

## Original article

# Acoustic impedance inversion in coal strata using the priori constraint-based TCN-BiGRU method

Suzhen Shi<sup>1</sup>\*, Youchao Qi<sup>3</sup>, Weiming Chang<sup>2</sup>, Li Li<sup>2</sup>, Xuejun Yao<sup>2</sup>, Jingxue Shi<sup>2</sup>

<sup>1</sup>State Key Laboratory of Coal Resources and Safe Mining, China University of Mining and Technology-Beijing, Beijing 100083, P. R. China

<sup>2</sup>School of Geoscience and Surveying Engineering, China University of Mining and Technology-Beijing, Beijing 100083, P. R. China

<sup>3</sup>Nanchang Urban Planning & Design Institute Group Co., Ltd, Nanchang 330038, P. R. China

### Keywords:

Acoustic impedance inversion  
deep learning  
priori constraint  
TCN-BiGRU network

### Cited as:

Shi, S., Qi, Y., Chang, W., Li, L., Yao, X., Shi, J. Acoustic impedance inversion in coal strata using the priori constraint-based TCN-BiGRU method. *Advances in Geo-Energy Research*, 2023, 9(1): 13-24.  
<https://doi.org/10.46690/ager.2023.07.03>

### Abstract:

Acoustic impedance inversion is a key technique for the seismic exploration of coalfield, which can determine subsurface lithological changes and coal seam distribution. The traditional method is highly subjective, has poor generalizability, and interpretation can be time and labor consuming. Due to the powerful nonlinear interpretation and feature extraction capabilities of neural networks, deep learning technology has demonstrated potential for geophysical exploration. To predict acoustic impedance accurately and efficiently, this study proposes the use of the initial geological model as the priori constraint for training. The low-frequency feature extraction capability of a bidirectional gated recurrent unit network and the high-frequency feature extraction capability of a temporal convolutional network are used to establish a new acoustic impedance inversion method in coal strata with a priori constraint data. The temporal convolutional network-bidirectional gated recurrent unit method was applied to data from the Xinjing Mining Area in Shanxi province, northern China. The results displayed good precision by accurately predicting the distribution and thickness variation of local coal seams. Compared with the traditional model-based method and the method using temporal convolutional network-bidirectional gated recurrent unit network, the proposed priori constraint-based temporal convolutional network-bidirectional gated recurrent unit network has better feature expression capability and provides more detailed coal seam information. In conclusion, the new method can improve the accuracy of acoustic impedance inversion, which is of great significance for coalfield seismic exploration.

## 1. Introduction

Acoustic impedance (AI) inversion utilizes seismic, well logging, and horizon data, as well as geological information, to interpret and image subsurface lithological changes, coal seam distribution, and structural development. It is a key technique used in coalfield seismic exploration (Wang et al., 2005). Traditional inversion methods rely too much on well data, especially in complex areas, and the accuracy is lower (Mustafa et al., 2021).

Machine learning now represents the mainstay of tools applied to geoscience research (Li et al., 2023). Deep learning (DL) is a new and efficient technology that has undergone

widespread development and utilization in academia and industry in recent years. The most notable difference between DL and ordinary machine learning is that its neural network contains more hidden processing layers (Janiesch et al., 2021). DL uses low-level feature information to represent high-level feature information and involves multiple processing layers or multiple non-linear transformations and repetitive structures to abstract high-dimensional data (Wang et al., 2020). Seismic inversion based on DL has gradually addressed the problems that traditional inversion methods could not solve. It can automatically extract complex feature parameters from large volumes of geological data, which is highly significant for

accelerating the development of intelligent AI inversion. DL network algorithms for AI inversion can be divided into two main categories: convolutional neural networks (CNNs) and recurrent neural networks (RNNs).

Geophysicists have applied many convolutional neural networks and related variants to AI inversion. Das et al. (2019) used a CNN-based AI inversion method on the Volve field dataset from offshore Norway, the inversion results manifested a high correlation (82%) with the AI log. Biswas et al. (2019) used CNN to estimate AI based on post-stack seismic traces, this method achieved accurate AI evaluation results and improved efficiency. Wu et al. (2020) used a fully convolutional residual network for seismic AI inversion, which can effectively predict the AI and show robustness against noise and phase difference. Wu et al. (2021a) proposed a semi-supervised learning workflow on account of a generative adversarial network (GAN) for AI inversion. The inversion results showed that the proposed method, due to making use of both labeled data and unlabeled data, is more approach to field data than conventional CNN-based inversion methods. Chen et al. (2021a) built a deep convolutional neural network to implement seismic AI inversion. The network has achieved a high level of inversion accuracy under a relatively fast computational speed and demonstrated effective through a comprehensive analysis of field data. Wu et al. (2021b) improved the 1 dimensional (1D) network by replacing it with a 2 dimensional (2D) CNN and incorporating constraints from the initial AI model. The 2D CNN exhibited stronger robustness to noise, achieved better recovery of thin layers, and produced more consistent lateral impedance models compared to the 1D CNN with the same network structure and training logs. Yoo et al. (2022) proposed a domain adaptation model to AI inversion and predicted it more accurately compared to the standard CNN model especially in regions without well logging data. Tao et al. (2023) obtained an AI profile by inputting seismic profile and background impedance into a well-trained self-attention U-Net. The self-attention U-Net proved to be robust against noise. Furthermore, it showed superior spatial continuity compared to deconvolution with recursive inversion and Total Variation regularization.

These studies have demonstrated that traditional CNN methods can predict AI from seismic data at limited frequency bands, but there is room for improvement. When traditional CNNs train seismic and well logging data, the continuity of geophysical data and correlations between similar points are not considered, and the long-term information is not accurately captured. Temporal convolutional networks (TCNs) are the variants of CNNs. TCNs can effectively extract and represent sequential features at different abstract levels in the input sequence through dilated causal convolution operations and residual blocks, so as to capture high-frequency information. Mustafa et al. (2019) proposed a workflow that utilizes a temporal convolutional network (TCN)-based network architecture to predict AI by treating the problem as a sequence modeling task. The workflow overcomes the issue of overfitting encountered in CNN. Later, Mustafa et al. (2021) introduced local spatial context information in a TCN and then performed sequence modeling on seismic traces, achieving highly precise

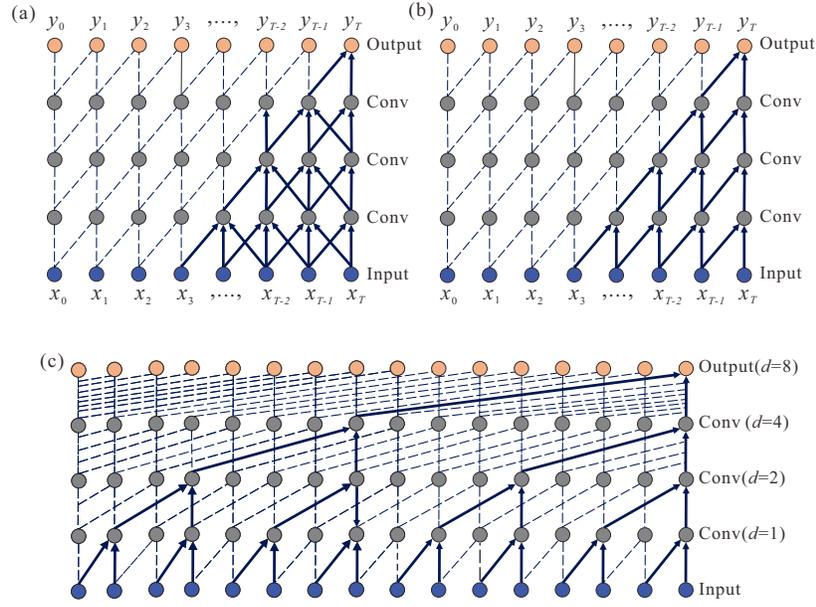
estimation of AI. Smith et al. (2022) used a TCN to invert seismic data with coherent noise and produced better results than a model-based inversion. Wang and Chen (2022) proposed an AI inversion method based on a data-driven TCN model and the results show that the TCN model have better reconstruction effects in complex strata. Marques et al. (2022) compared different DL networks (LSTM, TCN, CNN, GAN) for AI inversion. The TCN exhibited the best performance and generalization in the inversion results due to its multiple convolutional layers in every Temporal Block and a mass of neurons.

For large volumes of geophysical data, diverse physical attributes, and sequential data, RNNs are suitable for processing sequential data and have been widely used in seismic inversion. A traditional RNN is prone to vanishing or exploding gradient problems with an increase in time units and number of steps (Yoon et al., 2020). Long short-term memory (LSTM) networks were developed to overcome these issues with RNNs. Zhao et al. (2023) proposed a data-driven, high-resolution AI inversion method based on a bidirectional LSTM, which produced accurate and reliable inversion results. Due to the complex structure and many internal parameters of LSTM networks, they have considerable computational requirements. Therefore, LSTM networks were simplified and upgraded to form a gated recurrent unit (GRU). Alfarraj and AlRegib (2019) connected three layers of GRUs in series to construct a sequence modeling sub-module in the inversion model, which is used to capture low-frequency trends of AI. Song et al. (2021) used a series of GRUs to develop a global feature extraction layer to enrich low-frequency information in AI inversion results.

The inversion results of most current methods based on a single network suffer from poor accuracy, instability, and low resolution. Moreover, well logging label data of real mining areas are always limited and sparse, and it is difficult to enhance the accuracy of inversion results by obtaining more well logging labels (Adler et al., 2021). To resolve the above issues, a reasonable initial geological model (the priori constraint) was introduced into the network as a constraint for training, and an AI inversion method using the priori constraint-based temporal convolutional network-bidirectional gated recurrent unit (TCN-BiGRU) network was developed in this work. This method was then applied to the Xinjing Mining Area to predict coal seam distribution.

## 2. Methods

TCNs can effectively extract and represent sequential features at different abstract levels in the input sequence through dilated causal convolution and residual connections, so they can capture high-frequency information. Bidirectional gated recurrent units (BiGRUs) are used to capture the long-term sequential correlation of data, which can effectively extract useful features from low-frequency information. By combining the advantages of TCNs and BiGRUs, high-frequency and low-frequency sequential information were comprehensively utilized to improve the accuracy of AI inversion.



**Fig. 1.** Convolution diagram (Wang and Chen, 2022). (a) Traditional convolution, (b) causal convolution and (c) dilated causal convolution.

## 2.1 TCN

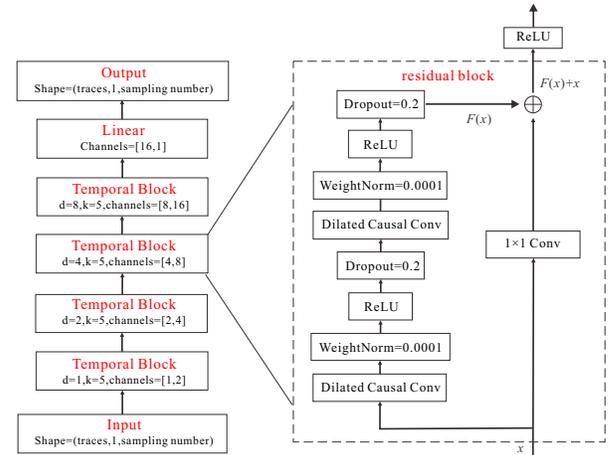
The TCN network structure is composed of several Temporal Blocks. Each Temporal Block contains a residual block and a ReLU function, which can better capture the feature changes in sequence data at different scales (Yang et al., 2023).

Sequential data is a data column recorded in chronological order. Traditional convolution considered data points independent of each other (Fig. 1(a)). The causal convolution (Fig. 1(b)) captures the causal correlation of sequential data, and the results obtained are associated with previously observed data. The calculation process of the causal convolution can be written as:

$$p(x) = \prod_{t=1}^T p(x_t | x_1, \dots, x_{t-1}) \quad (1)$$

where  $t$  is any time point between 1 and  $T$ ,  $x_t$  is the  $t$ -th component of input sequential data  $X^T = \{x_1, x_2, \dots, x_T\}$ , and  $p(x)$  is the output data of the current observation point.

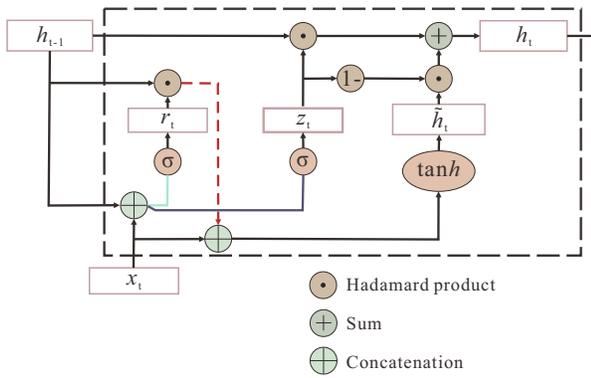
Since the size of the causal convolutional receptive field depends on the convolutional layers and the convolutional kernel, enough causal convolutional layers are needed to capture more data features when the sequence is extended. When more information can be traced back, more feature extraction layers should be stacked, and the layers of the network model deepens. This makes the model more cumbersome and complex and also increases the hyperparameters. To deal with this, the TCN model introduces a dilated convolutional structure to form a dilated causal convolution. Compared with the causal convolution (Fig. 1(b)), the dilated causal convolution allows for a sampling interval, and the sampling rate is controlled by the expansion coefficient  $d$ . When  $d = 1$ , each data point would be used as the input. When  $d = 2$ , data points with an interval of 2 would be used for the input, and so on. Generally speaking, the larger the number of network layers, the larger



**Fig. 2.** Structure of the TCN model.

the  $d$ , that is, the size of the effective window increases at an exponential rate with the number of convolutional layers. Thus, when the number of convolutional layers remains the same, the dilated causal convolution enlarges the receptive field of the network, greatly reduces the network parameters (Sun, 2020). Fig. 1(c) shows a dilated causal convolution with a convolutional kernel size of 2 and expansion coefficients of 1, 2, 4, and 8.

As shown in Fig. 2, The Temporal Block contains a residual block (Zhang et al., 2023) and a ReLU function. After input data  $x$  enters the Temporal Block, feature information in  $x$  is extracted through the first dilated causal convolution. Then weight normalization scales the weights to keep their values within a small range which can reduce the impact of weight differences, improve model stability and generalization ability. In the next part, the ReLU function is then used for activation.



**Fig. 3.** Internal structure of a GRU.

## 2.2 TCN-BiGRU

Later, the output of the ReLU enters the dropout regularization which is used to alleviate the occurrence of overfitting and improve the generalization ability of the network. This is followed by a second dilated causal convolution, performing the same operation as before to get the output value  $F(x)$ . Finally, the original input  $x$  is subjected to a 11 convolution operation, and then the residual calculation is performed with  $F(x)$  to obtain the final output value  $F(x) + x$ .

## 2.3 BiGRU

GRUs are a unique version of an RNN, which can save traversed features and learn to discard irrelevant features. Unlike a standard RNN, GRU selectively forgets information, so it can be trained to retain old information. Moreover, the GRU model has fewer gates and parameters than the LSTM model, resulting in less computation and faster processing (Shen et al., 2021). Fig. 3 shows the internal structure of a GRU.

A GRU has two important components, the reset gate and update gate. The effect of the reset gate is to integrate feature data input from the current moment with previously retained information. The effect of the update gate is to determine how much previously retained information to save in the current moment. The unique characteristic of a GRU is that the gated units constantly retain and remove information, allowing it to retain long-term sequence information. The formulas are as follows:

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \quad (2)$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \quad (3)$$

$$\tilde{h}_t = \tanh[W x_t + U(r_t \odot h_{t-1}) + b] \quad (4)$$

$$h_t = z_t \odot h_{t-1} + (1 - z_t) \odot \tilde{h}_t \quad (5)$$

where  $z_t$  is the output of the update gate at time point  $t$ ,  $r_t$  is the output of the reset gate at time point  $t$ ,  $x_t$  is the  $t$ -th component of the input sequence data  $x$ ,  $h_{t-1}$  are the hidden layer states at the previous time point  $t-1$ ,  $h_t$  are the hidden layer states at time point  $t$ ,  $\tilde{h}_t$  are the candidate hidden layer

states at time point  $t$ ,  $\odot$  is Hadamard product which means multiplication of the corresponding elements of matrixes,  $W_z$  and  $U_z$  are the weight matrices of the update gate,  $W_r$  and  $U_r$  are the weight matrices of the reset gate,  $W$  and  $U$  are the weight matrices of the candidate hidden layer states,  $b_z$ ,  $b_r$ ,  $b$  are the biases of the update gate, reset gate, and candidate hidden layer states, respectively,  $\tanh$  and  $\sigma$  (Sigmoid) are activation functions. As Fig. 3 shows,  $z_t$  controls how many hidden states  $h_{t-1}$  and the input  $x_t$  flow into the hidden states  $h_t$ . The larger the value of  $z_t$ , the more information flows in  $h_t$ . The  $z_t$  can also control how many hidden states  $h_t$  should be updated by the candidate hidden states  $\tilde{h}_t$ . The  $r_t$  controls how many hidden states  $h_{t-1}$  and the input  $x_t$  flow into the candidate hidden states  $\tilde{h}_t$ . The larger the value of  $r_t$ , the more information flows in  $\tilde{h}_t$ .

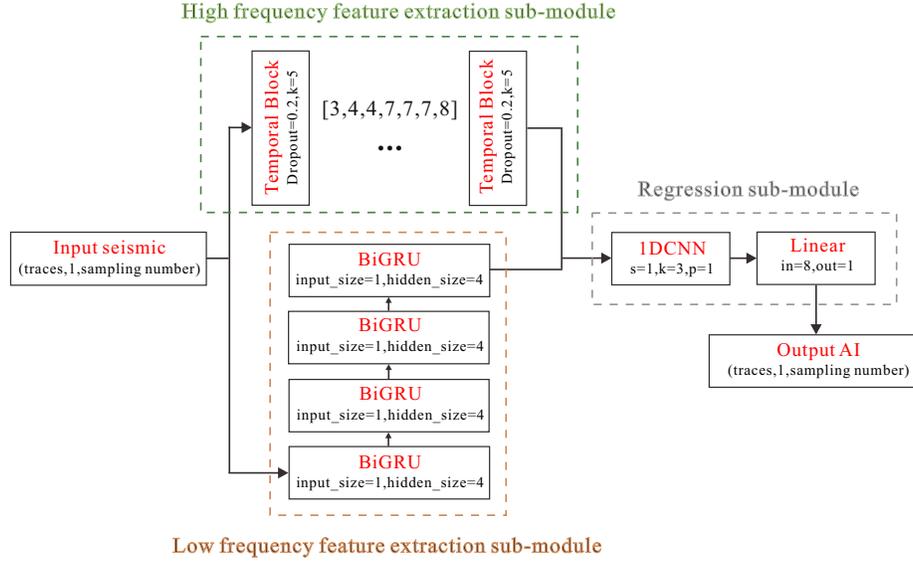
A BiGRU consists of two unidirectional GRUs with opposite directions (forward and backward propagation), and its output is determined jointly by the two GRUs, therefore, BiGRU has the ability to learn from both past and future information. BiGRU can discover the regularity in input sequential data through forward and backward propagation, and can better handle abnormal situations such as noise and missing values in input sequences.

Based on the above mechanisms, the BiGRU can control the flow and retention of information, making the network more capable of capturing and utilizing long-term dependencies in sequential data and the BiGRU provides more comprehensive contextual information and stronger memory capabilities, resulting in better performance in processing sequential data. Finally, the BiGRU model for predicting AI with 6 BiGRU hidden layers and a fully connected layer is constructed. Its input dimension is 1, the size of the each hidden layer is 7, and the output dimension of the fully connected layer is 1.

A TCN-BiGRU network model combines the advantages of both models. By training the TCN-BiGRU network model, geophysical data can be captured more rationally, and the nonlinear relationship between input seismic data and output AI can be obtained more accurately. Fig. 4 shows that the TCN-BiGRU network model consists of three main parts: the high-frequency feature extraction sub-module, the low-frequency feature extraction sub-module, and the regression sub-module.

The high-frequency feature extraction sub-module consists of seven Temporal Blocks, and the output channels of each Temporal Block are specified in the square bracket (Fig. 4). Each Temporal Block contains 2 dilated causal convolutions. The convolutional kernel size is set to 5, the stride is set to 1, the padding is 4, 8, 16, 32, 64, 128, and 256, respectively. The expansion coefficient  $d$  is 0, 2, 4, 8, 16, 32, and 64, respectively. The convolutional layer can operate on a small window of input features and obtain the high-frequency part of the AI. However, Temporal Blocks cannot preserve states like BiGRUs can, therefore, they cannot capture low-frequency trends.

The low-frequency feature extraction sub-module consists of 4 BiGRUs (Fig. 4). When inputting training data into the BiGRU, the data feature changes are recorded based on the



**Fig. 4.** Structure of the TCN-BiGRU network model.

sequence, and states are calculated based on future and past predictions. The concatenation of four BiGRUs is equivalent to an upgraded deep network, and is capable of extracting more complex nonlinear relationships between features and labels. Furthermore, BiGRUs commonly produce the low-frequency part of the AI.

The regression sub-module includes a 1D CNN and a simple fully connected layer (Fig. 4). The convolutional kernel size of the 1D CNN is set to 3, the step size is set to 1, and the padding is set to 1. The 1D CNN can learn local and global correlation features in input sequences. Then, the fully connected layer combines and models features extracted from the 1D CNN at a higher level and reduces the dimension of the 1D CNN output. In brief, the regression sub-module helps improve the model's understanding and predictive performance of the input data.

## 2.4 Inversion principle

Both seismic and well logging data can be regarded as sequential data. Using post-stack seismic data as the input, predicted AI as the output, and actual well logging AI as the labels, AI inversion can be considered as sequence modeling between seismic data and well logging AI labels. Supposing that  $X = \{x_1, x_2, \dots, x_n\}$  is a set of post-stack seismic data used as features, in which  $x_i$  is the post-stack seismic record of the  $i$ -th trace, and  $Y = \{y_1, y_2, \dots, y_n\}$  are the corresponding AI labels. The formulas are as follows:

$$\theta^{l+1} = \theta^l - \eta \cdot V_{\theta} L(\theta^l) \quad (6)$$

$$\hat{\theta} = \arg_{\theta} \min L(y_i, f_{\theta}(x_i)) \quad (7)$$

where  $\theta$  are a set of weights and biases,  $l$  is the current number of epochs,  $\eta$  is the learning rate,  $V_{\theta} L(\theta^l)$  is the loss function gradient,  $y_i$  is the value of the  $i$ -th label data which refers to real AI,  $f_{\theta}(x_i)$  is the value of the  $i$ -th predicted AI,  $L(y_i, f_{\theta}(x_i))$

is the loss value between predicted AI  $f_{\theta}(x_i)$  and the real AI  $y_i$ , and  $\hat{\theta}$  is the final parameter set of the inversion model when the loss value cannot be reduced any further. First, the subset (the  $i$ -th trace of seismic data) of  $X$  is preprocessed and input into the neural network model, and the predicted AI  $f_{\theta}(x_i)$  is obtained by forward propagation. The loss function between the predicted AI  $f_{\theta}(x_i)$  and real AI  $y_i$  is calculated to determine the gradient  $V_{\theta} L(\theta^l)$ . And the iterative updating of model hyperparameter  $\theta$  is then achieved by backpropagation. Finally, the above process is repeated until the loss function cannot be reduced any further.  $\arg_{\theta} \min$  is used to obtain the optimal parameter set  $\hat{\theta}$  when  $L(y_i, f_{\theta}(x_i))$  reaches its minimum value.

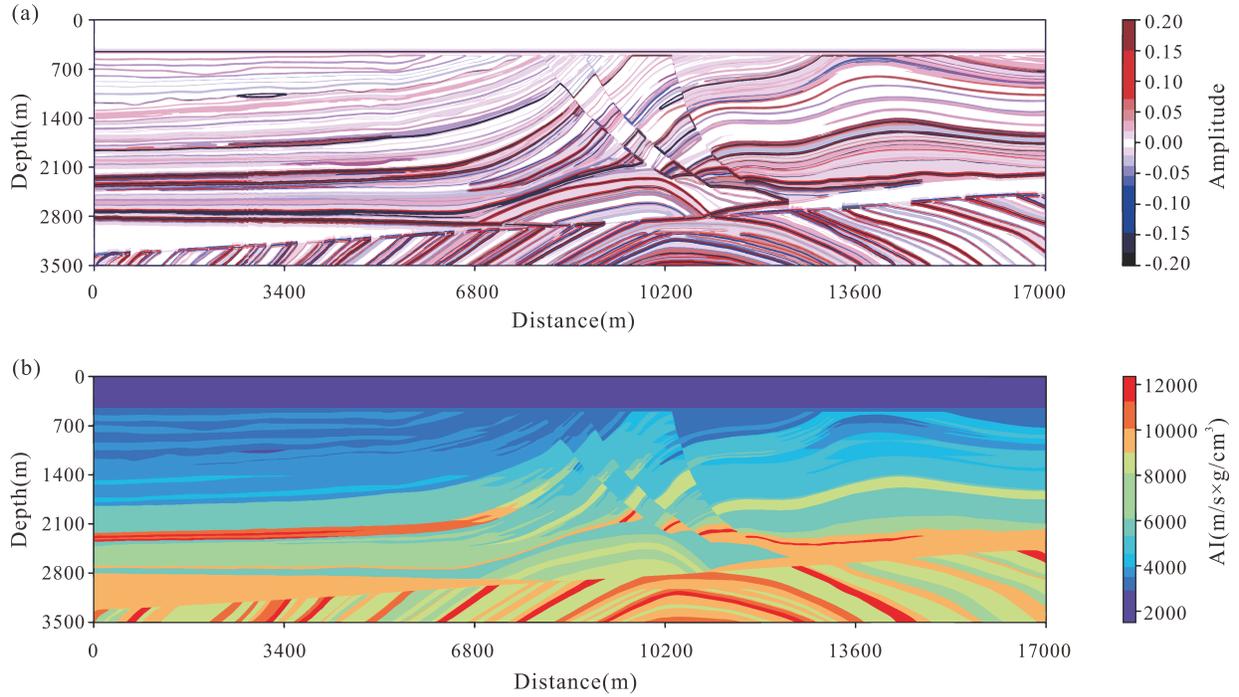
## 3. Inversion experiment

The publicly available data used in this study were obtained from the Marmousi seismic model created by the French Institute of Petroleum, which has been widely used to research the validity and rationality of advanced geophysical methods, including seismic inversion, modeling, and imaging (Martin et al., 2006). The model's data includes 199 geological layers, and its geologic body includes eroded rivers, structural overturns, reservoir sweet spots, and sedimentary development, with various lithological compositions. The main part of the central structurally complex area is an anticline structure and overturning surface. The Marmousi model data covers an area 17 km long and 3.5 km deep. A single trace contains 701 recorded points, and the whole model contains 2,721 seismic traces and corresponding AI.

### 3.1 Data processing

#### 3.1.1 Data removal

To facilitate processing and visualization using Python, the Marmousi model was converted from SEG-Y format to NPY



**Fig. 5.** Marmousi model data. (a) Seismic profile (Marmousi model) and (b) AI profile (Marmousi model).

format. The array shape of the seismic profile and AI profile in NPY format were (2,721, 701), i.e., the profile contains 2,721 traces of seismic data or AI, each trace of data having 701 recorded points. Fig. 5 shows that the first 101 points (corresponding to a depth of less than 500 m) recorded in each trace of the seismic and AI profile of the Marmousi model are null values, which are not useful for network training. After cutting them out, the array shape was (2,721, 600).

### 3.1.2 Dividing the dataset

Fig. 5 shows that the left and right ends have relatively simple structures, with a plane of unconformity at the bottom, and a complex middle, with anticlines, normal faults, and unconformity structures. To adequately test the predictive efficacy of the model, the middle traces from 499 to 2,399 were extracted for training. To simulate the actual mining areas, the extraction of the seismic data and the AI follows the principle of sparse and uniform. So the seismic data and the AI were extracted at equal intervals of 20 traces as the training set. The rest of the 1,880 traces were used in the validation set. The entire data (2,721×600) was used as a test set.

### 3.1.3 Data normalization

The different sampling points, data volumes, and value ranges of the input seismic data and AI make the identification and processing of model data difficult and lead to reduced efficiency in computational and convergence. Z-score normalization was used to standardize the original data (Abdi and Williams, 2010). The formula is as follows:

$$\hat{X} = \frac{X - X_m}{S_1} \quad (8)$$

$$\hat{Y} = \frac{Y - Y_m}{S_2} \quad (9)$$

where  $\hat{X}$  is the normalized seismic data of the training set,  $X_m$  is the mean value of the  $X$ ,  $\hat{Y}$  is the normalized AI of the training set,  $Y_m$  is the mean value of the  $Y$ ,  $S_1$  is the standard deviation of the  $X$ ,  $S_2$  is the standard deviation of the  $Y$ . The processed datasets conform to the standard normal distribution, that is, the mean value of each dataset is 0, and the standard deviation of each dataset is 1.

## 3.2 Model training and validation

Mean squared error (MSE) is widely used in regression problems as a loss function. It describes the difference between predicted AI  $f_{\theta}(x_i)$  and the real AI  $y_i$  and is expressed as follows:

$$L(y_i, f_{\theta}(x_i)) = \frac{1}{n} \sum_{i=1}^n (y_i - f_{\theta}(x_i))^2 \quad (10)$$

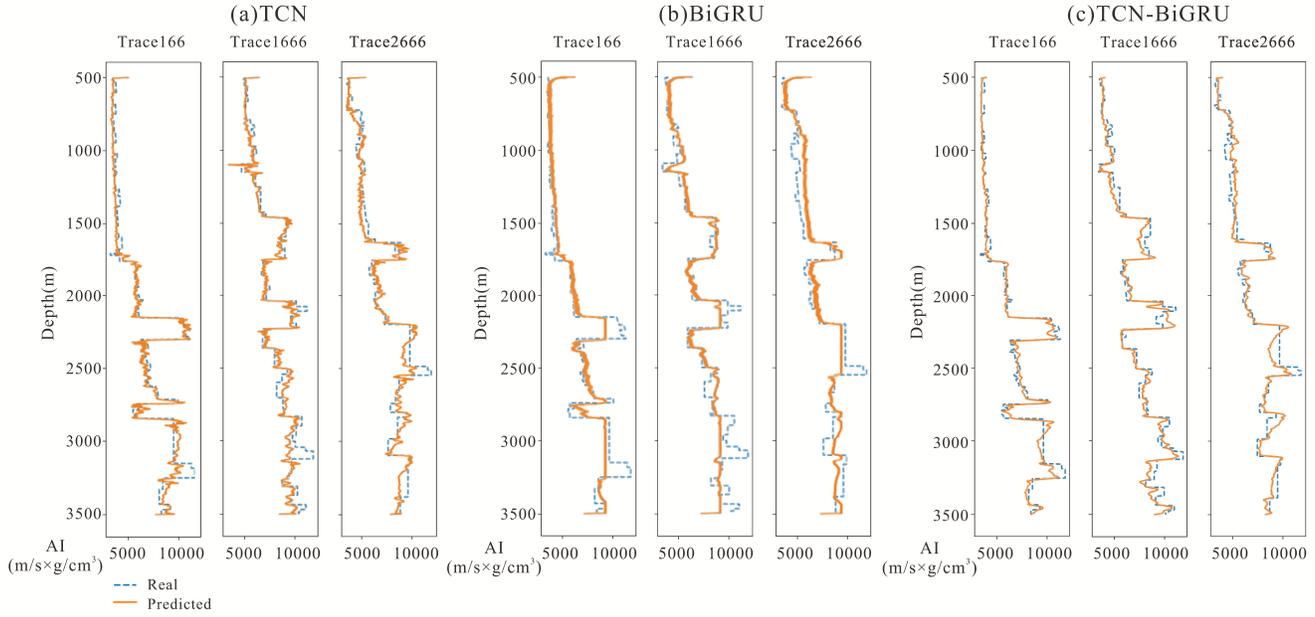
where  $n$  is the number of seismic traces in the training set.

Adam (Adaptive moment estimation) algorithm is an adaptive estimation algorithm based on low-order moments that can replace the conventional stochastic gradient descent method. It can calculate the adaptive learning rate of different parameters according to the first and second moment estimation of the gradient. It has high computational efficiency and low memory requirements (Jais et al., 2019). So Adam algorithm was used as the neural network optimization algorithm.

Network parameters were optimized by minimizing the loss function through multiple epochs. The TCN model was trained for 1,000 epochs and had a learning rate of 0.001 and batch size of 15. The BiGRU model was trained for 100 epochs and had a learning rate of 0.001 and batch size of 45. The

**Table 1.** Training evaluation of the TCN, BiGRU, and combined models.

Model	Epochs	Learning rate	Min trainloss	Min valloss
TCN	1,000	0.005	0.173	0.175
BiGRU	100	0.001	0.093	0.068
TCN-BiGRU	220	0.005	0.032	0.025

**Fig. 6.** Single-trace prediction results of the TCN, BiGRU, and combined models. (a) TCN, (b) BiGRU and (c) TCN-BiGRU.

TCN-BiGRU model was trained for 220 epochs and had a learning rate of 0.005 and a batch size of 10. Table. 1 lists the minimum loss of the various models.

The Marmousi seismic data in the test set was inputted into the above three saved models to obtain the predicted AI curves. Fig. 6 demonstrates the predicted AI of the three models at the 166<sup>th</sup>, 1,666<sup>th</sup>, and 2,666<sup>th</sup> traces as well as the real values at corresponding positions (yellow lines are predicted results, and blue lines are real results). The correlation coefficients of the TCN, BiGRU, and TCN-BiGRU models were 0.820, 0.829, and 0.957, respectively. Fig. 6(a) shows the single-trace prediction results of the TCN model, which reflect the overall trend, but there is sharp curve jitter. And the proportion of high-frequency components in the curves is high. Fig. 6(b) shows the single-trace prediction results of the BiGRU model, which reflect the lithological changes of the thick layers, but the curves are flat and do not reflect the high-frequency information of the thin layers. Fig. 6(c) shows that the single-trace prediction results of the TCN-BiGRU model are markedly better than the other two models. The predicted results have a high degree of matching with the real data in the key strata, and the thin layers are accurately depicted.

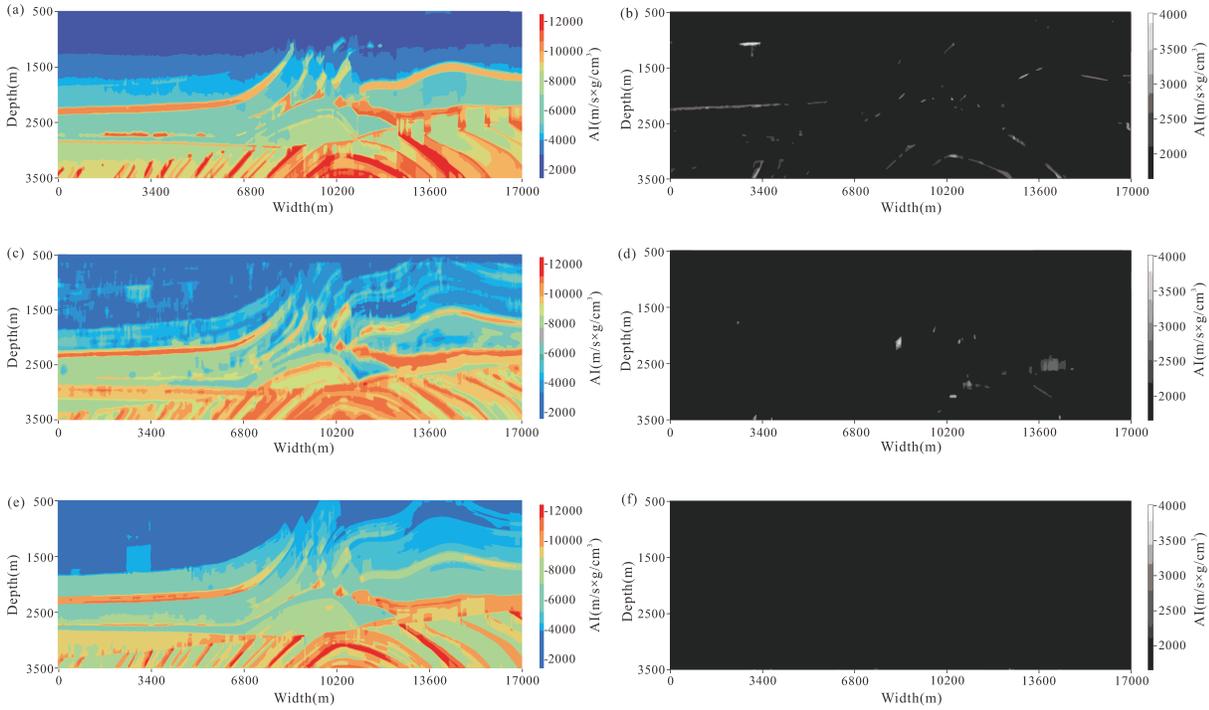
The seismic data was imported into the three models to obtain the predicted AI profiles and absolute error profiles. Fig. 7(a) shows that the TCN model can identify the structural

developments and strata trends, but there are issues of vertical discontinuity. Fig. 7(b) shows that the errors are mainly concentrated in areas with developed structures. Fig. 7(c) shows that the BiGRU model clearly identifies changes in lithological interfaces and strata structures, produces excessive homogeneity of stratum, and is unable to identify thin layers. Fig. 7(d) shows that the errors mainly focus on the thin layers. Fig. 7(e) shows that the TCN-BiGRU model clearly identifies changes in lithological interfaces and strata structures, and thin layers can be distinguished. Fig. 7(f) shows that the difference between the predicted AI and the real AI is very small. In a word, the TCN-BiGRU model has greater accuracy than the single network models.

