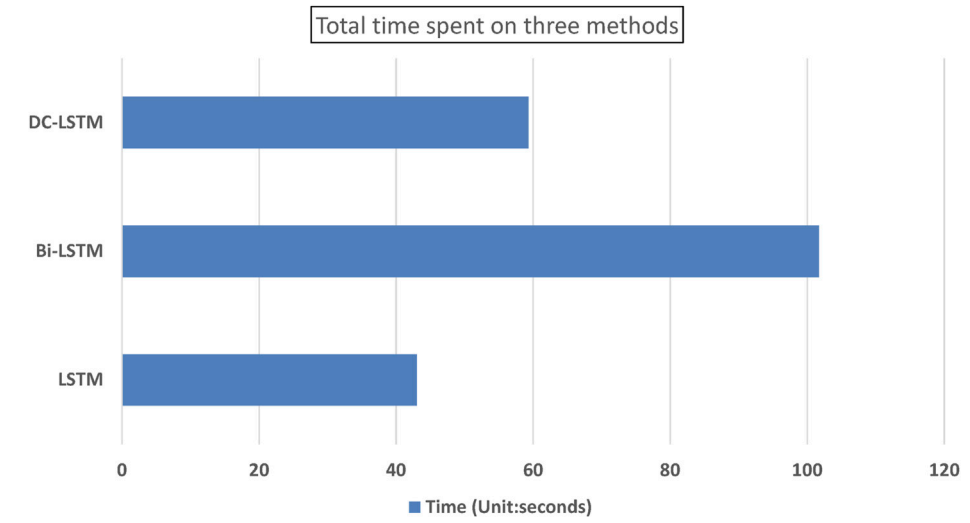


FIGURE 14
Total time spent on three methods.

4 Conclusion

This paper presents a novel approach for predicting well logs using artificial intelligence while drilling, utilizing seismic impedance and three related LSTM networks. The proposed approach incorporates the constraints of sedimentary formation seismic impedance at the drill bit and the changing trend of impedance through the upper and lower formations. This optimization of the prediction process results in improved computational efficiency. Through experimentation, the Bi-LSTM and DC-LSTM derivative networks are demonstrated to have higher prediction accuracy than the base LSTM. Notably, the DC-LSTM network, developed in this paper, is more efficient in reducing the number of training parameters and computational time without compromising prediction accuracy, as compared to the Bi-LSTM. Field data tests of the three networks were conducted using two wells, and the results demonstrate their effectiveness and efficiency.

The data-driven approach utilizing the three networks is effective for predicting well logs in laterally homogeneous target formations, particularly in the case of clastic sedimentation. However, if the formation exhibits significant lateral heterogeneity, such as fractures and faults near the drilling bits, the prediction accuracy may be compromised. Therefore, the challenge of making reliable predictions under such conditions with limited quality data and rapidly changing trends is an important research topic for the future.

Data availability statement

The datasets presented in this article are not readily available because Oilfield sensitive. Requests to access the datasets should be directed to yungui.xu@outlook.com.

Author contributions

HW was responsible for the preliminary literature research, the writing of the whole manuscript, and the drawing of the figure. YX and ST were responsible for the late revision and submission of the manuscript. LW, WC, and XH were responsible for providing inspiration and ideas for the manuscript.

Funding

National Natural Science Foundation of China: U20B2016. Study on high temperature high pressure seismic rock model and integrated intelligent reservoir characterization methods for a mid-deep reservoir in Yingqiong basin.

Conflict of interest

LW was employed by Shale Gas Project Management Department of CNPC Chuanqing Drilling Engineering Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Adamowski, J., and Chan, H. F. (2011). A wavelet neural network conjunction model for groundwater level forecasting. *J. Hydrology* 407 (1-4), 28–40. doi:10.1016/j.jhydrol.2011.06.013

Ahmed Ali Zerrouki, T. A., and Baddari, K. (2014). Prediction of natural fracture porosity from well log data by means of fuzzy ranking and an artificial neural network in Hassi Messaoud oil field, Algeria. *J. Petroleum Sci. Eng.* 115, 78–89. doi:10.1016/j.petrol.2014.01.011

Alizadeh, B., Najjari, S., and Kadkhodaie-Ilkhchi, A. (2012). Artificial neural network modeling and cluster analysis for organic facies and burial history estimation using well log data: A case study of the south pars gas field, Persian gulf, Iran. *Comput. Geosciences* 45, 261–269. doi:10.1016/j.cageo.2011.11.024

da Silva, J. A. C. I., and Saggioro, N. J. (2013). *Recurrent neural network based approach for solving groundwater hydrology problems*. London, United Kingdom: IntechOpen.

Deng, H., Zhang, L., and Wang, L. (2019). Global context-dependent recurrent neural network language model with sparse feature learning. *Neural Comput. Appl.* 31 (2), 999–1011. doi:10.1007/s00521-017-3065-x

Graves, A., and Jaitly, N. (2014). *Towards end-to-end speech recognition with recurrent neural networks*. Beijing, China: International Conference on Machine Learning.

Graves, A., and Schmidhuber, J. (2005). *Framewise phoneme classification with bidirectional LSTM networks*. Montreal, QC: 2005 IEEE International Joint Conference on.

Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. *Neural Comput.* 9 (8), 1735–1780. doi:10.1162/neco.1997.9.8.1735

Hrynaszkiewicz, I. (2010). A call for BMC Research Notes contributions promoting best practice in data standardization, sharing and publication. *BMC Res. notes* 3, 235. doi:10.1186/1756-0500-3-235

Iturrarán-Viveros, U., and Parra, J. O. (2014). Artificial Neural Networks applied to estimate permeability, porosity and intrinsic attenuation using seismic attributes and well-log data. *J. Appl. Geophys.* 107, 45–54. doi:10.1016/j.jappgeo.2014.05.010

Jin, Z. D. C. Y. M. (2018). *Synthetic well logs generation via recurrent neural networks*. Beijing, China: Petroleum Exploration and Development (4), 629–639.

Lokeshkumar, R., Jayakumar, K., Prem, V., and Nonghuloo, M. S. (2020). Analyses and modeling of deep learning neural networks for sequence-to-sequence translation(Article). *Int. J. Adv. Sci. Technol.* 29 (5), 3152–3159.

Long, W., Chai, D., and Aminzadeh, F. (2016). *Pseudo density log generation using artificial neural network*. Anchorage, AK: SPE Western Regional Meeting.

Mo, X., Zhang, Q., and Li, X. (2015). *Well logging curve reconstruction based on genetic neural networks*. Jilin Changchun, China: Jilin University. College of Geoexploration Science and Technology.

Rolon, L., Mohaghegh, S. D., Ameri, S., Gaskari, R., and McDaniel, B. (2009). Using artificial neural networks to generate synthetic well logs. *J. Nat. Gas Sci. Eng.* 1 (4-5), 118–133. doi:10.1016/j.jngse.2009.08.003

Sahoo, S., Russo, T. A., Elliott, J., and Foster, I. (2017). Machine learning algorithms for modeling groundwater level changes in agricultural regions of the U.S. *U.S. Water Resour. Res.* 53 (5), 3878–3895. doi:10.1002/2016wr019933

Salehi, M. M., Rahmati, M., Karimnezhad, M., and Omidvar, P. (2017). Estimation of the non records logs from existing logs using artificial neural networks. *Egypt. J. Petroleum* 26 (4), 957–968. doi:10.1016/j.ejpe.2016.11.002

Schuster, M., and Paliwal, K. K. (1997). Bidirectional recurrent neural networks. *IEEE Trans.*45 (11), 2673–2681. doi:10.1109/78.650093

Shan, L., Liu, Y., Tang, M., Yang, M., and Bai, X. (2021). CNN-BiLSTM hybrid neural networks with attention mechanism for well log prediction. *J. Petroleum Sci. Eng.* 205, 108838. doi:10.1016/j.petrol.2021.108838

Silva, A. A., Neto, I. A. L., Misságia, R. M., Ceia, M. A., Carrasquilla, A. G., and Archilha, N. L. (2015). Artificial neural networks to support petrographic classification of carbonate-siliciclastic rocks using well logs and textural information. *J. Appl. Geophys.* 117, 118–125. doi:10.1016/j.jappgeo.2015.03.027

Silversides, K., Melkumyan, A., Wyman, D., and Hatherly, P. (2015). Automated recognition of stratigraphic marker shales from geophysical logs in iron ore deposits. *Comput. Geosciences* 77, 118–125. doi:10.1016/j.cageo.2015.02.002

Singh, U. K. (2011). Fuzzy inference system for identification of geological stratigraphy off Prydz Bay, East Antarctica. *J. Appl. Geophys.* 75 (4), 687–698. doi:10.1016/j.jappgeo.2011.08.001

Tamim, N., Laboureur, D. M., Hasan, A. R., and Mannan, M. S. (2019). Developing leading indicators-based decision support algorithms and probabilistic models using Bayesian network to predict kicks while drilling. *Process Saf. Environ. Prot. Trans. Institution Chem. Eng. Part B* 121, 239–246. doi:10.1016/j.psep.2018.10.021

Wang, G., Carr, T., Ju, Y., and Li, C. (2014). Identifying organic-rich Marcellus Shale lithofacies by support vector machine classifier in the Appalachian basin. *Comput. GEOSCIENCES* 64, 52–60. doi:10.1016/j.cageo.2013.12.002

Wang, J., Cao, J., Liu, Z., Zhou, X., and Lei, X. (2020). Method of well logging prediction prior to well drilling based on long short-term memory recurrent neural network(Article). *J. Chengdu Univ. Technol. Sci. Technol. Ed.* 47 (2), 227–236. doi:10.3969/j.issn.1671-9727.2020.02.11

Wang, W., Qiu, Z. S., Zhong, H. Y., Huang, W. A., and Dai, W. H. (2017). Thermo-sensitive polymer nanospheres as a smart plugging agent for shale gas drilling operations. *J. Xi'an Shiyou Univ. Nat. Sci. Ed.* 32 (2), 116–125. doi:10.1007/s12182-016-0140-3

Zhang, C. J., Chen, L., Tong, Y., and Liu, Z. (2015). *Cleaning uncertain data with a noisy crowd*. Hong Kong, China: Hong Kong University of Science and Technology.

Frontiers in Earth Science

Investigates the processes operating within the major spheres of our planet

Advances our understanding across the earth sciences, providing a theoretical background for better use of our planet's resources and equipping us to face major environmental challenges.

Discover the latest Research Topics

See more →

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact