APPLIED GEOPHYSICS, Vol.21, No.2 (June 2024), P. 1−14, 11 Figures.
DOI:10.1007/s11770-024-1082-y

Multitask Weighted Adaptive Prestack Seismic Inversion

Cheng Jian-yong[1,2], Yuan San-yi[1], Sun Ao-xue[1], Luo Chun-mei[1]*, Liu Hao-jie[3], and Wang Shang-xu[1]

Abstract: Traditional deep learning methods pursue complex and single network architectures without considering the petrophysical relationship between different elastic parameters. The mathematical and statistical significance of the inversion results may lead to model overfitting, especially when there are a limited number of well logs in a working area. Multitask learning provides an effective approach to addressing this issue. Simultaneously, learning multiple related tasks can improve a model's generalization ability to a certain extent, thereby enhancing the performance of related tasks with an equal amount of labeled data. In this study, we propose an end-to-end multitask deep learning model that integrates a fully convolutional network and bidirectional gated recurrent unit for intelligent prestack inversion of "seismic data to elastic parameters." The use of a Bayesian homoscedastic uncertainty-based loss function enables adaptive learning of the weight coefficients for different elastic parameter inversion tasks, thereby reducing uncertainty during the inversion process. The proposed method combines the local feature perception of convolutional neural networks with the long-term memory of bidirectional gated recurrent networks. It maintains the rock physics constraint relationships among different elastic parameters during the inversion process, demonstrating a high level of prediction accuracy. Numerical simulations and processing results of real seismic data validate the effectiveness and practicality of the proposed method.
Keywords: Prestack seismic inversion; Multitask learning; Fully convolutional neural network; Bidirectional gated recurrent neural network

Introduction

The prestack seismic inversion technique retains the characteristics of seismic reflection amplitudes that vary with incident angle or offset distance [1]. Through prestack inversion, various geological parameters, such as P- and S-wave velocities, density, Poisson's ratio, and fluid factor, can be determined for detailed reservoir studies. Joint analysis of these parameters can reduce ambiguity in seismic interpretation and effectively predict underground lithology, physical properties, and hydrocarbon content. The initial application of the prestack seismic inversion technique focused on qualitatively describing reservoir types. To obtain elastic parameters that directly reflect subsurface lithological information and serve the needs of oil and gas exploration, the technique has evolved into a series of quantitative prestack inversion methods for describing different types of reservoir parameters.

Traditional prestack seismic inversion methods can be broadly categorized into two types. The

Manuscript received by the Editor December 18, 2023; revised manuscript received April 3, 2024, 2024
This work was supported by National Key R & D Program of China(2018YFA0702501), National Natural Science Foundation of China (41974140), Science and Technology Management Department, China National Petroleum Corporation (2022DQ0604-01), China National Petroleum Corporation - China University of Petroleum (Beijing) Strategy.
1. State Key Laboratory of Petroleum Resources and Engineering, China University of Petroleum (Beijing), Beijing, China. 102249
2. PetroChina Changqing Oilfield Company No.8 Oil Production Company, Xian 718600, China
3. Shengli Geophysical Research Institute of Sinopec, Dongying 257000, China.
* Corresponding author: Luo Chun-mei,(Email: lcm@cup.edu.cn).

first type is based on the Zoeppritz equation and its approximations to calculate the reflection coefficients and predict hydrocarbon reservoirs by studying the influence of elastic parameter variations on seismic amplitudes. However, different amplitude versus offset approximation formulas have specific assumptions and applicable conditions. Deviations may occur when these conditions are not fully satisfied. The second type is the prestack-full waveform inversion (FWI) method based on the wave equation. This approach uses complete wavefield information to solve the inverse problem, resulting in higher inversion accuracy than the former approach. Nevertheless, they are prone to becoming trapped in local optima, and the substantial computational and storage requirements limit their practical application. Therefore, FWI is currently primarily used for low-frequency velocity modeling in seismic imaging as a background field to achieve more accurate imaging results. Yuan et al. [12] effectively applied FWI to reservoir characterization at a low cost.

With the development of deep learning (DL) and big data, an increasing number of DL methods have been applied in the field of geophysical exploration, including first arrival picking [2][3], velocity modeling [4], FWI [5], seismic facies recognition [6][7], fault identification [8] [9], wave impedance inversion [10][11][12], and reservoir prediction [13][14]. Some researchers have successfully applied DL techniques to prestack inversion, yielding promising outcomes. Biswas et al. [15] used a physics-guided convolutional neural network (CNN) for prestack inversion and accurately predicted elastic parameters through unsupervised learning. Du et al. [16] constructed an anisotropic inversion model for prestack seismic inversion using residual networks to extract features from input data and obtain P-wave impedance, S-wave impedance, and rock physics parameters. Wang et al. used band-limited inversion data, seismic velocity data after correctedcorrection, and other information to generate full-band inversion data, including P-wave impedance and P-to-S -velocity ratio, which serveserved as inputs for subsequent inversion of low-frequency models. ExperimentThe experimental results indicateindicated that the low-frequency models based on deep learningDL effectively reflectreflected changes in sedimentary facies and improveimproved the accuracy of inversion results.Sun et al. [17] used reversible neural networks to learn the amplitude versus offset forward modeling process and inverted low-to-mid-frequency velocities and densities from input data. This method randomly generates easily obtainable datasets without requiring precise training samples and initial models, thereby reducing the dependence on initial models to a certain extent. This method demonstrated good inversion results for low-to-mid-frequency velocities and densities using synthetic and field data. Cao et al. [18] employed sequential Gaussian simulation and elastic distortion algorithms to generate sufficient and diverse datasets. They used a combination of U-Net and fully connected neural networks to predict elastic parameters. Sparse reflection coefficients were used as physical constraints to enhance network prediction accuracy. The inversion results were superior to those of the traditional DL methods.

The aforementioned DL methods demonstrate excellent performance in prestack seismic inversion and perform well on field data. However, they typically require a substantial amount of labeled training data to enhance model prediction accuracy. Owing to the high drilling costs, the limited number of wells constrains model generalization ability. Moreover, they do not specifically design appropriate network structures for different elastic parameter inversion tasks and usually adopt a single-task and single-output format for different inversion tasks. There are empirical physical formulas linking different parameters that show certain correlations. Therefore, the inversion processes of multiple elastic parameters can be regarded as multitask learning (MTL) process. MTL, an important part of machine learning, is inspired by human learning behavior. In the process of learning new skills, it attempts to apply the knowledge learned through one task to help learn another related skill. The main idea is to simultaneously learn multiple related tasks and share specific domain knowledge inherent in different tasks by sharing some parameters. This approach enhances a model's generalization ability across multiple tasks and improves its overall performance. Traditional transfer learning refers to transferring knowledge learned in one domain to another, with a sequence in the learning process. In MTL, information among different types of tasks can be shared, and knowledge can be transferred among different tasks. Therefore, MTL is also called parallel transfer learning. Compared with traditional single-task learning (STL), MTL has several advantages. 1) MTL can aggregate data from multiple tasks and use their correlation to achieve augmentation; different tasks can explore domain knowledge and reuse information, thereby effectively reducing the demand for data and

alleviating the adverse effects of data sparsity. 2) By exploring and synthesizing domain knowledge and feature information from different tasks, MTL can provide more robust embedded representations for different tasks, effectively reducing the risk of model overfitting and improving model performance in each task. 3) MTL can simultaneously learn multiple models, thereby reducing training time and storage, updating, and maintenance costs in the subsequent learning; also, MTL can balance the noise of different task data and improve a model's generalization ability.

In this study, we propose a deep MTL method that combines a fully convolutional network (FCN) and bidirectional gated unit (Bi-GRU) to achieve end-to-end intelligent inversion from seismic data to elastic parameters. The MTL is designed to model both mid-to-long-term dependencies and nonlinear relationships present in seismic data. Furthermore, the MTL loss function is designed based on homoscedastic uncertainty to achieve adaptive learning of weights for different inversion tasks. Finally, the proposed method is applied to field data from eastern China to verify its effectiveness.

Method and Model

Multitask FCN–Bi-GRU Deep Neural Network Model

Information Sharing Layer

FCN is a variant of CNN that incorporates convolutional layers and can selectively retain pooling layers based on different tasks. Its main characteristic is the use of convolutional layers instead of fully connected layers in traditional CNNs. By preserving the spatial position information of the original input data, FCN uses deconvolutional layers to increase the feature vector size. The upsampled feature vectors are then used for classification or regression predictions for each feature point. In addition, FCN can incorporate skip connections from different depth layers to ensure robustness and accuracy.

The information sharing layer uses convolutional kernels of different sizes to extract shared high-dimensional temporal features from the input seismic data. The convolutional formula is as follows:

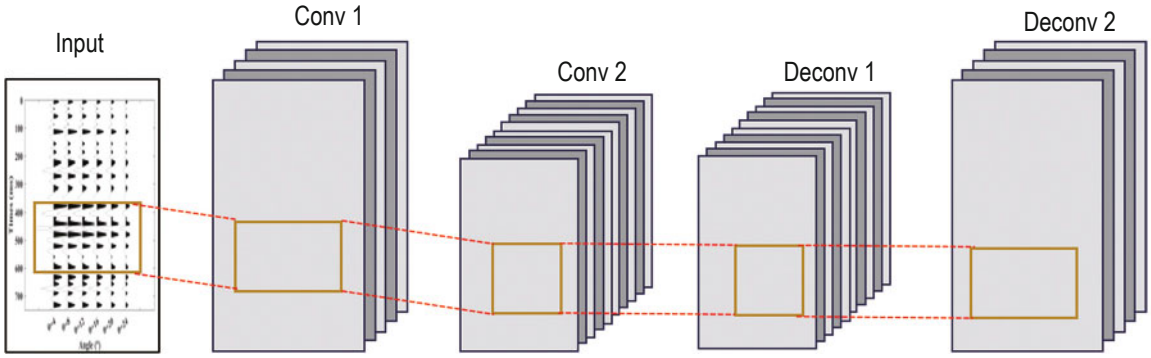

Figure 1. FCN architecture.

$$x_j^l = f\left(\sum_{i \in M_j} x_i^{l-1} * W_{ij}^l + b_j^l\right), \quad (1)$$

where x_j^l denotes the feature output matrix, f represents the activation function, W_{ij}^l and b_j^l denotes the convolution kernel weights and biases, respectively, and * represents the convolution operation.

The input data are the seismic angle gathers of CDP (Common Depth Point) points, which include two dimensions: time and angle. Two sets of different convolutional blocks were used to extract local features of different scales from the seismic data. Because the input data size changes during feature extraction, deconvolution layers are employed to up-sample the features so that they have the same sampling rate as the label data.

Special Task Layer

Bi-GRU is an extended form of a recurrent neural network comprising two stacked GRUs in both directions. Its structure is shown in Figure 2. When performing sequence modeling, it uses hidden state vectors to consider information from both forward and backward directions, capturing long-distance dependencies in the sequence data through contextual computation to determine the final output. The purpose

of this special task layer is to establish a nonlinear mapping relationship between shared high-dimensional temporal features and real-well labels comprising a series of Bi-GRUs.

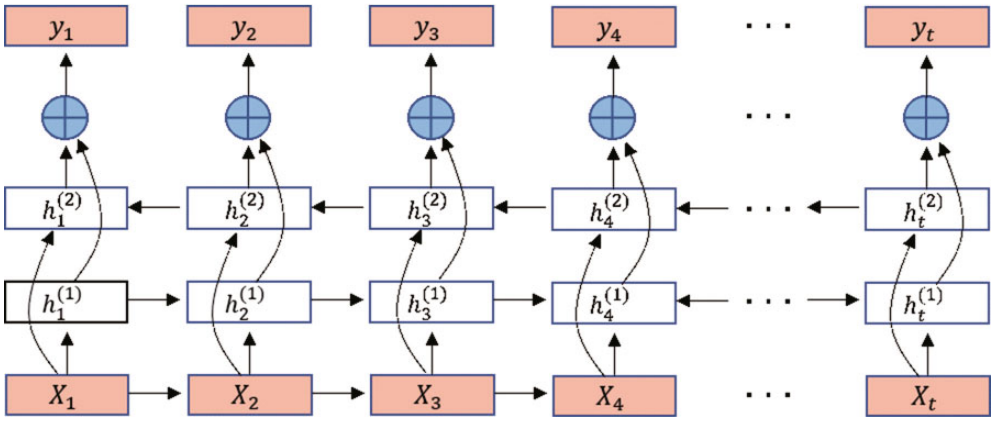

Figure 2. Bi-GRU architecture.

In Figure 2, for any time step t, there is a forward propagation layer $\vec{h}_t$ and a backward hidden state layer $\overleftarrow{h}_t$. The forward and backward hidden state updating processes are as follows:

$$\vec{\boldsymbol{h}}_t^{(1)} = \boldsymbol{f}\left(\boldsymbol{W}^{(1)}\boldsymbol{x}_t + \boldsymbol{U}^{(1)}\boldsymbol{h}_{t-1}^{(1)} + \boldsymbol{b}^{(1)}\right), \quad (2)$$

$$\overleftarrow{\boldsymbol{h}}_t^{(2)} = \boldsymbol{f}\left(\boldsymbol{W}^{(2)}\boldsymbol{x}_t + \boldsymbol{U}^{(2)}\boldsymbol{h}_{t+1}^{(2)} + \boldsymbol{b}^{(2)}\right), \quad (3)$$

where W, U, and b are model parameters. At time t, the forward and backward recurrent layers update their hidden states to obtain the results for the forward propagation layer $\vec{h}_t$ and the backward hidden state layer $\overleftarrow{h}_t$. These results were concatenated into a new vector that contained information from the seismic data in both the forward and backward directions. This vector was then used as the hidden state h_t for the next input layer. For models with multiple hidden layers, this vector can be used as input for subsequent classification or regression tasks.

Multitask FCN–Bi-GRU Model

When using DL models for seismic inversion, it is often difficult to fully extract effective and potential complex relationships between seismic data and elastic parameters using a single network. Considering the correlation between different elastic parameters, we propose a multitask FCN–Bi-GRU model that combines convolutional and recurrent networks. The basic network structure is shown in Figure 3. The model uses a parameter hard-sharing mechanism to realize MTL, and the network input data are normalized prestack seismic data. First, the shared local morphological features of the input seismic data are extracted by the FCN in the shared

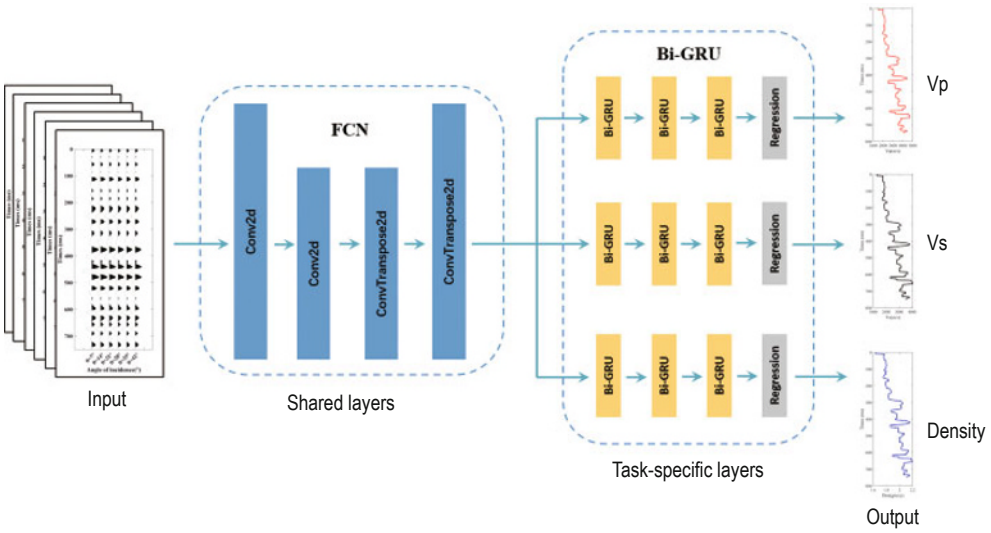

Figure 3. Basic FCN–Bi-GRU structure.

layer, and temporal feature vectors are constructed. The results are then fed into subnetworks of different task layers for capturing the trend and information of high-dimensional time-series features with depth modeling the internal feature dynamics, which facilitates the learning of the dynamic change law inside elastic parameters. Finally, the fully connected layer is used to reduce the number of channels in the output vectors, which is then used for the subsequent label loss computation and elastic parameter model prediction. The network model considers the seismic local morphological feature related to the stratigraphic sedimentary law and can establish a nonlinear relationship between this feature and the labels of different elastic parameters, thereby enabling complete learning of the spatiotemporal features of seismic data.

Design of Loss Function Optimization

Uncertainty is an important research direction in DL and can be conceptualized as epistemic or aleatoric uncertainty. The former refers to the cognitive bias caused by insufficient data. When there is insufficient data and the training data distribution is not representative of the overall data distribution, there is bias in model training, and this uncertainty can be improved by increasing the training data. Aleatoric uncertainty refers to cognitive bias caused by the task or data. This is characterized by the fact that it does not improve with increasing data. Aleatoric uncertainty can be further divided into 1) heteroscedastic uncertainty, which refers to the model prediction bias caused by mislabeling, and 2) homoscedastic uncertainty [19], which refers to the bias that does not depend on the data but is caused by the same data for different tasks. One of the ideas of MTL to adjust the weight coefficients for different tasks is based on the uncertainty weighting method, which makes the overall multitask model training smooth and effective by assigning relatively small weights to difficult tasks based on the homoscedastic uncertainty in the contingent uncertainty for modeling.

The loss function of a conventional MTL network model highly depends on the weight coefficients of individual task losses, which are usually manually adjusted. It is defined as follows:

$$Loss = \sum_{i=1}^{M} \omega_i L_i + \frac{\lambda}{2} \| w \|^2, \tag{4}$$

where M denotes the number of tasks, ω_i denotes the hyperparameter in the loss function representing the weight of different tasks that need to be manually adjusted, and λ denotes the regularization parameter for the model's weight parameters, which is used to prevent overfitting. Because MTL model performance is highly influenced by the weight of each task loss function, one challenge in MTL is balancing the weight coefficients among different tasks. Conventional approaches simply add up the losses of tasks or set a unified loss weight, and sometimes further manual adjustment may be performed. However, in MTL, different tasks often have different optimization goals and data distributions. Therefore, simply adding task loss functions or manually adjusting weight coefficients may not yield optimal results in MTL. Considering the intrinsic rock physics relationships among different elastic parameters, we designed an MTL model loss function based on homoscedastic uncertainty to balance multiple elastic parameter inversion tasks and avoid manual adjustment of weight coefficients. Assuming the model input data x, weight parameter w, noise σ^2 in the output values, and network model output $f^w(x)$. The probability estimation of regression tasks can be represented as follows:

$$p\left(y|f^w(x)\right) = N\left(f^w(x), \sigma^2\right), \tag{5}$$

The maximum likelihood estimation of Equation 5 is as follows:

$$logp\left(y|f^w(x)\right) \propto -\frac{1}{2\sigma^2} \| y - f^w(x) \|^2 - log\sigma, \tag{6}$$

We define the MTL likelihood function as follows:

$$p\left(y_1, \ldots, y_k | f^w(x)\right) = p\left(y_1|f^w(x)\right) \ldots p\left(y_k|f^w(x)\right), \tag{7}$$

where y_1, y_2, y_3 represent the outputs of three tasks (in this study, P-velocity, S-velocity, and density). Then, we obtain

$$\begin{aligned} p\left(y_1, y_2, y_3 | f^w(x)\right) &= p\left(y_1|f^w(x)\right) \cdot p\left(y_2|f^w(x)\right) \cdot p\left(y_3|f^w(x)\right) \\ &= N\left(y_1; f^w(x), \sigma_1^2\right) \cdot N\left(y_2; f^w(x), \sigma_2^2\right) \cdot N\left(y_3; f^w(x), \sigma_3^2\right) \end{aligned}, \tag{8}$$

Maximizing the logarithm of the likelihood function is equivalent to minimizing its negative logarithm, which

results in the following MTL loss function:

$$Loss\left(w,\sigma_1,\sigma_2,\sigma_3\right)=-logp\left(y_1,y_2,y_3|f^w\left(x\right)\right)$$

$$\propto\frac{1}{2\sigma_1^2}\|y_1-f^w\left(x\right)\|^2+\frac{1}{2\sigma_2^2}\|y_2-f^w\left(x\right)\|^2+\frac{1}{2\sigma_3^2}\|y_3-f^w\left(x\right)\|^2+\log\sigma_1\sigma_2\sigma_3, \quad (9)$$

$$=\frac{1}{2\sigma_1^2}L_1(w)+\frac{1}{2\sigma_2^2}L_2(w)+\frac{1}{2\sigma_3^2}L_3(w)+\log\sigma_1\sigma_2\sigma_3$$

By minimizing this loss function, the weight coefficients can be automatically adjusted. σ_1^2, σ_2^2,and σ_3^2 represent the noise in the outputs of the three tasks, and larger noise leads to smaller weight coefficients for the corresponding tasks. The term log σ_1 σ_2 σ_3 is a regularizer for the output noise terms to prevent excessively small weight coefficients. By minimizing this loss function, the manual adjustment process of task weight coefficients is eliminated, and the optimal solution is obtained through the adaptive adjustment of weight coefficients during network training.

Evaluation Criteria for Inversion Results

To quantitatively evaluate the subsequent inversion results, two evaluation criteria are introduced in this section. The first is the Pearson correlation coefficient (PCC), which measures the degree of correlation between the true and inversion results. Its value ranges from −1 to 1, and it is defined as follows:

$$PCC\left(y,\hat{y}\right)=\frac{1}{N}\frac{\sum_{i=1}^{N}\left(y_i-\mu_y\right)\left(\hat{y}_i-\mu_{\hat{y}}\right)}{\sqrt{\sum_{i=1}^{N}\left(y_i-\mu_y\right)^2}\sqrt{\sum_{i=1}^{N}\left(\hat{y}_i-\mu_{\hat{y}}\right)^2}}, \quad (10)$$

where y represents the true well log curve, $\hat{y}'$ represents the results obtained from different inversion methods, N denotes the sequence length (the number of time samples for a single trace), and μ_y and $\mu_{\hat{y}}$ denote the means of the true and predicted values, respectively. PCC reflects the correlation between the true and predicted results. A higher correlation indicated that the predicted results were highly similar to the true results in the waveform, indicating better inversion results, whereas a lower correlation indicated poorer inversion results.

Because different elastic parameters have different dimensions and orders of magnitude, the second evaluation criterion uses the mean absolute percentage error (MAPE) to measure prediction accuracy. MAPE reflects the absolute error percentage and confidence level between the inversion and true results, ranging from 0 to 1, and it is defined as follows:

$$MAPE=\frac{100\%}{N}\sum_{t=1}^{N}\left|\frac{\hat{y}_i-y_i}{y_i}\right|, \quad (11)$$

A smaller MAPE indicates that the inversion results are closer to the true results numerically, indicating better inversion results, whereas a larger MAPE suggests significant differences from the true results and poorer inversion performance.

Numerical Examples

Prestack Elastic Parameter Inversion Under Noise-Free Conditions

Figures 4a–c show the P-velocity, S-velocity, and density of the two-dimensional (2D) Marmousi model, respectively. Six incident angles of 4°, 8°, 12°, 16°, 20°, and 24° were selected. The reflection coefficient sequences dependent on angles were calculated using the Aki–Richards equation [20]. The convolution between a zero-phase Ricker wavelet with a frequency of 25 Hz and the angle reflection coefficient sequence generates prestack seismic gather. Figure 4d shows prestack common-angle data for an incident angle of 20°, where the black and red curves represent the training and testing data, respectively.

The following formula is used to normalize the prestack seismic data:

$$\overline{X}=\frac{X-X_\mu}{X_\sigma}, \quad (12)$$

where X_μ and X_σ denote the mean and standard

deviation of the input data, and X denotes normalized data. The FCN–Bi-GRU model was trained using the corresponding seismic data and pseudo-well logs. The iteration period was set to 500, the learning rate was set to 0.005, and an Adam optimizer with adaptive moment estimation was used for gradient descent. The regularization parameter λ for the model weight was set to 1×10^{-4}. The obtained elastic parameter prediction model was applied for testing on the normalized prestack seismic data, and the elastic parameter results were predicted by reverse normalization of the network output. Based on the above process, testing was performed under both STL and MTL conditions. The FCN–Bi-GRU model was used, and the network parameters were the same as those outlined in the previous section. The only difference is that in the STL inversion case, there is only one special task layer that inverts a single elastic parameter (either P-velocity, S-velocity, or density) at a time.

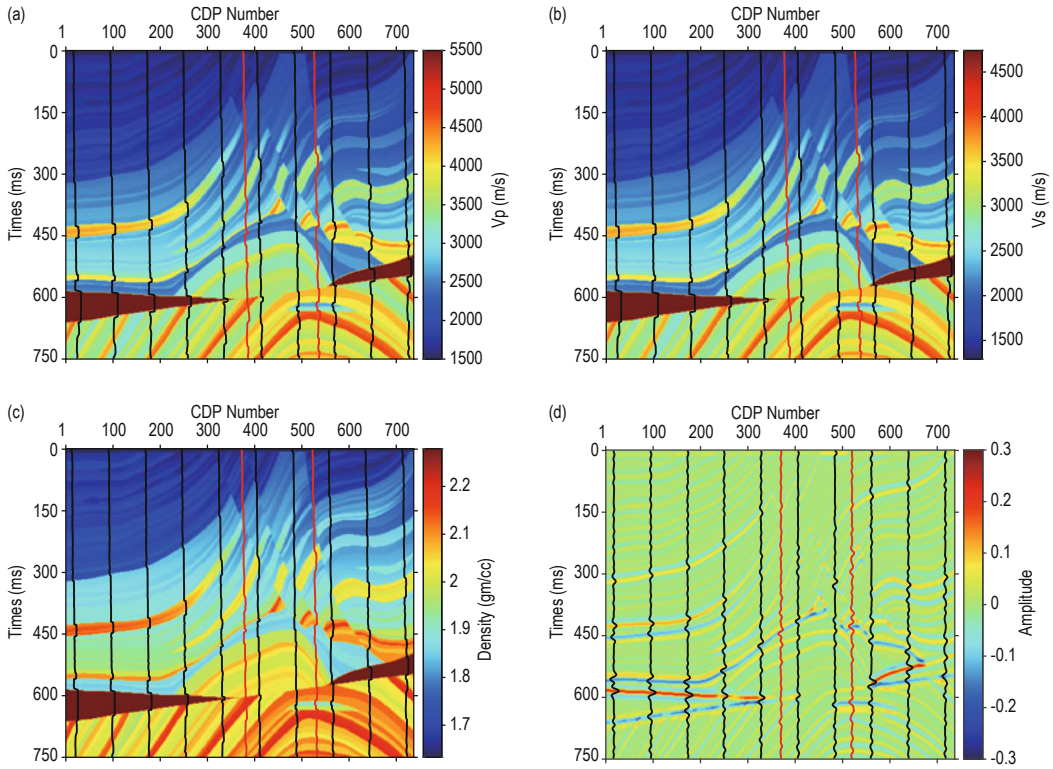

Figure 4. Vp (P-wave velocity), Vs (S-wave velocity) model, density model, and prestack seismic data. (a) P-velocity; (b) S-velocity; (c) Density; (d) Prestack seismic data.

Figure 5 shows the STL and MTL inversion results. From the 2D inversion profiles, it can be preliminarily observed that the STL inversion results in the complex geological structure area from CDP 350–500 having poor profiles, with more vertical artifacts and lower lateral continuity. Compared with the STL inversion, the MTL inversion results have relatively better profiles, with reduced vertical artifacts and strong spatial structural consistency, indicating good correspondence between the underground geological structures and P-velocity, S-velocity, and density. However, there were still some vertical artifacts near the region, with significant vertical changes in the structure around CDP 480. This is attributable to two reasons: 1) different dimensional information exists between the prestack seismic data and elastic parameter model, resulting in weak spatial correspondence, especially in terms of frequency components, where low-frequency components are difficult to learn; 2) owing to the trace-by-trace inversion, the lateral variation in the geological formation was not considered; thus, when there is a large deviation in the elastic parameter result of a single trace, the lateral continuity of the predicted profile deteriorates.

We performed four sets of tests. Test 1 represents STL; to demonstrate the superiority of the proposed loss function optimization, three additional comparative tests were conducted with MTL: Tests 2 and 3 involved manually adjusting the weight coefficients. In Test 2, the

weight coefficients were set to be equal, meaning that each task contributed equally to the loss function. In Test 3, the weight coefficients were manually optimized to determine the optimal values. Test 4 used a loss function designed with homoscedastic uncertainty to adaptively adjust the weight coefficients. The specific parameters are shown in Table 1. The tests used 10 noise-free prestack seismic datasets uniformly extracted as training data, with pseudo-well logs serving as labels. The inversion results of different methods were obtained.

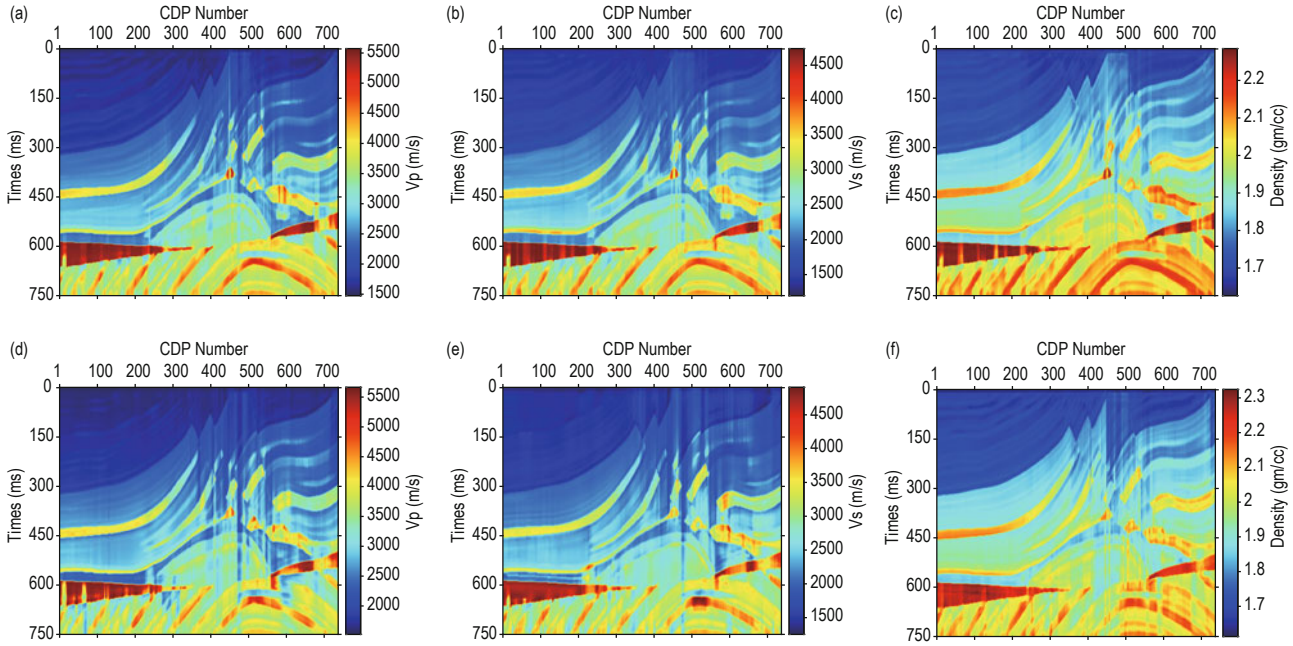

Figure 5. Comparison of STL and MTL inversion results.
(a) Vp (MTL); (b) Vs (MTL); (c)Density (MTL); (d) Vp (STL); (e) Vs (STL); (f) Density (STL).

Table 1 Weight coefficients for different experimental loss functions

Test	Training methods	Weight coefficients
Test 1	STL	——
Test 2	STL and equal weight	0.33 / 0.33 / 0.33
Test 3	MTL and optimal weight	0.21 / 0.35 / 0.44
Test 4	MTL and uncertainty	——

Figure 6 shows the evaluation results of the four sets of weight coefficient experiments. The MTL inversion results have higher correlation coefficients and lower absolute errors than the STL inversion results. Further, among the three sets of comparative experiments for MTL, the uncertainty weight performed better than the

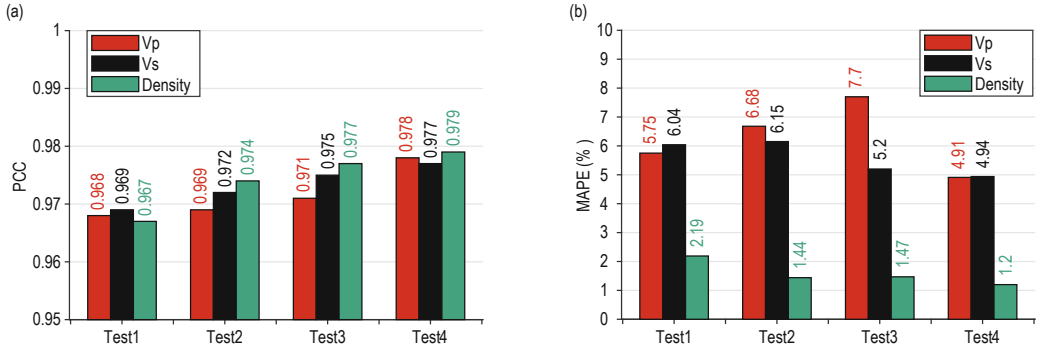

Figure 6. Evaluation metrics for inversion results under different training modalities.
(a) PCC for different experiments; (b) MAPE for different experiments.

optimal weight found through fine-grained grid search, indicating the effectiveness of the proposed loss function optimization. The main reason may be that the best weights found using the grid search method are limited by the search resolution, resulting in certain limitations in terms of weight search accuracy. In addition, optimizing the weights through uncertainty allows for dynamic changes in the weight coefficients of different tasks during network training, thereby improving the network optimization process.

The joint analysis of the elastic parameter inversion results for STL (Test 1) and MTL (Test 4) is shown in Figure 7, where the black curve represents the rock physics cross plot with true elastic parameter labels. The cross plot of the elastic parameter inversion results for MTL presents a relatively narrow distribution and is closer to the true cross plot curve. This indicates that using MTL can achieve knowledge sharing of rock physics for different elastic parameter inversion tasks and maintain the rock physics relationship among different elastic parameters during the inversion process, thereby improving the performance of individual tasks and obtaining elastic parameter inversion results that are more consistent with underground geological structures.

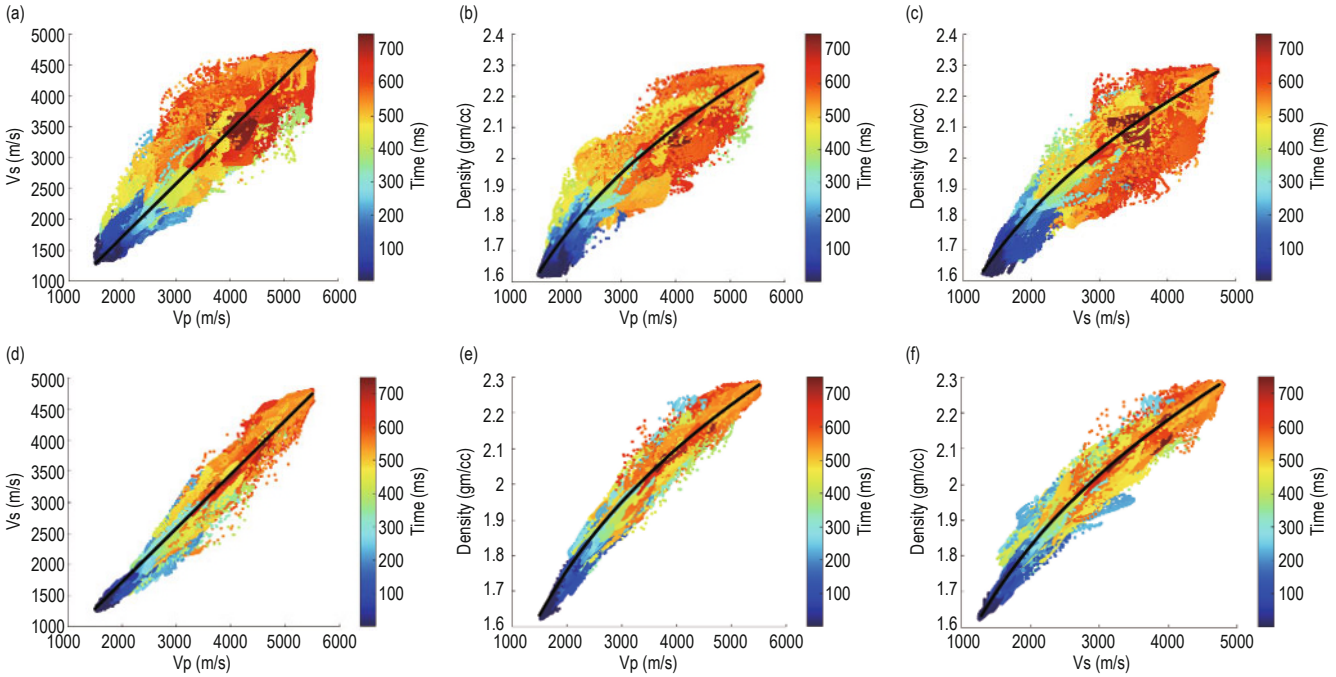

Figure 7. Comparison of STL and MTL inversion elasticity parameter cross plot. (a) Vp–Vs cross plot (STL); (b) Vp–Density cross plot (STL); (c) Vs–Density cross plot (STL); (d) Vp–Vs cross plot (MTL); (e) Vp–Density cross plot (MTL); (f) Vs–Density cross plot (MTL).

Prestack Elastic Parameter Inversion Under Noise

Real seismic data collected in the field were often subject to external interference, resulting in inevitable disturbances. To further illustrate that MTL has better noise resistance than STL, different levels of random noise are added to the original seismic data. The networks are separately trained on different noise levels and then extrapolated to the corresponding noisy data to obtain inversion results at different noise levels. Figure 8 shows the STL and MTL inversion results under a signal-to-noise ratio (SNR) of 5. Tables 2 and 3,

Table 2 Quantitative evaluation of inversion results based on MTL

SNR	Vp-MAPE (%)	Vs-MAPE (%)	Den-MAPE (%)
15	5.19	5.64	1.36
10	5.35	5.98	1.89
5	5.71	5.92	1.35
2	5.74	7.61	1.73

respectively, summarize the quantitative evaluation of the STL and MTL inversion results of seismic data with different SNRs.

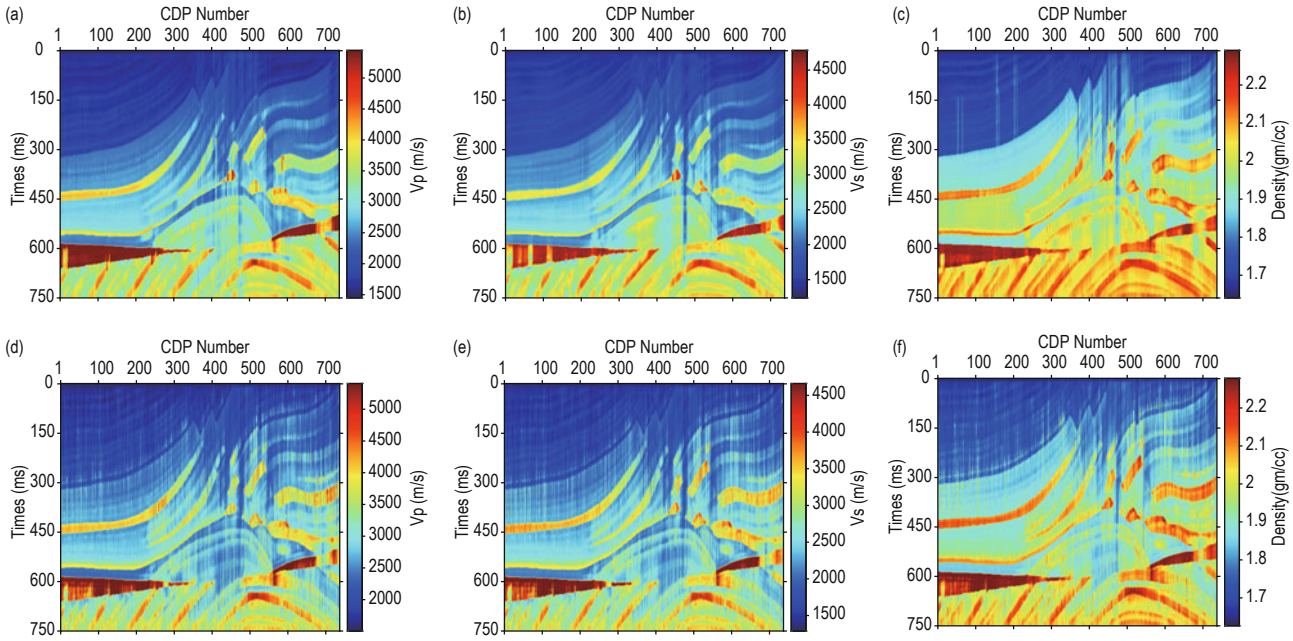

Figure 8. Comparison of STL and MTL inversion results at SNR = 5.
(a) Vp (MTL); (b)Vs (MTL); (c) Density (MTL); (d) Vp (STL); (e) Vs (STL); (f)Density (STL).

Table 3 Quantitative evaluation of inversion results based on STL

SNR	Vp-MAPE (%)	Vs-MAPE (%)	Den-MAPE (%)
15	5.35	5.69	1.57
10	8.20	7.67	2.07
5	9.36	9.01	2.39
2	9.77	9.71	2.48

From the inversion results of the noisy data in Figure 8, the inversion profile obtained by MTL is cleaner and more continuous, indicating that MTL networks are more suitable for prestack data with a low SNR. The STL inversion shows the "hanging noodles" phenomenon at the anticline between 4.5 and 6 s, indicating a deviation in the prediction results. In contrast, the MTL inversion improves the lateral continuity, and the "hanging noodles" phenomenon disappears at the anticline, better reflecting the true underground geological structure. From the quantitative evaluation of the inversion results corresponding to Tables 2 and 3, under the same noise-level conditions, the MTL inversion has better adaptability to noisy data and produces better inversion results. When the SNR is relatively high (SNR = 15), the improvement in prediction accuracy resulting from MTL is not significant, with MAPE increasing by only 0.16%, 0.05%, and 0.21%. However, when the SNR was relatively low (SNR = 2), MTL significantly improved the prediction accuracy, with MAPE increasing by 4.42%, 2.1%, and 0.75%. The tests on noisy data show that MTL can use the correlation among data to achieve data enhancement, balance the noise of different tasks, and improve the accuracy of inversion results for noisy data.

Field Data Processing

The proposed inversion method was applied to real data from an oilfield. The prestack data were obtained from a potential oil reservoir block in an eastern oilfield exploration area of China. The reservoir in this block is a heavy oil reservoir with meandering river deposition,

braided river deposition, and a fan delta. The shallow reservoir has loose characteristics with high porosity and permeability. The deep reservoir has either low porosity and permeability or medium porosity and permeability characteristics. Figure 9 shows a well-to-well seismic profile passing through six wells with a total of 428 traces, and each trace records 901 time samples. The time range is 1.6–2.2 s, with a time sampling interval of 1 ms. To test and validate the proposed method, five wells are selected as training wells, whereas the remaining one well (CDP 12) is used as the testing well.

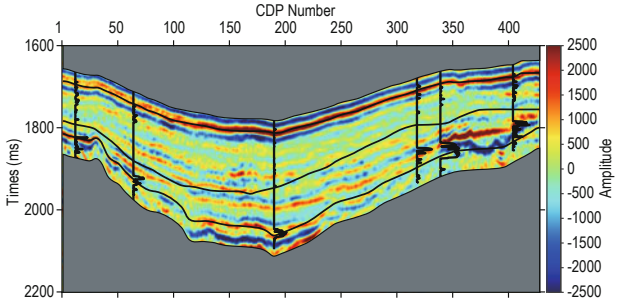

Figure 9. Well-to-well seismic profile.

Using the well-to-well logging data and the near-well data shown in the above figure to train the multitask FCN–Bi-GRU model, we performed 500 iterations. The other settings are the same as in the synthetic data example, except for the iteration number and training dataset. The trained network model was then applied to the entire seismic profile to obtain inversion results. Figure 10 shows the MTL inversion results for P-velocity, S-velocity, and density. From the 2D inversion results, the overall trend is an increase from shallow to deep, which is consistent with real geological rules. This forms a solution with a distinct layered background and has more significant geological significance, effectively reflecting the lateral distribution characteristics and vertical stacking relationship of the reservoir. In addition, the inversion results showed high resolution in both vertical and horizontal directions with good lateral continuity.

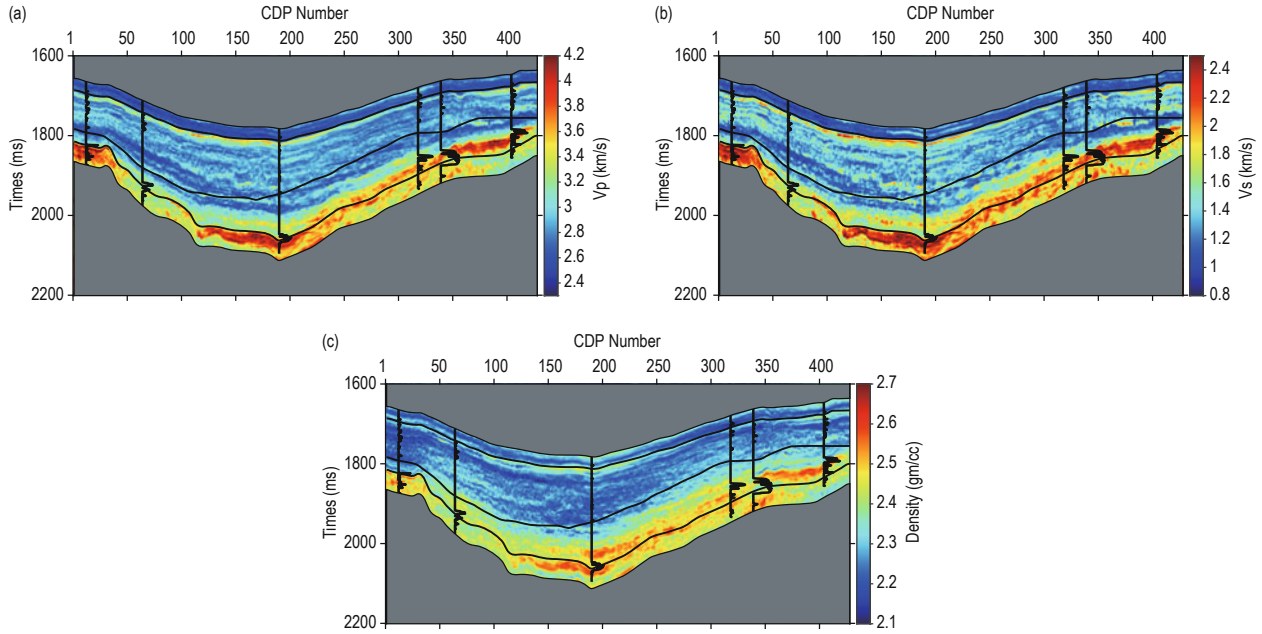

Figure 10. Multitask FCN–Bi-GRU elastic parameter inversion results (a) Vp, (b) Vs, and (c) Density.

We further validate and analyze the inversion results of the proposed method by comparing the inversion results of the training well trace (CDP 64) and validation well trace (CDP 12). As shown in Figure 11, the MTL inversion results at CDP 64 match well with the true results and provide a more pronounced characterization of details, indicating that the network has been sufficiently trained on the training well positions. Both the traditional and MTL inversion methods well fit the true well curve in the range of 1.6–1.7 s at CDP 12. However, in the formation area with a larger vertical span from 1.7 to 1.9 s, the MTL inversion results deviate more from the true curve than the traditional method's results, and its predicted results are underestimated compared with the true values. At the blind well location, the correlation coefficient between the predicted and true

P-velocities is 0.90, that between the predicted and true S-velocities is 0.8523, and that between the predicted and true densities is 0.8736.

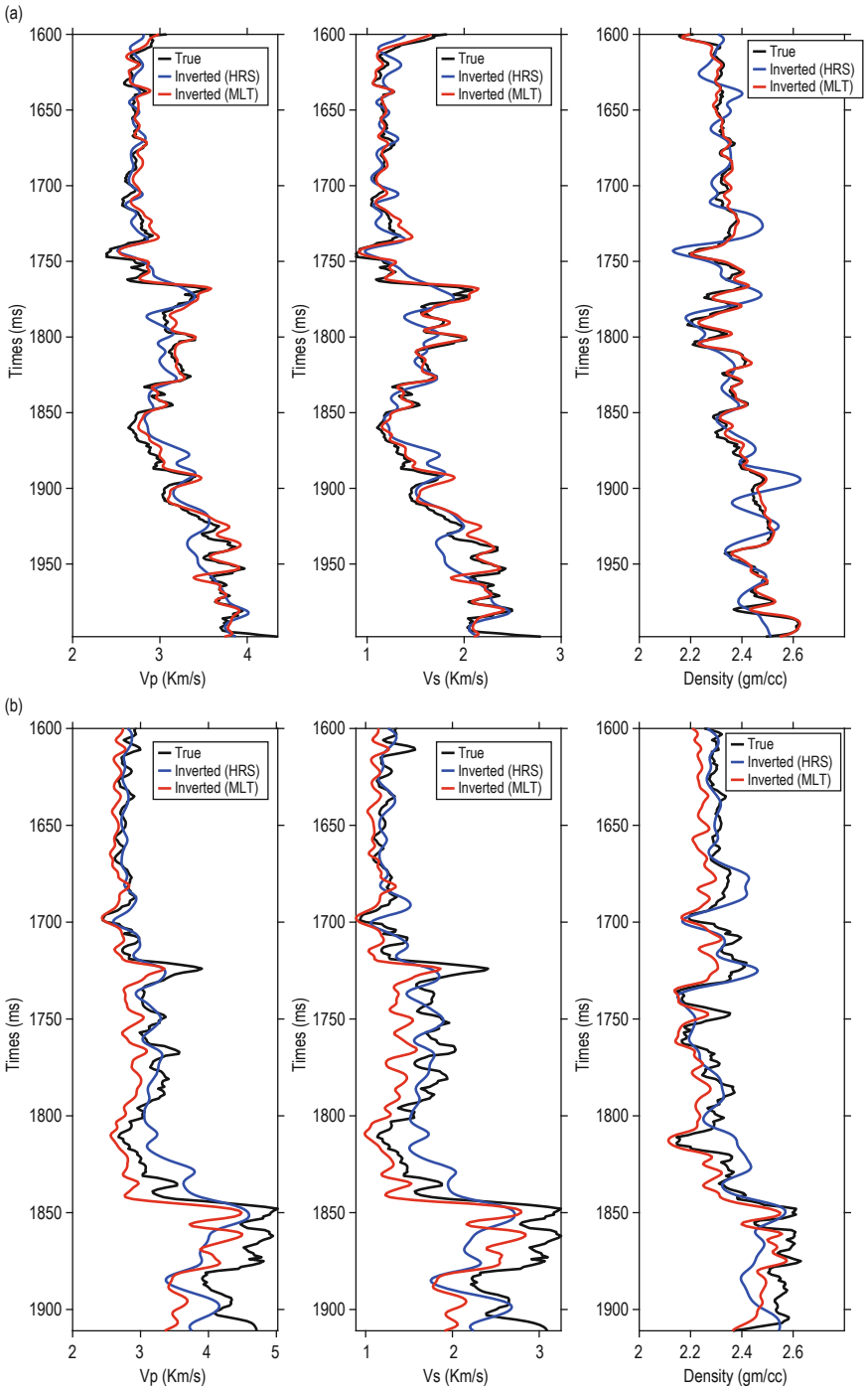

Figure 11. Comparison of real results (black curves), traditional inversion results (blue curves), and multitask inversion results (red curves) at (a) CDP 64 location in training well and (b) CDP 12 location in test well.

Conclusion

Considering a DL seismic inversion algorithm, we attempt to use MTL for prestack elastic parameter inversion and propose a weight-adaptive DL prestack seismic inversion method that combines the advantages of CNNs and Bi-GRUs. The following findings and

conclusions are obtained from tests performed on synthetic and real data.

The proposed method uses MTL for high-accuracy prestack inversion and can maintain the rock physics relationships among different elastic parameters during the inversion process. It has a high accuracy for different elastic parameters and strong adaptability.

The weight-adaptive prestack inversion method uses a loss function based on homoscedastic uncertainty, which can automatically balance the loss weight of different elastic parameter inversion tasks during the MTL process, reduce uncertainty in the inversion process, and improve inversion accuracy.

Notably, although the proposed method has achieved good results, it does not consider the distinguishing effect of elastic parameters on different geological bodies. Both geological facies and elastic parameters describe the same underground reservoir from different angles. In the future, geological facies can be introduced as prior information into neural network training to form a phase-controlled prestack inversion, thereby further improving the elastic parameter prediction accuracy. In addition, the main parameters in this present study are P-wave velocity, S-wave velocity, and density. In the future, the MTL method can be expanded to include additional parameters. By exploiting the correlation among different parameters, suitable network structures can be designed to achieve joint predictions of multiple parameters using MTL, thereby further improving the accuracy of different prestack parameter predictions.

References

Abu-Mostafa Y S. Learning from hints in neural networks[J]. Journal of complexity, 1990, **6**(2): 192–198.

Yuan S, Liu J, Wang S, et al. Seismic waveform classification and first-break picking using convolution neural networks[J]. IEEE Geoscience and Remote Sensing Letters, 2018, **15**(2): 272–276.

Wang J, Xiao Z, Liu C, et al. Deep learning for picking seismic arrival times[J]. Journal of Geophysical Research: Solid Earth, 2019, **124**(7): 6612–6624.

Yang F, Ma J. Deep-learning inversion: A next-generation seismic velocity model building method DL for velocity model building[J]. Geophysics, 2019, **84**(4): R583–R599.

Araya-Polo M, Jennings J, Adler A, et al. Deep-learning tomography[J]. The Leading Edge, 2018, **37**(1): 58–66.

Qian F, Yin M, Liu X Y, et al. Unsupervised seismic facies analysis via deep convolutional autoencoders[J]. Geophysics, 2018, **83**(3): A39–A43.

He S, Song Z, Zhang M, et al. Incremental semi-supervised learning for intelligent seismic facies identification[J]. Applied Geophysics, 2022, **19**(1): 41–52.

Huang L, Dong X, Clee T E. A scalable deep learning platform for identifying geologic features from seismic attributes[J]. The Leading Edge, 2017, **36**(3): 249–256.

Wu X, Liang L, Shi Y, et al. FaultSeg3D: Using synthetic data sets to train an end-to-end convolutional neural network for 3D seismic fault segmentation[J]. Geophysics, 2019, **84**(3): IM35–IM45.

Das V, Pollack A, Wollner U, et al. Convolutional neural network for seismic impedance inversion[J]. Geophysics, 2019, **84**(6): R869–R880.

Yuan S, Jiao X, Luo Y, et al. Double-scale supervised inversion with a data-driven forward model for low-frequency impedance recovery[J]. Geophysics, 2022, **87**(2): R165–R181.

Yuan S, Wang S, Luo Y, et al. Impedance inversion by using the low-frequency full-waveform inversion result as an a priori model[J]. Geophysics, 2019, **84**(2): R149–R164.

Sang W, Yuan S, Han H, et al. Porosity prediction using semi-supervised learning with biased well log data for improving estimation accuracy and reducing prediction uncertainty[J]. Geophysical Journal International, 2023, **232**(2): 940–957.

Cui R, Cao D, Liu Q, et al. VP and VS prediction from digital rock images using a combination of U-Net and convolutional neural networks Velocity prediction from digital rock by DL[J]. Geophysics, 2021, **86**(1): MR27–MR37.

Biswas R, Sen M K, Das V, et al. Prestack and poststack inversion using a physics-guided convolutional neural network[J]. Interpretation, 2019, **7**(3): SE161–SE174.

Du J, Liu J, Zhang G, et al. Pre-stack seismic inversion using SeisInv-ResNet[M]//SEG Technical Program Expanded Abstracts 2019. Society of Exploration Geophysicists, 2019: 2338–2342.

Sun Y, Liu Y, Zhang M, et al. Inversion of low-to medium-frequency velocities and densities from AVO data using invertible neural networks[J]. Geophysics,

2022, **87**(3): A37–A42.

Cao D, Su Y, Cui R. Multi-parameter pre-stack seismic inversion based on deep learning with sparse reflection coefficient constraints[J]. Journal of Petroleum Science and Engineering, 2022, **209**: 109836.

Kendall A, Gal Y, Cipolla R. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7482–7491.

Aki K, Richards P G. Quantitative seismology[M]. 2002.

Cheng Jian-yong received a bachelor's degree in surveying and prospecting technology and engineering from Yangtze University, Hubei, China, in 2020 and received a master's degree in geological resources and geological engineering from China University of Petroleum-Beijing, Beijing, China, in 2023. His research interests include artificial intelligence and reservoir prediction.

E-mail: 17371273547@163.com.