### 3.3 Evaluation of inversion results

To select the most suitable model for inversion, it was necessary to conduct a more precise numerical evaluation of the above inversion networks. Two metrics that are often used to analyze the accuracy of a regression model are the coefficient of determination ( $R^2$ ) and the root MSE (RMSE), which are expressed as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - f_{\theta}(x_i))^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (11)$$



**Fig. 7.** AI profiles predicted by the TCN, BiGRU, and combined models. (a) Predicted profile (TCN model), (b) error profile (TCN model), (c) predicted profile (BiGRU model), (d) error profile (BiGRU model), (e) predicted profile (TCN-BiGRU model) and (f) error profile (TCN-BiGRU model).

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - f_{\theta}(x_i))^2} \quad (12)$$

where  $\bar{y}$  is the mean value of the entire predicted AI,  $N$  is the number of the entire seismic traces.  $R^2$  is a metric used to evaluate the goodness of fit of regression models. When  $R^2$  is closer to 1, the prediction accuracy is higher. RMSE describes

**Table 2.** Quantitative evaluation of the TCN, BiGRU, and combined models.

Model	Training time	Epochs	RMSE	$R^2$
TCN	10 min 38 s	1,000	876.46	0.87
BiGRU	5 min 36 s	100	877.60	0.87
TCN-BiGRU	20 min 30 s	220	477.58	0.96

the error between the real and predicted values (Chen et al., 2021b; Zhang et al., 2022), which represents the stability of the network model.

$R^2$  and RMSE of the TCN, BiGRU, and TCN-BiGRU models were calculated by using the test set. Table. 2 lists the  $R^2$  and RMSE of the various models. The RMSE and  $R^2$  of the TCN-BiGRU model introduced in the study were 477.58 and 0.96, respectively. Compared with the single network models, the TCN-BiGRU model had notably higher prediction accuracy and greater stability. And there was only a small increase in training time required.

## 4. Mining area application

### 4.1 Actual data inversion based on TCN-BiGRU model

This study used the Xinjing Mining Area in the city of Yangquan in northern China's Shanxi Province as the subject area. Coal seams No. 3, 8, and 15 in the study area were developed, and one dataset was prepared for model training and testing. The connected well seismic dataset (Fig. 8) comprises well logging data from eight wells on a measuring line: 3-178, 3-170, 3-169, 3-168, 3-167, 3-155, 3-148, and 3-137.

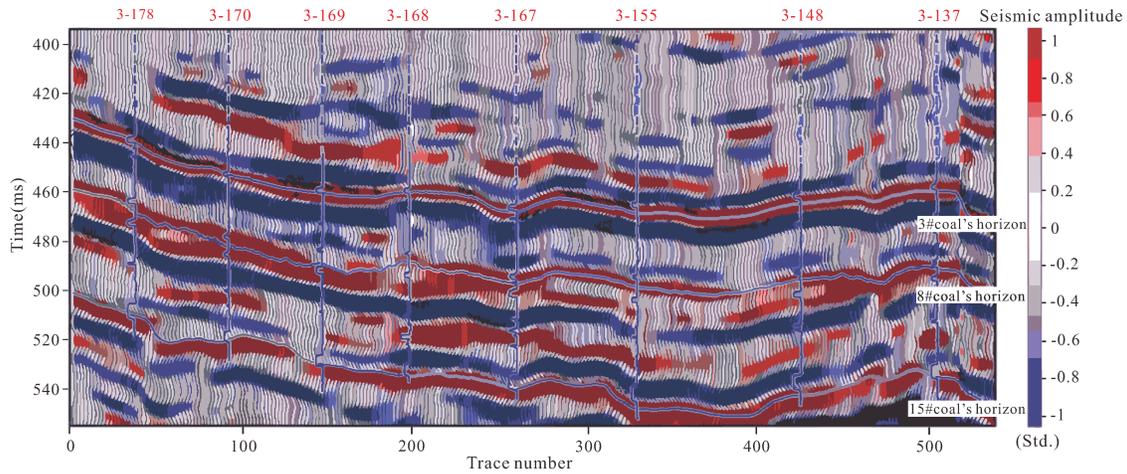
First, the data was preprocessed.

(1) Seismic data and well logging data were normalized according to Eqs. (8) and (9).

(2) Well logging data and seismic data were calibrated so that the well logging data sequence in the depth domain and the seismic data sequence in the time domain were linked according to the corresponding target layers.

(3) To use more data while ensuring the same length, 700 sample points from each well logging AI were ultimately selected as the labels. And the near-well seismic traces were selected as the features. The array shape of the labels or features was (8, 1, 700). Then dividing 70% of the labels and features into the training set and 30% into the validation set.

The training set and validation set were input into the TCN-BiGRU model and the parameters were optimized by minimizing the loss function through multiple epochs. The Adam algorithm was used as the neural network optimization



**Fig. 8.** Xinjing Mining Area connected well seismic dataset.

algorithm and MSE as the loss function. The number of training epoch was 500, the learning rate was 0.005, the batch size was 10, and the final training time was 66 s. The training set and validation set converged to 0.032 and 0.025, respectively.

The actual seismic data was then input into the trained TCN-BiGRU model to obtain the predicted AI inversion result (Fig. 9).

As Fig. 9 shows, the main coal seams, lithological forms, and lithological interfaces can be observed in the profile, but they are inconsistent with well logging verification. The TCN-BiGRU model failed to obtain satisfactory inversion result when applied to actual data, but the training dataset showed a good fit, with the training loss converging to 0.03. This does not mean that the TCN-BiGRU model developed in this study does not work. Our analysis indicates that the result was caused by the following: (1) Using data from only eight wells for training made it difficult for the TCN-BiGRU model to learn the complex relationship between the entire seismic data and AI; (2) the lack of horizontal constraints in the training data led to poor continuity of the coal seams; (3) the frequency band of the actual seismic data was generally lower than that of the well logging AI, causing low accuracy.

#### 4.2 Actual data inversion using the priori constraint-based TCN-BiGRU model

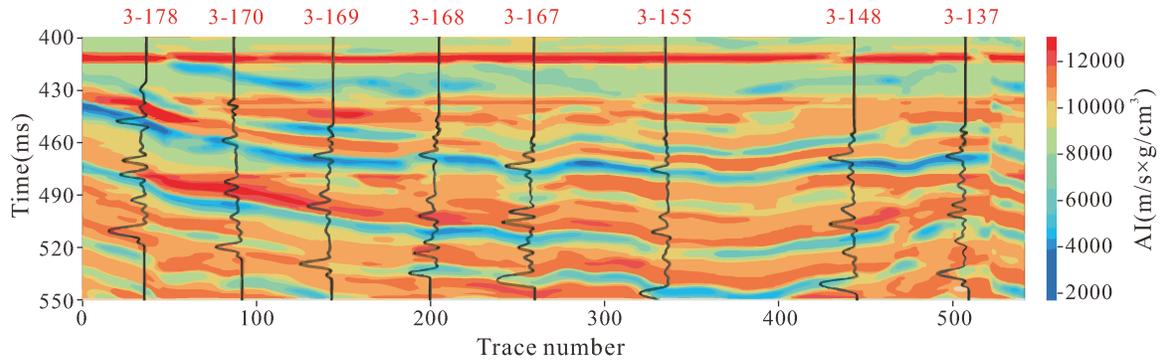
Data-driven intelligent approaches exhibit remarkable flexibility, computational efficiency, and accuracy when dealing with complex multi-scale problems (Xie et al., 2023). Inspired by the data-driven methods with the priori constraint (Yuan et al., 2019; Chen et al., 2021a), the inversion result was improved by inputting the initial geological model (the priori constraint) into the network model.

The initial geological model (Fig. 10(a)) was constructed from seismic data, horizon, and well logging data. The actual near-well seismic traces and the near-well logging AI of the initial geological model were used as the features and then input into the TCN-BiGRU model. The array shape of the features was (8, 2, 700). After optimizing the network

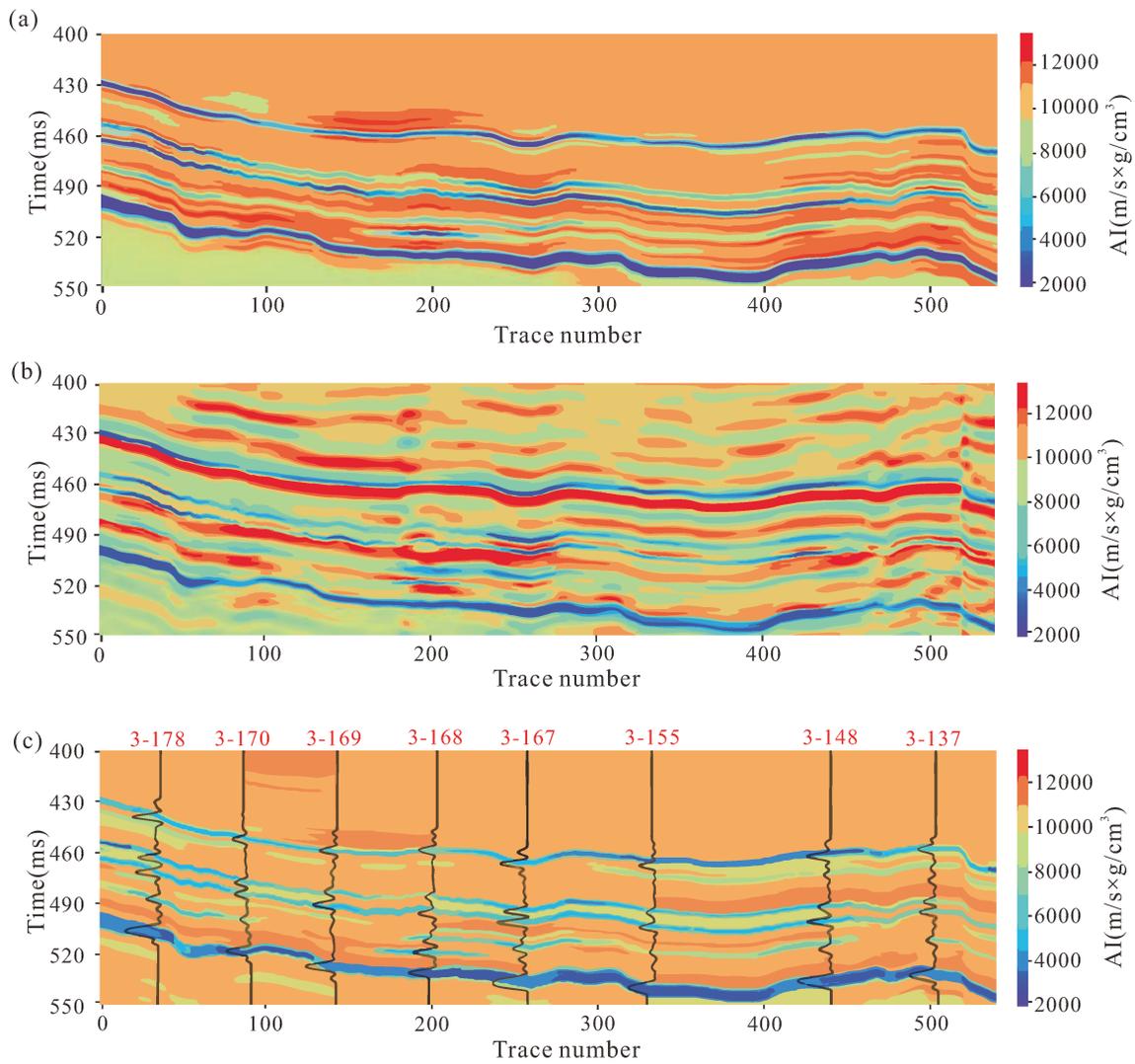
parameters by minimizing the loss function through multiple epochs, The Adam algorithm was used as the neural network optimization algorithm and MSE as the loss function. The number of training rounds was 500, the learning rate was 0.005, the batch size was 10, and the final training time was 58 s. The training set and validation set converged to 0.096 and 0.323, respectively. The actual seismic data was then input into the TCN-BiGRU model to obtain the predicted AI inversion result (Fig. 10(c)).

To further test the efficacy of the priori constraint-based TCN-BiGRU, traditional model-based AI inversion was performed on the mining area using seismic data, well logging, and horizon data. The rick wavelet with a main frequency of 50 Hz was used, and horizons 3, 8, and 15 were used as constraints. The hard constraint was set to 40% which means that the difference between the inversion result and the initial geological model does not exceed 40%. Kriging interpolation technology was used to construct the initial model with a low pass filter (300-350 Hz). The traditional model-based AI inversion result is shown in Fig. 10(b).

As Fig. 10(c) shows, the initial geological model provided high-frequency feature information and low-frequency trend constraints for the priori constraint-based TCN-BiGRU model, therefore, the accuracy of the inversion result is quite high. The near-well logging information of the initial geological model enables the inversion to better depict thin coal seams, and the horizon information improves coal seam identification. Coal seams No. 3, 8, and 15 were accurately identified, and they achieved a good fit with the well logging verification. Compared with the traditional model-based AI inversion result (Fig. 10(b)), the main coal seam fluctuations are consistent, with thickness variation observable in both. However, more information on other layers can be observable based on the inversion result of the priori constraint-based TCN-BiGRU model. For example, a layer of stable sandstone under coal seam No. 3 is observable. The AI inversion result of the priori constraint-based TCN-BiGRU model clearly shows the distribution, thickness variation, and formation fluctuation of the coal seams.



**Fig. 9.** AI prediction result of the TCN-BiGRU model.



**Fig. 10.** Initial geological model and inversion results. (a) Initial geological model, (b) model-based AI inversion result and (c) AI inversion result of the priori constraint-based TCN-BiGRU model.

## 5. Conclusions

This study used DL technology for AI inversion, tested and verified the Marmousi seismic model data and actual seismic data, utilized non-linear interpretation and feature extraction capability of deep neural networks, and improved the features to obtain highly accurate AI inversion results. The following conclusions were drawn from our research:

- 1) Our DL inversion network model combined the low-frequency feature extraction capability of the BiGRU network and the high-frequency feature extraction capability of the TCN network, enabling it to successfully learn the mapping relationship between the seismic data and AI.
- 2) The Marmousi seismic model was used as the experimental data and applied to the TCN-BiGRU inversion network model. The obtained AI inversion results clearly identify changes in lithological interfaces, strata structures, and thin layers, which are more accurate than that obtained using the single neural network model.
- 3) Actual data from the Xinjing Mining Area in Shanxi Province was used as experimental data, and the near-well logging AI of the initial geological model were used as the priori constraint, which was input into the TCN-BiGRU inversion network model to constrain the training process. This initial geological model provided information on thin layers and horizon information, improved the accuracy and horizontal continuity of inversion, provided more precise depictions of thin coal seams, and accurately identified coal seams No. 3, 8, and 15. These findings were consistent with well logging verification.

