Transformer and Convolutional Hybrid Neural Network for Seismic Impedance Inversion

Chunyu Ning, Bangyu Wu, *Member, IEEE*, and Baohai Wu

***Abstract*—The inversion of elastic parameters especially P-wave impedance is an essential task in seismic exploration. Over the years, deep learning methods have made significant achievements in seismic impedance inversion, and convolutional neural networks (CNNs) become the dominating framework relying on extracting local features effectively. In fact, the elastic parameters temporal correlation consists of local and global characteristics, with the latter as a general trend in vertical direction due to gravity and diagenesis (vertical mechanical compression). Therefore, considering the excellent performance in capturing global dependencies of Transformer, we design an improved transformer encoder, a transformer and convolutional hybrid neural network (trans-CNN), for seismic impedance inversion. The designed network not only has the ability of transformer capturing global features with the facilitation of parallel computing but also the advantage of extracting local features of CNNs. With sparse well log data as labels, it can infer the absolute impedance from seismic data without an initial model. We also devise a relative time interval prediction self-supervised task to assist the network in better extracting seismic data features without adding any labels. Therefore, a multitask framework composed of self-supervised and supervised learning is used to train the network. We first conduct experiments on the Marmousi2 and overthrust model. The prediction profiles show that the proposed trans-CNN has better inversion and transfer learning ability than several comparable networks. We then test the proposed network on a field data, the experiments further suggest that trans-CNN can obtain stable inversion results with better horizontal continuity and high vertical resolution.**

***Index Terms*—Impedance inversion, seismic data, self-supervised learning, transformer.**

I. Introduction

SEISMIC inversion is a process to obtain physical properties and spatial structure of strata based on seismic image. It plays a vital role in the stratigraphic characterization of underground space and the evaluation of reservoir physical properties [1], [2]. P-wave impedance is a key elastic parameter for the exploration and identification of a reservoir. It indicates the property of the reservoirs and is indispensable in hydrocarbon prediction.

Manuscript received 8 August 2023; revised 20 December 2023; accepted 13 January 2024. Date of publication 25 January 2024; date of current version 12 February 2024. This work was supported in part by the Natural Science Basic Research Program of Shaanxi under Program 2023-JC-YB-269 and in part by the National Natural Science Foundation of China under Grant 41974122. *(Corresponding author: Bangyu Wu.)*

Chunyu Ning and Bangyu Wu are with the School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: ncy0312@stu.xjtu.edu.cn; bangyuwu@xjtu.edu.cn).

Baohai Wu is with The CGG GeoSoftware (Beijing), Beijing 100016, China (e-mail: baohaiwu@163.com).

Digital Object Identifier 10.1109/JSTARS.2024.3358610

Nowadays, for the seismic inversion methods, there are two major categories: model-driven and data-driven inversions. Model-driven seismic inversion usually relies on good low-frequency initial models and precise physical priors, which include deterministic methods [3], [4], [5], [6], [7] and stochastic methods [8], [9]. Due to frequency band mismatch, approximation of physical principles (forward modeling), and unavoidable noise in field seismic data [10], [11], it is often difficult for traditional model-driven methods to stably obtain high-resolution inversion results. Over the past few years, with the vigorous growth of machine learning and big data analysis, data-driven inversion methods especially deep learning have achieved tremendous progress in seismic inversion. In 2019, deep neural network model was used to invert multiple elastic parameters [12]. As the rapid growth of convolutional neural networks (CNNs), they are investigated intensively to solve the seismic inversion problem. It showed that CNNs have great potential in predicting high-accuracy impedance from band-limited seismic signals [13], [14], [15]. Almost at the same time, the full convolutional neural network (FCN) was proposed, which substituted all the fully connection layers with convolutional layers and made a breakthrough in seismic inversion [16]. Wu et al. [17] incorporated residual convolutions into FCN (FCRN) and transfer learning to improve model generalization ability for different synthetic models. Zheng et al. [18] extended FCRN to multitask learning impedance inversion with seismic data reconstruction as an auxiliary task. In addition, semisupervised methods such as CycleGAN, WcycleGAN, and geophysical-guided cycle-consistent GAN are also explored for seismic inversion tasks [19], [20], [21], [22].

The above methods generally use convolution to extract features in seismic data. Essentially, seismic traces can be regarded as time series and the magnitude of subsurface elastic parameters are increasing in a general trend due to gravity and diagenesis (vertical mechanical compression), which shows long-range and global characteristics. In this regard, the recurrent neural networks (RNNs) [23] and their variants long short-term memory (LSTM) [24], [25] and gated recurrent unit (GRU) [26] were successively utilized to model the long-term dependence in seismic sequences. GRU and bidirectional GRU were respectively combined with convolutional layers to extract information on seismic traces, enhancing the effectiveness of inversion networks [27], [28]. Nevertheless, the inherent gradient disappearance and inability to compute in parallel limit their capabilities for efficient applications. Besides, Mustafa et al. [29] applied the time convolutional network (TCN) to

© 2024 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

acoustic impedance inversion. TCN encapsulates the strengths of RNNs and CNNs and can capture low-frequency features with fewer network parameters. However, the above-mentioned sequence-to-sequence models are generally unable to accurately process long sequences. The attention mechanism can preserve and utilize all the information of the input during decoding, thus mitigating this problem. It assigns higher weights to the important and relevant features, enhancing the accuracy of the output prediction.

The attention mechanism imitates the human visual attention system and acquires the target zone that should be focused by searching the overall image. By adjusting the weight matrix of the attention mechanism, information within and between channels can be fetched according to the importance, such as squeeze and excitation network [30], efficient channel attention network [31]. Wu et al. [32] added feature-map attention and the channel-attention mechanisms to the impedance inversion network, which greatly improved the accuracy of the proposed ResANet. The self-attention takes the correlation between input vectors as the entry point, making the network extract full-scale information according to the correlation between every two positions. Transformer takes this a step further, based entirely on self-attention modules, drawing multiple global dependencies within a sequence. Most importantly, transformer is no longer focused solely on temporal association within series but pays more attention to learning multiple semantic information.

In 2017, transformer architecture was proposed for the first time and applied to natural language processing (NLP) [33]. Its excellent performance in sequence modeling and machine translation has made it rapidly gain attention in various fields. In 2018, transformer was applied in computer vision [34]. In 2020, the proposals of object detection model DETR [35] and image classification model ViT [36] ignite the rapid development of transformer. At present, transformer has been widely used in image classification [37], [38], target detection [39], [40], image segmentation [41], [42], [43], and other research directions, sweeping the entire field of computer vision. More and more scholars and engineers begin to take advantage of the transformer as a powerful tool for different tasks.

Considering the time dependence of seismic series and the metric of transformer, we propose to use transformer for modeling low-frequency trends between seismic trace and impedance. We design an improved convolution-augmented transformer encoder for seismic inversion by analogizing time sampling points to words in NLP. In addition, we also devise a self-supervised task to help the network extract latent effective information without adding any labeled samples. On the whole, we propose a Transformer and convolutional hybrid neural network, coined as trans-CNN, for seismic impedance inversion. Trans-CNN takes advantage of long-term feature acquisition and parallel computing in transformer, while also benefiting from the local feature extraction ability of CNN. To strengthen the model robustness when labeled data is deficient, we introduce a self-supervised learning task for the modeling of the time relationship inside a trace. Generally speaking, the proposed model is trained in a multitask framework for both self-supervised and supervised learning, improving the accuracy of results even in the case of limited labeled data. In the first task, the network aims to learn the relative time interval inside seismic trace by self-supervised learning. While the other task, the network predicts the impedance by supervised learning. Due to improved transformer encoder and CNN, trans-CNN can enhance information extraction ability and inversion accuracy while ensuring efficiency. The inversion results on two synthetic models indicate that trans-CNN can inverse impedance profiles with detailed strata and also has better transfer learning ability. For the field data, the inversion profile of trans-CNN is more consistent with the well logs than several comparable networks.

The rest of this article is organized as follows. Relevant theories used in trans-CNN are introduced in Section II. The network architecture and the loss function are introduced in Section III. Section IV demonstrates the performance of trans-CNN on the synthetic and field data tests. Section V is the discussion. Finally, Section VI concludes this article.

II. Components of Trans-CNN Network

A. Positional Encoding

Instead of processing sequential data in chronological order, transformer computes in parallel. Location information needs to be encoded in the data, that is, positional encoding. For time domain seismic inversion, it measures the distance between any two time points and generalizes to input embeddings with any length.

For the vector at the position i in the sequence, the position vector is defined as

$$p_t^i = \begin{cases} \sin(w_i t), & \text{if } i\%2 = 0 \\ \cos(w_i t), & \text{if } i\%2 = 1 \end{cases} \tag{1}$$

where the frequency of the trigonometric function is

$$w_i = \begin{cases} \frac{1}{10000^{(i-1)/d}}, & \text{if } i\%2 = 0 \\ \frac{1}{10000^{i/d}}, & \text{if } i\%2 = 1 \end{cases}. \tag{2}$$

Here, d is the channel number in convolutional layers and represents the input dimension of the attention mechanism as well.

B. Multihead Self-Attention

In Fig. 1, it shows the self-attention mechanism calculation flow. Data points in a trace are now represented by the vector $x_i \in R^{1\times d}$, where d is the same as above. Each x_i is multiplied by the corresponding matrix to get query $q_i \in R^{1\times d_k}$, key $k_i \in R^{1\times d_k}$ and value $v_i \in R^{1\times d}$, where d_k denotes the number of columns in query and key. The vital step of the attention mechanism is to calculate the similarity $A_{i,j}$ between q_i and k_j, and calculation is as follows:

$$A_{i,j} = \text{softmax}\left(\frac{q_i k_j^T}{\sqrt{d_k}}\right). \tag{3}$$

Finally, we can get the b_i^{h1} corresponding to the x_i

$$b_i^{h1} = \sum_{\text{j}=1}^{T} A_{i,j} v_j. \tag{4}$$

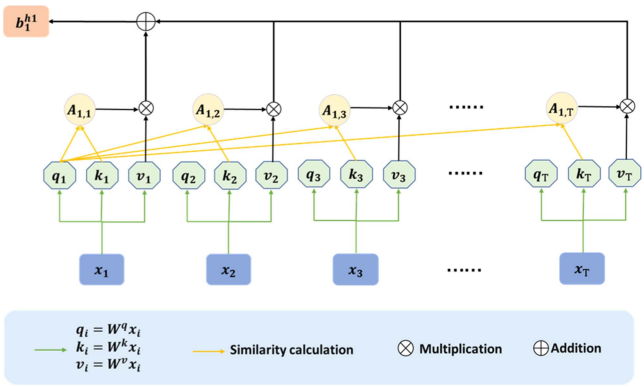

Fig. 1. Self-attention calculation flow (taking b_1^{h1} as an example).

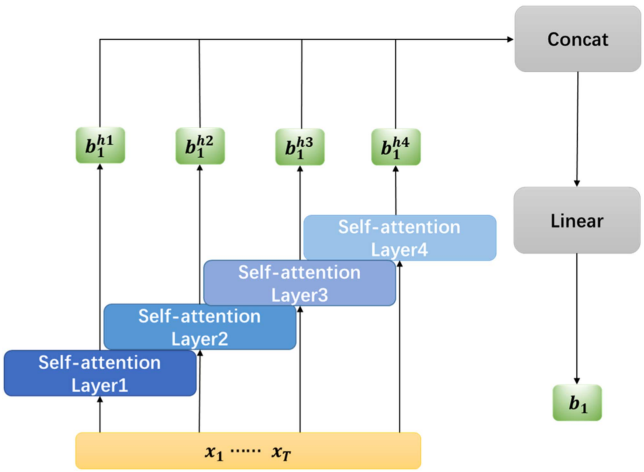

Fig. 2. Four-head self-attention calculation flow (taking b_1 as an example).

The self-attention mechanism in matrix form can be represented as follows:

$$\text{Attention}\,(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{5}$$

where all the q_i, k_i,v_i are packed together into matrix $Q \in R^{T\times d_k}$, $K \in R^{T\times d_k}$, and $V \in R^{T\times d}$. T denotes the length of seismic traces.

For single self-attention, there is one Q and K corresponding to a seismic trace, and the information included in the similarity matrix between Q and K is inadequate for complex geological structures in seismic data. Therefore, in order to obtain stronger information extraction capability, the number of Q, K, and V is increased in multihead self-attention. In general, the equation for the H-head self-attention can be represented in a similar way [33]

$$\text{MHAttn}\,(Q,K,V) = \text{Concat}\,(\text{head}_1,\ldots,\text{head}_H)\,W \tag{6}$$

$$\text{head}_m = \text{Attention}\left(QW_m^Q, KW_m^K, VW_m^V\right) \tag{7}$$

where the $W_m^Q \in R^{d_k\times D_k}$, $W_m^K \in R^{d_k\times D_k}$, $W_m^V \in R^{d\times d}$, and $W \in R^{HT\times d}$ are learned weights of the linear projection. In Fig. 2, an example of four-head self-attention module is illustrated.

We use the multihead self-attention to enhance feature extraction and especially long-term trends in seismic traces, in order to help the network achieve better accuracy and lateral continuity results. The head number in the multihead self-attention is set to 4, the dimension of the feed-forward layers to 1024, and the parameters d_k, D_k to 8.

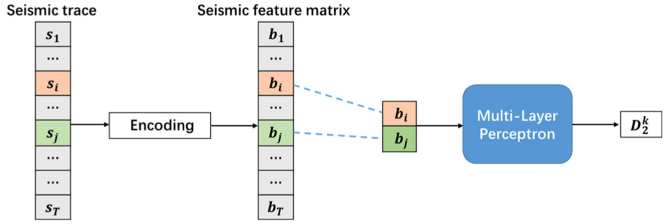

Fig. 3. Relative time interval prediction process.

C. Relative Time Interval Prediction Self-Supervised Learning

Deep learning is mainly to summarize the internal rules with various levels of representations, which usually requires the support of large-scale training samples. In the field of seismic inversion, insufficient label data has always been a problem for supervised deep-learning methods in practical applications due to limited well logs. We apply self-supervision to mitigate this problem. Self-supervised learning was initially used successfully in NLP as a pretext task to replace expensive annotations and obtain supervision from text instead [44], [45].

In trans-CNN, we use the measure of the relative location of temporal data random pairs as a self-supervised learning task. In this process, the network obtains the temporal information of seismic traces and improves the inversion performance without adding any additional labels [46]. Fig. 3 shows the relative time interval self-supervised learning process of a seismic trace. We can select the k_{th} time sampling pair i,j in a trace of length T, and their true distance is defined as

$$D_1^k = \frac{|i-j|}{T},\ i,j \in [1,\ T]\,. \tag{8}$$

After the trace passes through all encoding layers, each time sampling point is represented by a feature vector. Then we take out the vector b_i, b_j and send them into MLP as a random pair. The MLP calculates $D_2^k = \text{MLP}(b_i, b_j)^T$ to predict the distance between two points.

In this article, the number of random pairs M is related to the seismic trace length T. On synthetic models, M is 100 and on the field data M is 10. Finally, we define the relative time interval loss by mean absolute error as

$$\text{loss\_}t = \frac{1}{M}\sum_{k=1}^{M}\left|D_1^k - D_2^k\right|. \tag{9}$$

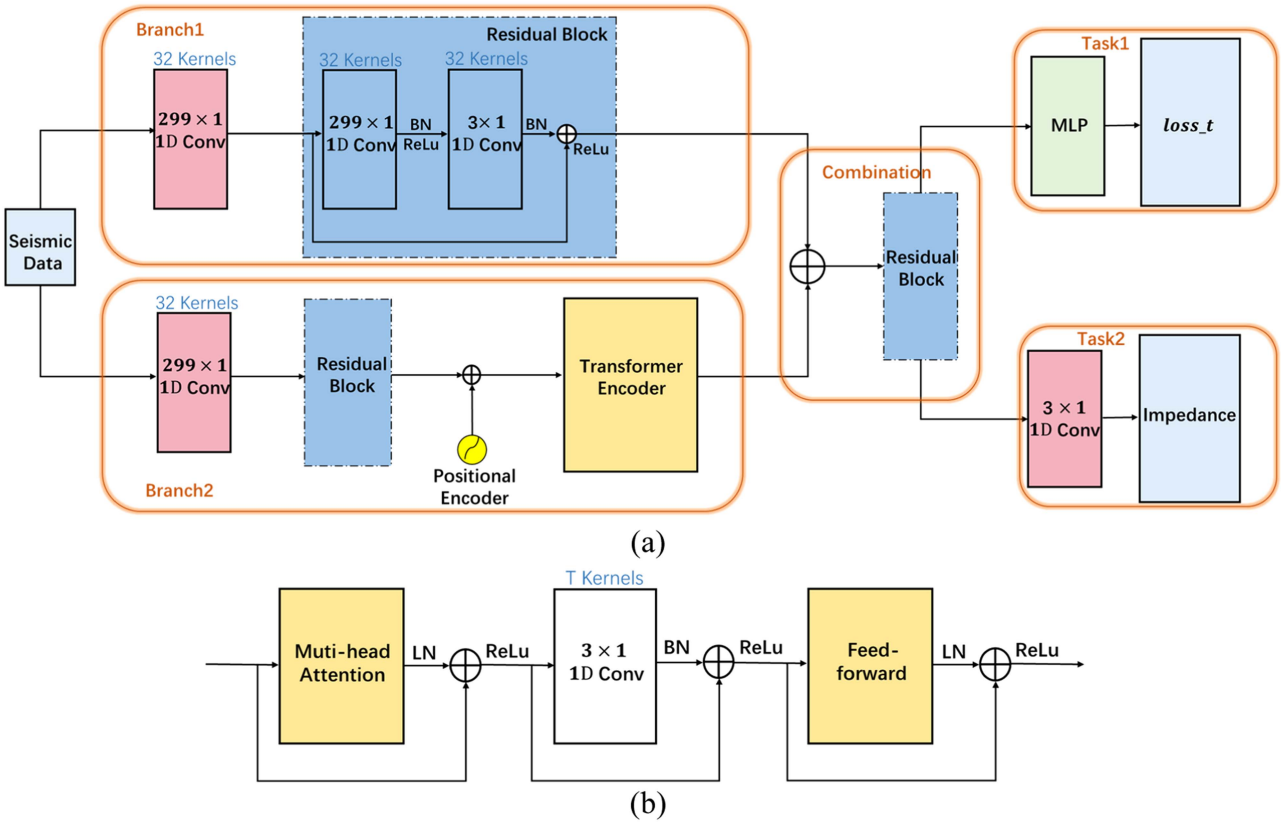

Fig. 4. Architecture of the trans-CNN. (a) Trans-CNN. (b) Convolution-augmented transformer encoder.

III. Trans-CNN Network for Seismic Impedance Inversion

A. Network Architecture

Fig. 4(a) shows the designed architecture of trans-CNN. This network consists of five main submodules with names Branch1, Branch2, Combination, Task1 and Task2. And, each submodule fulfills a different role in the overall network.

The Branch1 submodule extracts local features from seismic traces to model high-frequency details. The first part of Branch1 is a convolutional layer with 32 kernels of size 299×1, which changes according to the length of the source wavelet. It is followed by a residual block mainly consisting of two convolutional layers. One is 32 convolution kernels with size 299×1 and the other is 3×1. In residual blocks, batch normalization (BN) is applied to accelerate the training speed of the network after all convolutional layers, and the rectifiied linear unit (ReLu) helps the network depict complex nonlinear relationships.

The Branch2 submodule captures global dependencies of seismic traces and models their low-frequency content. In the Branch2, the convolutional layer and the residual block are to map the original seismic traces to corresponding feature matrices. Then the feature matrices are added positional information by positional encoder and sent into transformer encoder for global features extraction. While transformer focuses on capturing long-term dependencies, it may ignore important relative-offset-based local correlations. Therefore, we improve the structure of transformer encoder by adding a convolutional layer after attention to make the encoder pay attention to global and fine-grained correlations at the same time for more accurate inversion results. The details of convolution-augmented transformer encoder are shown in Fig. 4(b). Multihead self-attention, as Fig. 2, is the first and most important part of the transformer encoder. Then, a convolutional layer of T kernels with size 3×1 is added in the middle to capture the relative-offset-based local correlations, where T is the length of the seismic trace. The feed-forward layer follows at the end, which consists of two fully connected layers [47]. Moreover, the first part multihead self-attention and the last part Feed-forward layer adopts residual connection, layer normalization, and ReLu, while the middle convolutional layer uses BN.

In the combination submodule, there is a residual block same as previous to combine global and local information from two branches. The network starts to perform two tasks after the combination submodule. Task1 is applied to learn the relative time interval using MLP, which is mainly composed of two fully connected layers. Task2 predicts the impedance of input seismic traces, this task uses a 1 kernel convolutional layer of size 3×1 to decode the data.

B. Loss Function

The loss function in this article is composed of the relative time interval loss loss_t and the predicted impedance loss

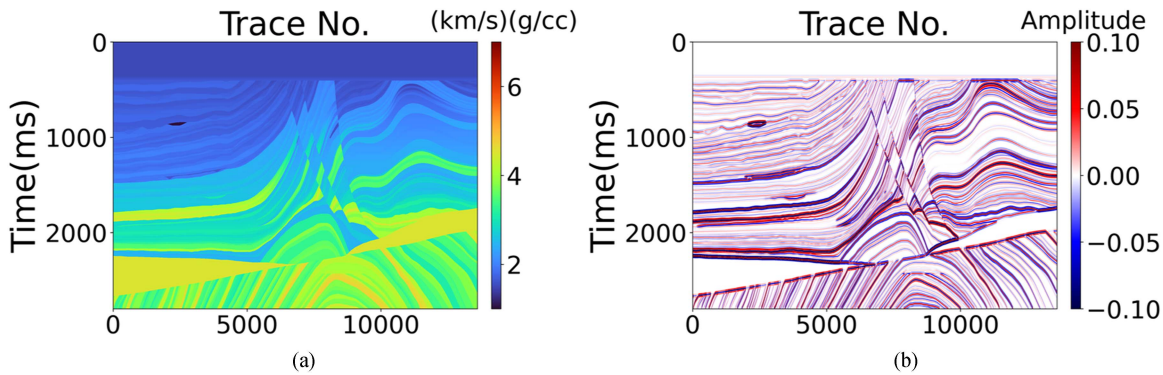

Fig. 5. Marmousi2 model. (a) Impedance. (b) Corresponding seismic data synthesized by source Ricker wavelet with 30 Hz 0° phase.

loss_ip. The loss_t is the output of Task1 in the network. And the loss_ip uses mean square error (MSE) to compute the loss of the predicted impedance (y_i') and the true value (y_i), written as

$$\text{loss\_}ip = \frac{1}{N}\sum_{i=1}^{N}(y_i' - y_i)^2. \tag{10}$$

We summarize the loss as

$$\text{Loss} = \text{loss\_}ip + \lambda \text{loss\_}t \tag{11}$$

where λ is a hyperparameter related to the input data.

IV. Experiments

The experiments are mainly divided into three parts according to data sources: the Marmousi2 synthetic model, the overthrust synthetic model used for transfer learning, and the field data are trained and predicted using FCRN, RNN-CNN, LSTM-CNN and trans-CNN networks respectively, where RNN-CNN and LSTM-CNN refer to replacing the transformer encoder in trans-CNN with a RNN/LSTM module.

A. Marmousi2 Model

The impedance of Marmousi2 model is shown in Fig. 5(a). It is used to calculate the reflection coefficient and then convolved with the source Ricker wavelet with 30 Hz 0° phase to synthesize corresponding seismic data [see Fig. 5(b)]. The Marmousi2 model is built to simulate the geological environment of continental drift. It includes several typical geological structures, for instance, great changes of velocity in both temporal and spatial directions and faults [48]. The model has 13 601 traces and 2800 time sampling points in each trace, with time intervals of 1 ms.

For this model, 136 traces are selected through isometric sampling to train the network, of which 90% are taken into the training dataset and 10% into the validation dataset. The validation dataset evaluates network performance every 50 epochs, for selecting the parameters with the least validation loss as the final parameters of trained networks. The batch size and epochs are designated as 5 and 1000. In the loss function, the value of hyperparameter λ is set to 1. We use Adam optimizer to update network parameters, setting its weight decay to 1×10^{-7} and learning rate to 0.001 respectively. Network training is executed under PyTorch framework and GPU-accelerated calculation is adopted.

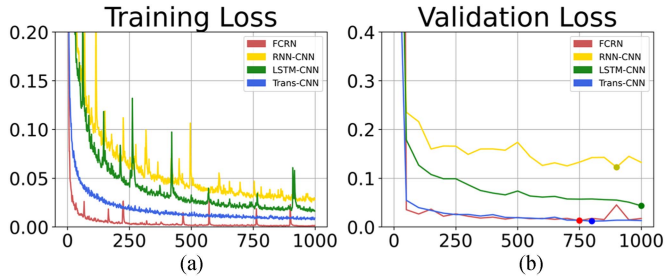

Fig. 6. Loss curves in training and validation process of the FCRN, RNN-CNN, LSTM-CNN, and trans-CNN on Marmousi2 model (the epochs of final parameters have been marked with dot). (a) Training loss. (b) Validation loss.

We can see the training loss curves from four networks in Fig. 6(a) and validation loss curves in Fig. 6(b). In Fig. 6(b), the epochs of final network parameters have been marked with dot on corresponding curves. For training loss curves, the trans-CNN is higher than FCRN with small margin. This is mainly because the loss function of trans-CNN includes two parts: relative time interval loss loss_t and predicted impedance loss loss_ip, while the network loss of FCRN only includes the latter. But it can be inferred from the validation loss curves in Fig. 6(b) that its generalization ability is the best among the four networks.

The impedance profiles predicted by all networks are shown in Fig. 7(a)–(d), and their corresponding residuals are shown below each one [see Fig. 7(e)–(h)]. From the 5000–10 000th traces of prediction profiles, which are more challenging to invert, we can observe that FCRN and RNN-CNN significantly deviate from the ground truth, while the LSTM-CNN and trans-CNN can predict the stratigraphic structures more accurately. It can be inferred that the latter two networks have better inversion ability when facing complex geological structures, and further inspection shows that trans-CNN achieves the best performance. In the black circles of Fig. 7(c) and (d), the trans-CNN result [see Fig. 7(d)] is with distinct structural edges and less white vertical strips. In the black circles of Fig. 7(e)–(h), there are strong changes in impedance, and only the result of trans-CNN does not have obvious errors. For the purpose of comparing the prediction details of a single trace, two traces (5729, 8542)

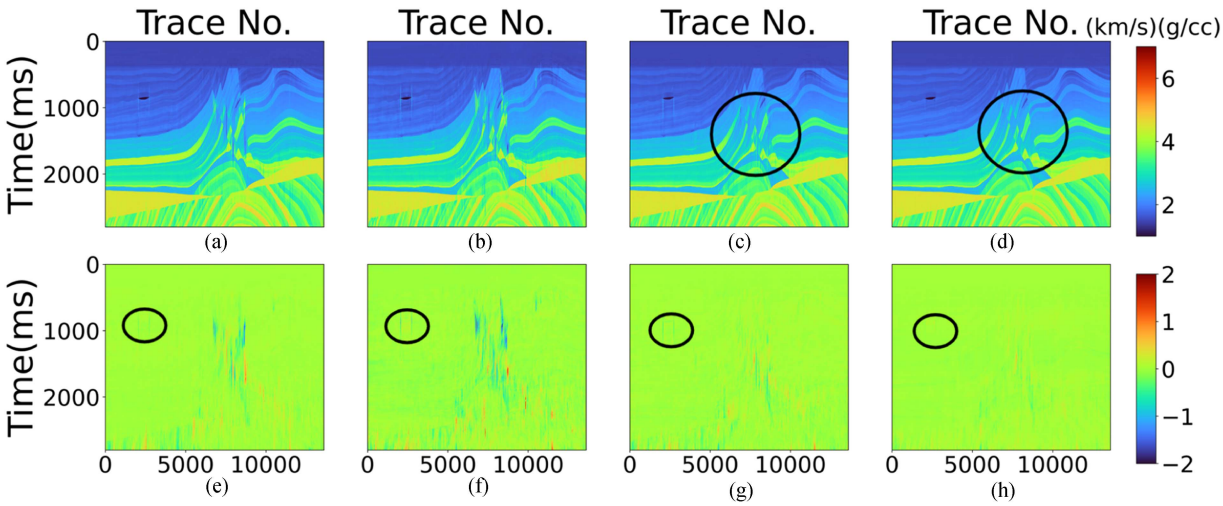

Fig. 7. Impedance prediction results and their residuals on Marmousi2 model. (a) FCRN. (b) RNN-CNN. (c) LSTM-CNN. (d) Trans-CNN. (e)–(h) Corresponding residuals in the first row.

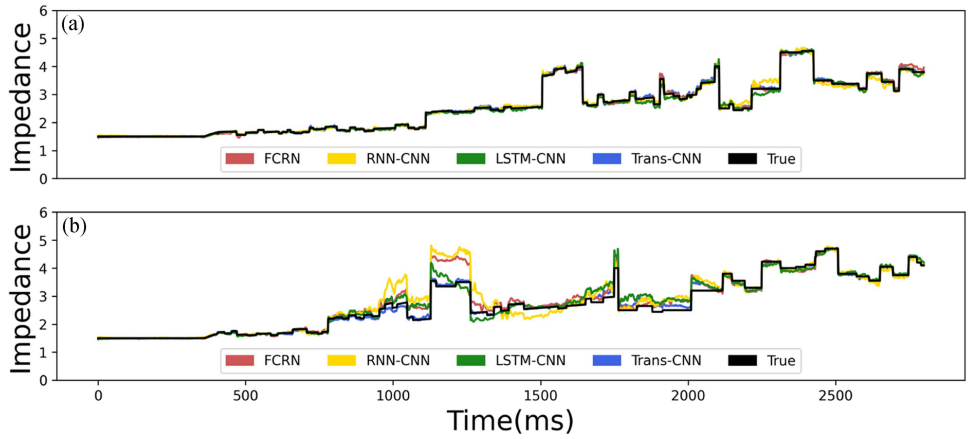

Fig. 8. Impedance prediction results of two traces on Marmousi2 model. (a) Trace 5729. (b)Trace 8542.

TABLE I
MSEs AND PCCs BETWEEN PREDICTED AND TRUE IMPEDANCE ON THE MARMOUSI2 MODEL (BEST VALUES ARE HIGHLIGHTED IN RED)

Indicators	FCRN	RNN-CNN	LSTM-CNN	Trans-CNN
MSE	0.0081	0.0181	0.0065	0.0027
PCC	0.9960	0.9909	0.9967	0.9986

TABLE II
TRAINING AND INFERRING TIME ON THE MARMOUSI2 MODEL

Indicators	FCRN	RNN-CNN	LSTM-CNN	Trans-CNN
Training	08 m 16 s	53 m 13 s	55 m 56 s	22 m 32 s
Inferring	57 s	11 m 42 s	12 m 46 s	1 m 40 s

between 5000 and 10 000th are selected for inspection [shown in Fig. 8(a) and (b)]. As shown in Fig. 8, the proposed network predicts reliable impedance for the complex stratigraphic changes. Table I shows the MSEs and Pearson correlation coefficients (PCCs) between prediction results and the ground truth in the above experiments. It can be seen that the MSE (0.0027) and PCC (0.9986) of trans-CNN are the best. In addition, the training and inferring time of networks are shown in Table II. It explains that transformer encoder has great advantages over RNN and LSTM in computing efficiency, especially in inference time.

B. Overthrust Model

Fig. 9 shows the dataset of the overthrust synthetic model. As can be seen, its geological structure is quite different from the Marmousi2 model. Therefore, we choose the overthrust model to compare the transfer learning ability of four networks: they are pretrained on the Marmousi2 model, and then fine-tuned by the overthrust model. This model consists of 401 traces with 2800 time points of 1 ms interval in each trace. We choose 5 traces (1th, 100th, 200th, 300th, and 400th) to participate in transfer learning. Due to insufficient labeled traces, cubic spline interpolation method is applied for sample augmentation. Cubic

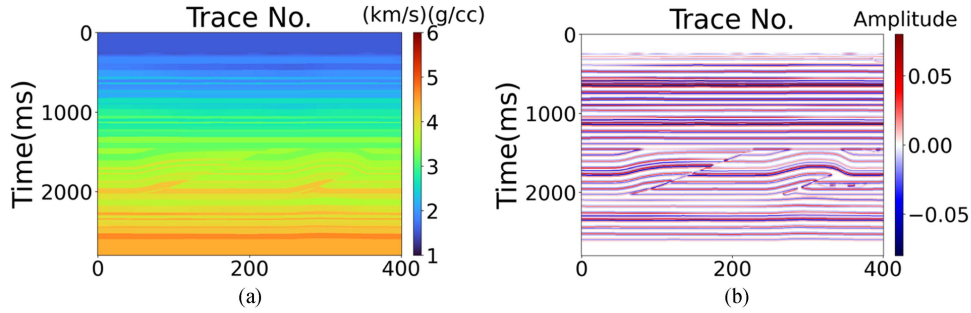

Fig. 9. Overthrust model. (a) Impedance. (b) Corresponding seismic data synthesized by source Ricker wavelet with 30 Hz 0° phase.

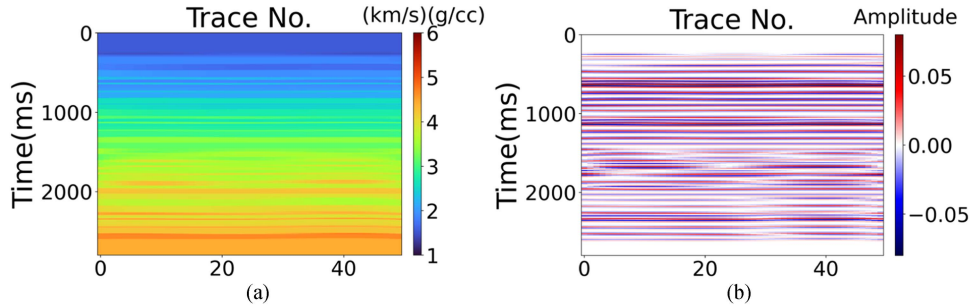

Fig. 10. Interpolated results of the overthrust model. (a) Interpolated impedance. (b) Corresponding interpolated seismic data.

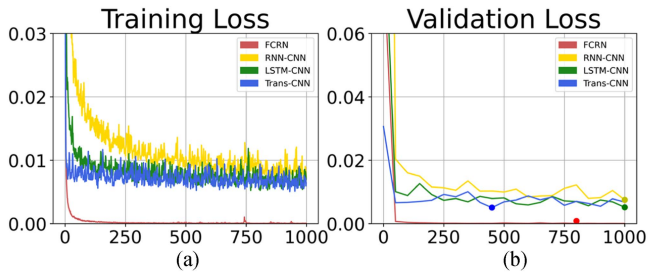

Fig. 11. Fine tuning loss curves in training and validation process of the FCRN, RNN-CNN, LSTM-CNN, and trans-CNN on the overthrust model (the epochs of final parameters have been marked with dot). (a) Training loss. (b) Validation loss.

TABLE III
MSEs AND PCCs BETWEEN PREDICTED AND TRUE IMPEDANCE ON THE OVERTHRUST MODEL (BEST VALUES ARE HIGHLIGHTED IN RED)

Indicators	FCRN	RNN-CNN	LSTM-CNN	Trans-CNN
MSE	0.0035	0.0030	0.0016	0.0007
PCC	0.9983	0.9981	0.9994	0.9997

TABLE IV
TRAINING AND INFERRING TIME ON THE OVERTHRUST MODEL

Indicators	FCRN	RNN-CNN	LSTM-CNN	Trans-CNN
Training	02 m 34 s	23 m 30 s	22 m 46 s	8 m 09 s
Inferring	0.7 s	22.6 s	24.2 s	3.6 s

spline interpolation uses the selected 5 original traces to get interpolated data from 50 traces (see Fig. 10). 90% of them are used for training and 10% for validation. The networks are estimated every 50 epochs for validation. And the hyperparameters of networks are the same as Section IV-A.

During fine-tuning, Fig. 11(a) and (b) shows the training and validation loss curves. What can be obtained is the FCRN losses are the smallest on the interpolated training and validation data. First, it is due to loss of FCRN only consist of loss_ip while losses of other three networks include loss_ip and loss_t. Then, the high similarity between the training and validation data makes the performance of four validation loss curves consistent with the training process. Fig. 12 illustrates that the generalization ability of trans-CNN exceeds the other three networks dramatically. From the prediction profiles of four networks [see Fig. 12(a)–(d)] and their residuals [see Fig. 12(e)–(h)], trans-CNN can accurately predict the structure of the thin strata and faults, while other prediction profiles are contaminated by vertical strip errors. To quantify the results, Table III shows the MSEs and PCCs between the predicted and true values. As can be seen, trans-CNN has the most outstanding performance on both MSE (0.0007) and PCC (0.9997). The training and inferring time of all networks are shown in Table IV.

C. Field Data

According to the above experimental results, trans-CNN has presented an advantage due to its global information extraction

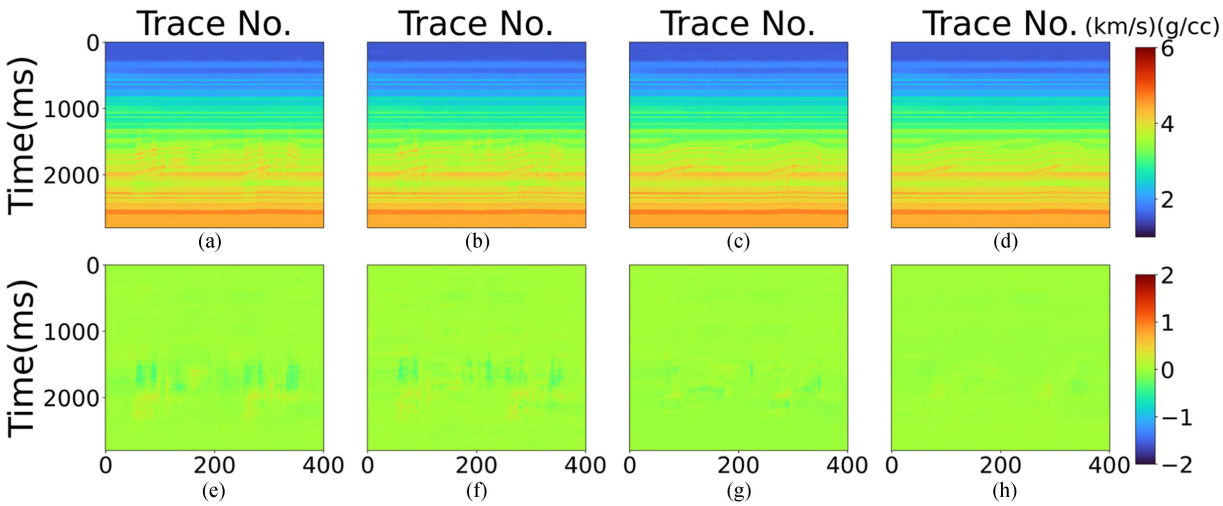

Fig. 12. Impedance prediction results with fine-tuning on the overthrust model. (a) FCRN. (b) RNN-CNN. (c) LSTM-CNN. (d) Trans-CNN. (e)–(h) Corresponding residuals in the first row.

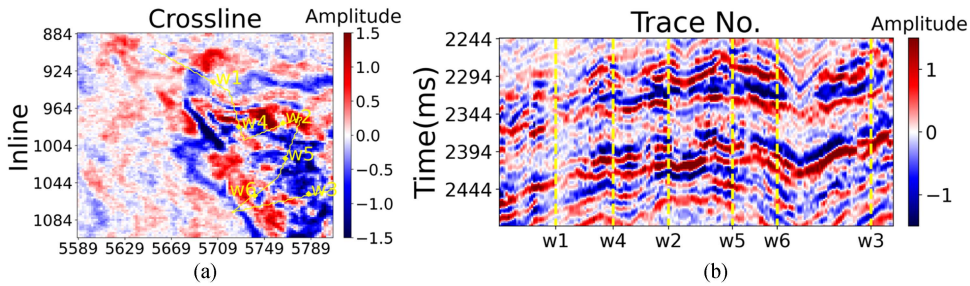

Fig. 13. Seismic data of the field data. (a) Time slice of 2300 ms, the yellow diamonds are well w1–w6 locations, and the yellow line depicts the position of cross-well profile. (b) Seismic data of cross well profile (dashed vertical lines indicate the location of wells).

on synthetic models. To further prove its effectiveness in application, we conduct experiments on a field data. A total of 12 100 (110 × 110) seismic traces are collected on the working area, having 1501 sampling points with 2 ms time interval in each trace. A random seismic data time slice is shown in Fig. 13(a). The 2244–2494 ms of traces are selected as our target layer.

The locations of 6 wells and their cross-well profile have been indicated in Fig. 13(a), and Fig. 13(b) shows the seismic data of cross-well profile. We apply a transfer learning strategy on this dataset to overcome the issue of insufficient labeled data. First, we interpolate the 6 wells impedance and use the convolution model to obtain the corresponding seismic data, they collectively form the synthetic model. The interpolated impedance and corresponding seismic profile of the synthetic model are shown in Fig. 14(a) and (b), and the source wavelet used for convolution is shown in Fig. 14(c). The networks are pretrained with the 185 seismic traces of cross-well profile (1.5% of total data) on the synthetic model, and it needs to be specifically stated that the data of 6 wells do not participate in the pretraining process. Then we use the same interpolation method for impedance and seismic data to obtain interpolated data and select the 24 traces closest to each well to fine-tune the network. With this augmentation operation, we are able to fine-tune the network by 144 seismic traces (1.2% of total data). In addition, network hyper parameters need to be adjusted as follows: the size 299 × 1 of convolution kernels is changed to 29 × 1, the Adam optimizer weight decay rate is changed to 1×10^{-6}, and the value of λ in loss fuction is changed to 0.5 according to ablation experiments. The remaining settings in the networks are same as Section IV-A.

In the pretraining and fine-tuning process of the experiments, we both make use of 90% data as a training dataset and 10% for validation. The training and testing curves obtained in the fine-tuning process are shown in Fig. 15. As we can see that the curves are consistent with those of the overthrust synthetic model. Fig. 16 shows the cross-well profiles predicted by four networks, and the impedance of wells is inserted in corresponding positions. It can be observed that there are many vertical stripes with no geological meaning in the results obtained by FCRN [see Fig. 16(a)] and RNN-CNN [see Fig. 16(b)], such as stratigraphic structure in the black boxes. And there are some incomplete stratigraphic structures that are in conflict with the continuous sedimentary model in Fig. 16(a), indicated by the arrow. For Fig. 16(c) and (d), the profile predicted by trans-CNN [see Fig. 16(d)] is more consistent with impedance in wells than LSTM-CNN [see Fig. 16(c)], such as the area indicated by the white boxes. Then, the LSTM-CNN profile shows more unrea-

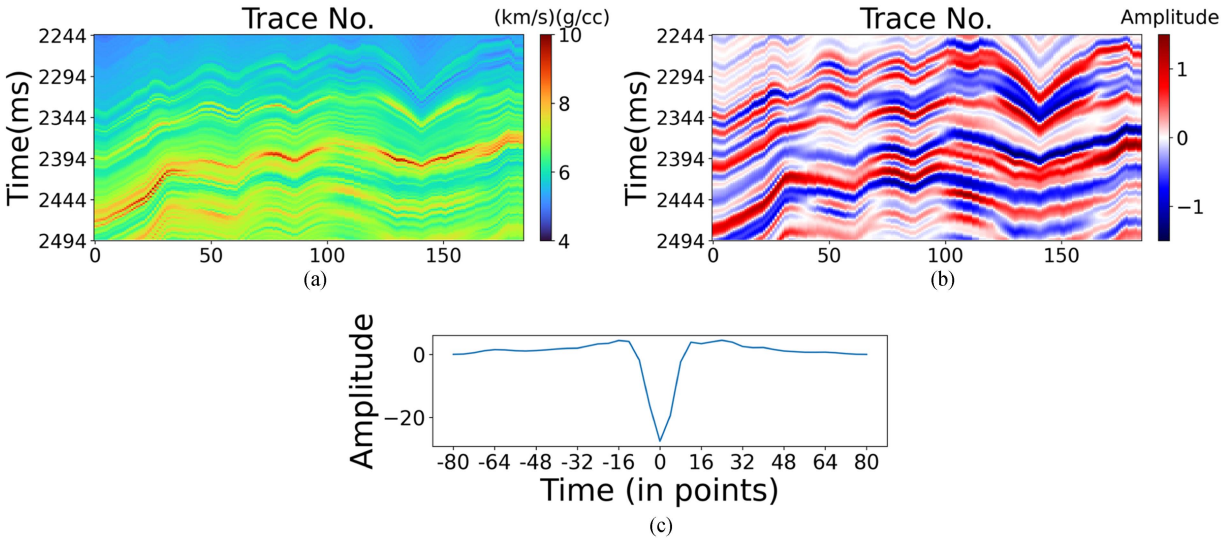

Fig. 14. Synthesize data for pretraining of the field data. (a) Impedance obtained by interpolation from wells. (b) Synthetic seismic data. (c) Source wavelet.

TABLE V
MSEs AND PCCs BETWEEN PREDICTED AND TRUE IMPEDANCE OF THE 6 WELLS ON THE FIELD DATA (BEST VALUES ARE HIGHLIGHTED IN RED)

MSE/PCC	FCRN	RNN-CNN	LSTM-CNN	Trans-CNN
W1	0.0666/0.9539	0.0852/0.9412	0.1016/0.9293	0.0815/0.9437
W2	0.0098/0.9916	0.0137/0.9878	0.0135/0.9876	0.0132/0.9879
W3	0.0360/0.9620	0.0334/0.9689	0.0266/0.9747	0.0220/0.9778
W4	0.0004/0.9997	0.0010/0.9991	0.0014/0.9987	0.0011/0.9991
W5	0.0004/0.9995	0.0016/0.9982	0.0018/0.9980	0.0010/0.9991
W6	0.0058/0.9945	0.0082/0.9928	0.0072/0.9931	0.0052/0.9951

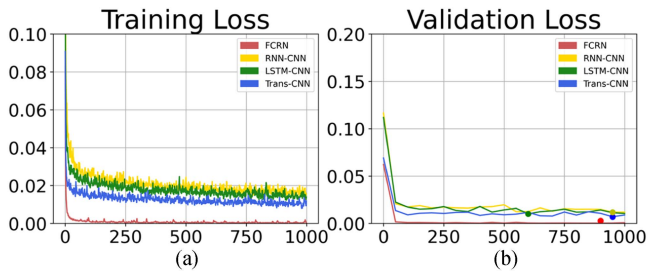

Fig. 15. Fine-tuning loss curves in the training and validation process of the FCRN, RNN-CNN, LSTM-CNN, and trans-CNN on the field data (the epochs of final parameters have been marked with dot). (a) Training loss. (b) Validation loss.

TABLE VI
AVERAGE MSE AND PCC BETWEEN PREDICTED AND TRUE IMPEDANCE OF RHE 6 WELLS ON FIELD DATA (BEST VALUES ARE HIGHLIGHTED IN RED)

Indicators	Single-task trans-CNN	Trans-CNN
MSE	0.0226	0.0674
PCC	0.9825	0.9383

sonable stratigraphic and structures than trans-CNN, especially the structure in black circle and vertical strips in black boxes. Moreover, the lateral continuity of trans-CNN is improved. The predicted and real impedance at 6 wells is plotted in Fig. 17. As we can see, the prediction values of the four networks at w1 are similar. There are large deviations with the ground truth due to rapid changes in impedance, such as time span between 2394 and 2444 ms (indicated by yellow box). Trans-CNN achieves the best performance at w3. The difference between predictions at w2 and w4–w6 is not obvious, and the predicted impedance of four networks are quite close to the ground truth. The MSEs and PCCs of 6 wells are shown in Table V. In general, trans-CNN and FCRN perform better than other networks. Taking PCC as the evaluation metric, the predicted values of trans-CNN are better at w3 (0.9778) and w6 (0.9951), while the PCC of FCRN is good at well w1 (0.9539), w2 (0.9916), w4 (0.9997), and w5 (0.9995).

To verify the function of self-supervised learning task in trans-CNN, we presented the results of the comparative experiment of single-task trans-CNN and our trans-CNN in Table VI and Fig. 18. Based on the combination of numerical indicators and figure, it can be seen that the single task trans-CNN performs better in terms of MSEs and PCCs on six wells, as it only updates network parameters according to the indicator MSE.

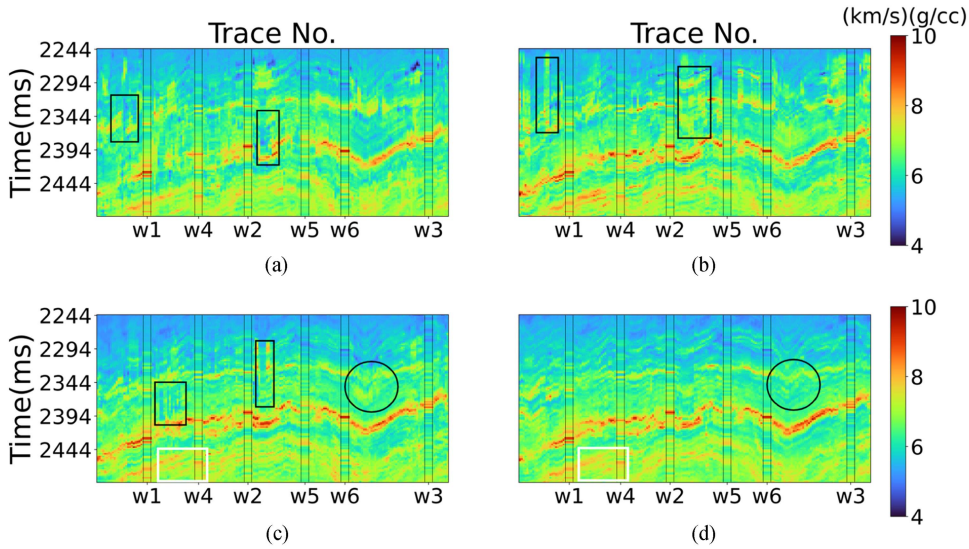

Fig. 16. Impedance prediction profiles on the field data. (a) FCRN. (b) RNN-CNN. (c) LSTM-CNN. (d) Trans-CNN.

TABLE VII
MSEs AND PCCs BETWEEN PREDICTED AND TRUE IMPEDANCE FOR THE BLIND WELL W4 (BEST VALUES ARE HIGHLIGHTED IN RED)

Indicators	FCRN	RNN-CNN	LSTM-CNN	Trans-CNN
MSE	0.0903	0.1786	0.1774	0.0674
PCC	0.9136	0.7675	0.7917	0.9383

However, its prediction results on the cross-well profile have a large number of vertical stripes. This clearly verifies that the self-supervised task can mitigate the overfitting and improve the generalization ability of the trans-CNN.

To further demonstrate the superiority of our network, we conduct blind well test on the field data. We choose w4 as the blind well and do not use any information about it in the pretraining and fine-tuning process. The predicted impedance of w4 is shown in Fig. 19. And, the MSEs and PCCs of networks are recorded in Table VII. It shows that the trans-CNN prediction result has the highest matching degree with ground truth on MSE (0.0674) and PCC (0.9383). Next, we use four networks to predict inline profile where w4 is located. We show the predicted impedance profiles in Fig. 20 and show the corresponding seismic data in Fig. 21. It is shown that trans-CNN predicted profiles [see Fig. 20(d)] have more complete stratigraphic structures with better lateral continuity.

Finally, we present the predicted results of the 3-D field data volume in Fig. 22. From the figure, it can be seen that the data volume predicted by trans-CNN has strong horizontal continuity and the time slice matches the one in Fig. 13(a).

V. DISCUSSION

At present, CNNs have become one of the fastest-growing fields in artificial intelligence. In CNNs, a unit in the networks only depends on a region of the input, which is one of the basic concepts, called receptive field. For a network, only the area within the receptive field is related to the unit value, so the accuracy of receptive field size is crucial. There are a number of ways to increase the receptive field size, of which increasing convolutional kernel size and stacking more convolutional layers are the most common ways. However, these methods have limitations in application. On one hand, the excessive increase of convolution kernel size will affect the feature extraction effectiveness; on the other hand, the unreasonable stack of the network depth may cause some original information loss in the training process. Therefore, we adopt transformer encoder architecture and use it to extract global features in trans-CNN, mitigating the limitations of receptive field size in CNN. In the experiments of synthetic models, it demonstrate that compared to RNN and LSTM, the improved transformer encoder effectively improves the accuracy and continuity of impedance inversion results and reduces training and inferring time. This is extremely crucial in exploring reservoir characteristics. In addition, the self-supervised learning task further improves the inversion performance under limited labels. When transfer learning is carried out for different datasets, trans-CNN shows the best transfer learning ability. In the field data experiments, we first use synthetic data to pretrain the network, which enables our network to learn the stable correspondence between impedance and seismic data, which is applicable to the entire data volume. During the fine-tuning process, due to the high accuracy of the interpolation data around the well, we only use these data to fine-tune networks.

Although trans-CNN performs the best among all the compared networks, there are still some directions for improvement. First, our network is trained under the multitask framework, so the mixed loss function composed of loss_t and loss_ip is used. In this article, the relative weight factor λ in the loss function is selected through ablation experiments, and cannot be guaranteed the optimal. In this regard, it can be improved to use some methods such as homescedic uncertainty [18] and gradient normalization [49] to treat the λ as a trainable parameter. Second,

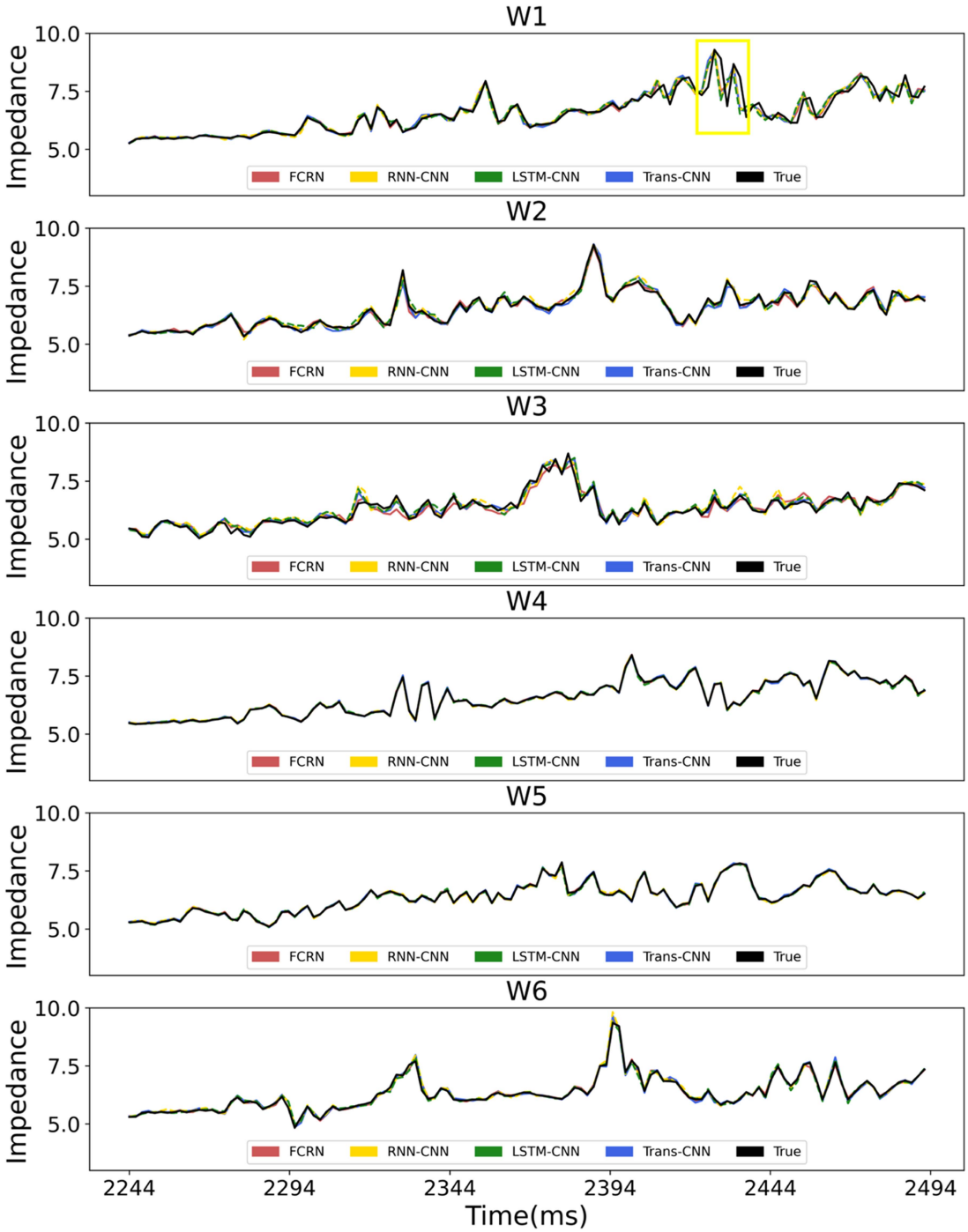

Fig. 17. Impedance prediction results of 6 wells on the field data.

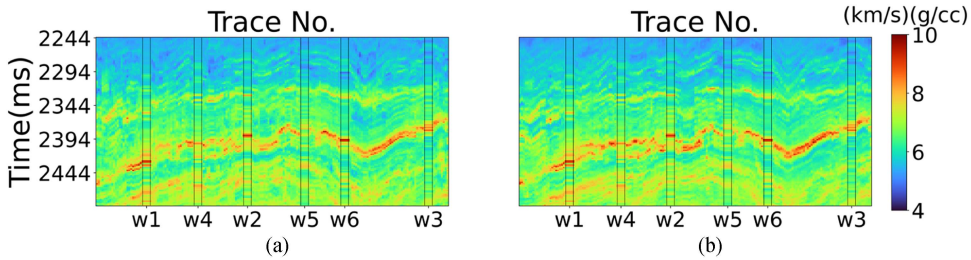

Fig. 18. Impedance prediction profiles on the field data. (a) Single-task trans-CNN. (b) Trans-CNN.

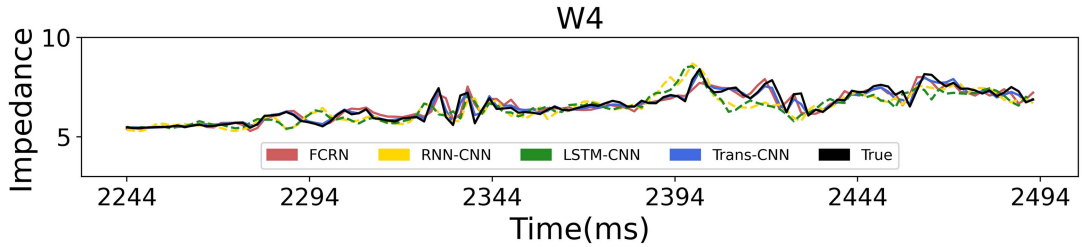

Fig. 19. Impedance prediction results of w4 on the field data for blind well test.

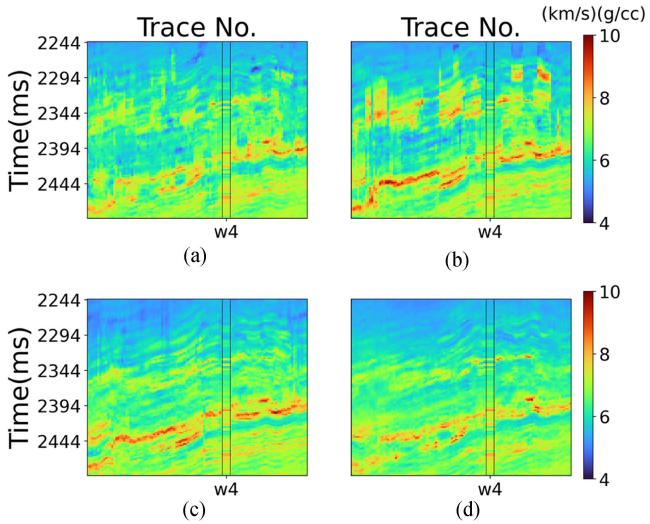

Fig. 20. Inline profile impedance prediction on the field data for blind well test. (a) FCRN. (b) RNN-CNN. (c) LSTM-CNN. (d) Trans-CNN.

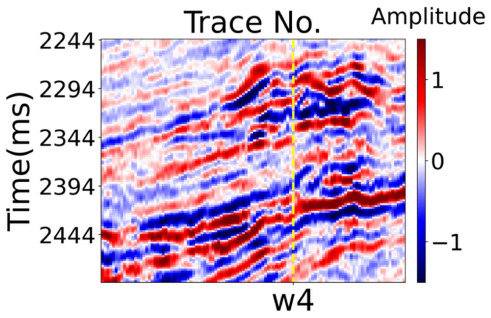

Fig. 21. Inline profile seismic data on the field data for blind well test (dashed vertical lines indicate the location of w4).

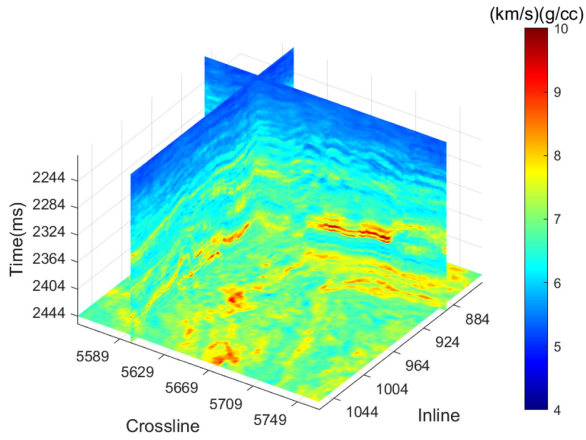

Fig. 22. Three-dimensional impedance prediction volume of trans-CNN on the field data.

we conduct antinoise tests on the networks. When the signal-to-noise ratio (SNR) is 20, trans-CNN and FCRN have similar antinoise performance, but when the SNR is lower, FCRN has stronger robustness. So the robustness of trans-CNN is a problem worth investigation. Third, the proposed network is improved on the basis of 1-D CNN. It extracts the information in each seismic trace independently but does not consider the spatial correlation among traces. Therefore, future research is going to combine the global feature extraction ability of transformer with 2-D CNN to further enhance the horizontal continuity and accuracy of predicted profiles. Last, prestack seismic data inversion can also be investigated to estimate multiple elastic parameters directly [50].

VI. Conclusion

In this article, we design an efficient hybrid neural network trans-CNN for seismic data impedance inversion. In terms of network architecture, we incorporate the transformer encoder according to the impedance inversion task and combine it with the convolutional module, so as to effectively extract global and local information at the same time. In addition, we add a relative time interval self-supervised learning task and use a multitask framework to enhance time dimension feature extraction without adding any labels.

The qualitative and quantitative comparisons on synthetic models indicate that trans-CNN obtains the most accurate prediction results. It is shown that the trans-CNN also has the best ability in transfer learning. Finally, we test our network on a field data. Faced with insufficient labeled data, we use synthesis and interpolation method for data augmentation and introduce a transfer learning strategy during network training. The proposed network obtains the most complete stratigraphic structure with a higher horizontal continuity and also achieves consistent performance on blind well test.

ACKNOWLEDGMENT

The authors would like to thank for the help of team members Z. Tang and X. Liu.

REFERENCES

[1] P. G. Lelièvre and D. W. Oldenburg, "A comprehensive study of including structural orientation information in geophysical inversions," *Geophysical J. Int.*, vol. 178, no. 2, pp. 623–637, 2009, doi: 10.1111/j.1365-246X.2009.04188.x.

[2] G. Mavko, T. Mukerji, and J. Dvorkin, *The Rock Physics Handbook*, Cambridge, U.K.: Cambridge Univ. Press, 2009, doi: 10.1017/CBO9780511626753.

[3] E. A. Robinson, "Predictive decomposition of seismic traces," *Geophysics*, vol. 22, no. 4, pp. 767–778, 1957, doi: 10.1190/1.1438415.

[4] D. A. Cooke and W. A. Schneider, "Generalized linear inversion of reflection seismic data," *Geophysics*, vol. 48, no. 6, pp. 665–676, 1983, doi: 10.1190/1.1441497.

[5] M. D. Sacchi, "Reweighting strategies in seismic deconvolution," *Geophysical J. Int.*, vol. 129, no. 3, pp. 651–656, 1997, doi: 10.1111/j.1365-246X.1997.tb04500.x.

[6] R. O. Lindseth, "Synthetic sonic logs—A process for stratigraphic interpretation," *Geophysics*, vol. 44, no. 1, pp. 3–26, 1979, doi: 10.1190/1.1440922.

[7] R. J. Ferguson and G. F. Margrave, "A simple algorithm for band-limited impedance inversion," *CREWES Annu.*, vol. 8, pp. 21–21–21–10, 1996.

[8] W. P. Gouveia and J. A. Scales, "Bayesian seismic waveform inversion: Parameter estimation and uncertainty analysis," *J. Geophysical Res., Solid Earth*, vol. 103, no. B2, pp. 2759–2779, 1998, doi: 10.1029/97JB02933.

[9] D. Carron, "High resolution acoustic impedance cross-sections from wireline and seismic data," in *Proc. SPWLA 30th Annu. Logging Symp. Soc. Petrophysicists Well-Log Analysts*, 1989.

[10] K. Baddari, J. Ferahtia, T. Aifa, and N. Djarfour, "Seismic noise attenuation by means of an anisotropic non-linear diffusion filter," *Comput. Geosci.*, vol. 37, no. 4, pp. 456–463, 2011, doi: 10.1016/j.cageo.2010.09.009.

[11] Z. T. Xu, Y. S. Luo, B. Y. Wu, and D. Y. Meng, "S2S-WTV: Seismic data noise attenuation using weighted total variation regularized self-supervised learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5908315, doi: 10.1109/TGRS.2023.3268554.

[12] Z. D. Zhang and T. Alkhalifah, "Regularized elastic full waveform inversion using deep learning," in *Proc. 81st EAGE Conf. Exhib.*, 2019, vol. 2019, pp. 1–5, doi: 10.3997/2214-4609.201901345.

[13] R. Biswas, M. K. Sen, V. Das, and T. Mukerji, "Pre-stack and post-stack inversion using a physics-guided convolutional Neural network," *Soc. Exploration Geophysicists Amer. Assoc. Petroleum Geologists*, vol. 7, no. 3, pp. 1–76, 2019, doi: 10.1190/INT-2018-0236.1.

[14] J. W. Fang, H. Zhou, Y. E. Li, Q. Zhang, and J. Zhang, "Data-driven low-frequency signal recovery using deep learning predictions in full-waveform inversion," *Geophysics*, vol. 85, no. 6, pp. 1–42, 2020, doi: 10.1190/geo2020-0159.1.

[15] Y. Wang, Q. Ge, W. Lu, and X. Yan, "Well-logging constrained seismic inversion based on closed-loop convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 8, pp. 5564–5574, Aug. 2020, doi: 10.1109/TGRS.2020.2967344.

[16] V. Das, A. Pollack, U. Wollner, and T. Mukerji, "Convolutional neural network for seismic impedance inversion," *Geophysics*, vol. 84, no. 6, pp. R869–R880, 2020, doi: 10.1190/geo2018-0838.1.

[17] B. Y. Wu, D. L. Meng, L. L. Wang, N. H. Liu, and Y. Wang, "Seismic impedance inversion using fully convolutional residual network and transfer learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 12, pp. 2140–2144, Dec. 2020, doi: 10.1109/LGRS.2019.2963106.

[18] X. Zheng, B. Y. Wu, X. S. Zhu, and X. Zhu, "Multi-task deep learning seismic impedance inversion optimization based on homoscedastic uncertainty," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2022, pp. 6149–6152, doi: 10.3390/app12031200.

[19] Y. Q. Wang, Q. Wang, W. K. Lu, Q. Ge, and X. F. Yan, "Seismic impedance inversion based on cycle-consistent generative adversarial network," in *Proc. SEG Tech. Prog. Expanded Abstr.*, 2019, pp. 2498–2502, doi: 10.1016/j.petsci.2021.09.038.

[20] A. Cai, H. B. Di, Z. Li, H. Maniar, and A. Abubakar, "Wasserstein cycle-consistent generative adversarial network for improved seismic impedance inversion: Example on 3D SEAM model," in *Proc. SEG Tech. Prog. Expanded Abstr.*, 2020, pp. 1274–1278, doi: 10.1190/segam2020-3425785.1.

[21] D. L. Meng, B. Y. Wu, N. H. Liu, and W. C. Chen, "Semi-supervised deep learning seismic impedance inversion using generative adversarial networks," in *Proc. Geosci. Remote Sens. Symp.*, 2020, pp. 1393–1396, doi: 10.1109/IGARSS39084.2020.9323119.

[22] H. H. Zhang, G. Z. Zhang, J. H. Gao, S. J. Li, J. M. Zhang, and Z. Y. Zhu, "Seismic impedance inversion based on geophysical-guided cycle-consistent generative adversarial networks," *J. Petroleum Sci. Eng.*, vol. 218, 2022, Art. no. 111003, doi: 10.1016/j.petrol.2022.111003.

[23] M. Alfarraj and G. Alregib, "Semisupervised sequence modeling for elastic impedance inversion," *Interpretation*, vol. 7, no. 3, pp. SE237–SE249, 2019, doi: 10.1190/segam2019-3215902.1.

[24] Z. Gao, C. Li, B. Zhang, X. Jiang, and Z. Xu, "Building large-scale density model via a deep-learning-based data driven method," *Geophysics*, vol. 86, no. 1, pp. M1–M15, 2021, doi: 10.1190/geo2019-0332.1.

[25] R. Guo, J. J. Zhang, D. Liu, Y. B. Zhang, and D. W. Zhang, "Application of bi-directional long short-term memory recurrent neural network for seismic impedance inversion," in *Proc. 81st EAGE Conf. Exhib.*, 2019, no. 1, pp. 1–5, doi: 10.3997/2214-4609.201901386.

[26] J. Wang and J. X. Cao, "Data-driven S-wave velocity prediction method via a deep-learning-based deep convolutional gated recurrent unit fusion network," *Geophysics*, vol. 86, no. 6, pp. M185–M196, 2021, doi: 10.1190/geo2020-0886.1.

[27] L. Song, X. Y. Yin, Z. Y. Zong, and M. Jiang, "Semi-supervised learning seismic inversion based on spatio-temporal sequence residual modeling neural network," *J. Petroleum Sci. Eng.*, vol. 208, no. Part D, 2022, Art. no. 109549, doi: 10.1016/j.petrol.2021.109549.

[28] H. H Zhang, G. Z. Zhang, J. H. Gao, S. J. Li, J. M. Zhang, and Z. Y. Zhu, "Seismic impedance inversion based on geophysical-guided cycle-consistent generative adversarial networks," *J. Petroleum Sci. Eng.*, vol. 218, 2022, Art. no. 111003, doi: 10.1016/j.petrol.2022.111003.

[29] A. Mustafa, M. Alfarraj, and A. Ghassan, "Estimation of acoustic impedance from seismic data using temporal convolutional network," in *Proc. SEG Int. Expo. Annu. Meeting*, 2019, Art. no. 5407, doi: 10.1190/segam2019-3216840.

[30] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141, doi: 10.1109/CVPR.2018.00745.

[31] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-net: Efficient channel attention for deep convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 11531–11539, doi: 10.1109/CVPR42600.2020.01155.

[32] B. Y. Wu, Q. Xie, and B. H. Wu, "Seismic impedance inversion based on residual attention network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4511117, doi: 10.1109/TGRS.2022.3193563.

[33] A. Vaswani et al., "Attention is all you need," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 6000–6010, doi: 10.48550/.arXiv.1706.03762.

[34] N. Parmar et al., "Image transformer," in *Proc. 35th Int. Conf. Mach. Learn.*, 2018, pp. 4055–4064, doi: 10.48550/.arXiv.1802.05751.

[35] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 213–229, doi: 10.48550/.arXiv.2005.12872.

[36] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," in *Proc. Int. Conf. Learn. Representations*, 2021, pp. 1–22, doi: 10.48550/.arXiv.2010.11929.

[37] Z. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 9992–10002, doi: 10.1109/ICCV48922.2021.00986.

[38] K. Han, A. Xiao, E. H. Wu, J. Y. Guo, C. J. Xu, and Y. H. Wang, "Transformer in transformer," in *Proc. 35th Conf. Neural Inf. Process. Syst.*, 2021, vol. 34, pp. 15908–15919, doi: 10.48550/.arXiv.2103.00112.

[39] Z. Q. Sun, S. C. Cao, Y. M. Yang, and K. Kitani, "Rethinking transformer-based set prediction for object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 3611–3620, doi: 10.1109/ICCV48922.2021.00359.

[40] D. Zhang, H. W. Zhang, J. H. Tang, M. Wang, X. S. Hua, and Q. R. Sun, "Feature pyramid transformer," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 323–339, doi: 10.48550/.arXiv.2007.09451.

[41] S. X. Zheng et al., "Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 6881–6890, doi: 10.1109/CVPR46437.2021.00681.

[42] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "SegFormer:Simple and effioient design for semantic segmentation with transformers," in *Proc. 35th Conf. Neural Inf. Process. Syst.*, 2021, pp. 12077–12090, doi: 10.48550/.arXiv.2105.15203.
[43] R. Strudel, R. Garcia, I. Laptev, and C. Schmid, "Segmenter: Transformer for semantic segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 7262–7272, doi: 10.48550/.arXiv.2105.05633.
[44] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Effificient estimation of word representations in vector space," in *Proc. Int. Conf. Learn. Representations*, 2013, doi: 10.48550/.arXiv.1301.3781.
[45] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Proc. 27th Conf. Neural Inf. Process. Syst.*, 2013, doi: 10.48550/.arXiv.1310.4546.
[46] Y. H. Liu, E. Sangineto, W. Bi, N. Sebe, B. Lepri, and M. D. Nadai, "Efficient training of visual transformers with small-size datasets," in *Proc. 35th Conf. Neural Inf. Process. Syst.*, 2021, pp. 23818–23830, doi: 10.48550/.arXiv.2106.03746.
[47] A. Gulati et al., "Conformer: Convolution-augmented transformer for speech recognition," in *Proc. Interspeech Conf.*, 2020, pp. 5036–5040, doi: 10.48550/.arXiv.2005.08100.
[48] A. Brougois, M. Bourget, P. Lailly, M. Poulet, P. Ricarte, and R. Versteeg, "Marmousi, model and data," Practical Aspects Seismic Data Inversion, cp-108-00002, 1990, doi: 10.3997/2214-4609.201411190.
[49] Z. Chen, V. Badrinarayanan, C. U. Lee, and A. Rabinovich, "GradNorm: Gradient normalization for adaptive loss balancing in deep multitask networks," in *Proc. 35th Int. Conf. Mach. Learn.*, 2018, pp. 793–802, doi: 10.48550/.arXiv.1711.02257.
[50] D. P. Cao, Y. Q. Su, and R. G. Cui, "Multi-parameter pre-stack seismic inversion based on deep learning with sparse reflection coefficient constraints," *J. Petroleum Sci. Eng.*, vol. 209, 2022, Art. no. 109836, doi: 10.1016/j.petrol.2021.109836.

Chunyu Ning received the bachelor's degree in statistics from Shangdong University, Weihai, China, in 2021. She is currently working toward the master's degree in applied statistics with Xi'an Jiaotong University, Xi'an, China.

Her research interests include deep learning for seismic inversion.

Bangyu Wu (Member, IEEE) received the B.S. degree in information engineering and the Ph.D. degree in information and communication engineering from Xi'an Jiaotong University, Xi'an, China, in 2005 and 2012, respectively.

From 2007 to 2011, he was a Visiting Scholar with the University of California at Santa Cruz, Santa Cruz, CA, USA. He also worked as a (Senior) Geophysicist at Statoil (Beijing) Technology Service Company Ltd., Beijing, China, from 2012 to 2015. He is currently an Associate Professor with the School of Mathematics and Statistics, Xi'an Jiaotong University. His research interests include seismic wave modeling and migration/inversion, signal processing, and machine learning.

Baohai Wu received the B.Sc. degree in geophysics from the Department of Geophysics, China University of Geosciences, Wuhan, China, in 2008, and the M.S. degree in geodetection and information technology from the Research Institute of Petroleum Exploration and Development, Beijing, China, in 2012.

He is currently a Technical Advisor with CGG GeoSoftware, Beijing. His research interests include rock physics, seismic inversion, and machine learning in geophysics.