Although the priori constraint-based TCN-BiGRU model has achieved good results, the number of labels still restricts the application of this method in different mining areas. The future direction of the research is to explore how to expand the label dataset and further improve the network structure to obtain more accurate AI inversion results in different mining areas.

## Acknowledgements

This research was jointly supported by the Fundamental Research Funds for the Central Universities (No. 2022JC-CXMT01) and the Open Fund of State Key Laboratory of Coal Resources and Safe Mining (No. SKLCRSM22DC02).

## References

- Abdi, H., Williams, L. J. Normalizing data, in *Encyclopedia of Research Design*, edited by N. J. Salkind, SAGE Publications, State of California, 2010.
- Adler, A., Araya-Polo, M., Poggio, T. Deep learning for seismic inverse problems: Toward the acceleration of geophysical analysis workflows. *IEEE Signal Processing Magazine*, 2021, 38(2): 89-119.
- Alfarraj, M., AlRegib, G. Semi-supervised learning for acoustic impedance inversion. Paper Presented at the SEG International Exposition and Annual Meeting, San Antonio, USA, 15-20 September, 2019.
- Biswas, R., Sen, M. K., Das, V., et al. Prestack and poststack inversion using a physics-guided convolutional neural network. *Interpretation*, 2019, 7(3): SE161-SE174.
- Chen, H., Gao, J., Zhang, W., et al. Seismic acoustic impedance inversion via optimization-inspired semisupervised deep learning. *IEEE Transactions on Geoscience and Remote Sensing*, 2021a, 60: 111.
- Chen, T., Zhang, Z., Zhou, X., et al. Prediction model of coalbed methane content based on well logging parameter optimization. *Coal Geology & Exploration*, 2021b, 49(3): 227-235. (in Chinese)
- Das, V., Pollack, A., Wollner, U., et al. Convolutional neural network for seismic impedance inversion. *Geophysics*, 2019, 84(6): R869-R880.
- Jais, I. K. M., Ismail, A. R., Nisa, S. Q. Adam optimization algorithm for wide and deep neural network. *Knowledge Engineering and Data Science*, 2019, 2(1): 41-46.
- Janiesch, C., Zschech, P., Heinrich, K. Machine learning and deep learning. *Electronic Markets*, 2021, 31(3): 685-695.
- Li, Y. E., O'Malley, D., Beroza, G., et al. Machine learning developments and applications in solid-earth geosciences: Fad or future? *Journal of Geophysical Research: Solid Earth*, 2023, 128(1): e2022JB026310.
- Marques, C. R., dos Santos, V. G., Lunelli, R., et al. Analysis of deep learning neural networks for seismic impedance inversion: A benchmark study. *Energies*, 2022, 15(20): 7452.
- Martin, G. S., Wiley, R., Marfurt, K. J. Marmousi2: An elastic upgrade for Marmousi. *The Leading Edge*, 2006, 25(2): 156-166.
- Mustafa, A., Alfarraj, M., AlRegib, G. Estimation of acoustic impedance from seismic data using temporal convolutional network. Paper Presented at the SEG International Exposition and Annual Meeting, San Antonio, USA, 15-20 September, 2019.
- Mustafa, A., Alfarraj, M., AlRegib, G. Joint learning for spatial context-based seismic inversion of multiple data sets for improved generalizability and robustness. *Geophysics*, 2021, 86(4): O37-O48.
- Shen, Y., Ma, Y., Deng, S., et al. An ensemble model based on deep learning and data preprocessing for short-term electrical load forecasting. *Sustainability*, 2021, 13(4): 1694.
- Smith, R., Nivlet, P., Alfayez, H., et al. Robust deep learning-based seismic inversion workflow using temporal convolutional networks. *Interpretation*, 2022, 10(2): SC41-SC55.
- Song, L., Yin, X., Zong, Z. Deep learning seismic impedance inversion based on prior constraints. *Oil Geophysical Prospecting*, 2021, 56(4): 716-727. (in Chinese)
- Sun, Z. H. Binary outer product expansion of convolutional kernels. *IEEE Journal of Selected Topics in Signal Processing*, 2020, 14(4): 871-883.
- Tao, L., Ren, H., Gu, Z. Acoustic impedance inversion from seismic imaging profiles using self attention U-Net. *Remote Sensing*, 2023, 15(4): 891.
- Wang, D., Chen, G. Seismic wave impedance inversion based on temporal convolutional network. *Earth Science*, 2022, 47(4): 1492-1506. (in Chinese)

- Wang, H., Yan, J., Fu, G., et al. Current status and application prospect of deep learning in geophysics. *Progress in Geophysics*, 2020, 35(2): 642-655. (in Chinese)
- Wang, B., Yin, X., Zhang, F. Elastic impedance inversion and its application. *Progress in Geophysics*, 2005, 20(1): 89-92. (in Chinese)
- Wu, B., Meng, D., Wang, L., et al. Seismic impedance inversion using fully convolutional residual network and transfer learning. *IEEE Geoscience and Remote Sensing Letters*, 2020, 17(12): 2140-2144.
- Wu, B., Meng, D., Zhao, H. Semi-supervised learning for seismic impedance inversion using generative adversarial networks. *Remote Sensing*, 2021a, 13(5): 909.
- Wu, X., Yan, S., Bi, Z., et al. Deep learning for multidimensional seismic impedance inversion. *Geophysics*, 2021b, 86(5): R735-R745.
- Xie, C., Du, S., Wang, J., et al. Intelligent modeling with physics-informed machine learning for petroleum engineering problems. *Advances in Geo-Energy Research*, 2023, 8(2): 71-75.
- Yang, W., Xia, K., Fan, S. Oil logging reservoir recognition based on TCN and SA-BiLSTM deep learning method. *Engineering Applications of Artificial Intelligence*, 2023, 121: 105950.
- Yoo, J., Kim, D., Choi, J., et al. Impedance inversion based on domain adaptation technique with reconstruction. *IEEE Geoscience and Remote Sensing Letters*, 2022, 19: 1-5.
- Yoon, D., Yeeh, Z., Byun, J. Seismic data reconstruction using deep bidirectional long short-term memory with skip connections. *IEEE Geoscience and Remote Sensing Letters*, 2020, 18(7): 1298-1302.
- Yuan, S., Wang, S., Luo, Y., et al. Impedance inversion by using the low-frequency full-waveform inversion result as an a priori model. *Geophysics*, 2019, 84(2): R149-R164.
- Zhang, T., Sun, L., Yang, Y., et al. A method of collapse column recognition based on improved residual network. *Coal Geology & Exploration*, 2023, 51(5): 171-179. (in Chinese)
- Zhang, Y., Wen, X., Wu, H., et al. Seismic acoustic impedance inversion using  $L_p$  quasi-norm sparse constraint and alternating direction multiplier algorithm. *Geophysical Prospecting for Petroleum*, 2022, 61(5): 856-864. (in Chinese)
- Zhao, D., Yang, N., Xiong, J., et al. Model-constrained and data-driven double-supervision acoustic impedance inversion. *Petroleum Science*, 2023, in press, <https://doi.org/10.1016/j.petsci.2023.03.019>.