Attention and Hybrid Loss Guided 2-D Network for Seismic Impedance Inversion

Qiao Xie, Bangyu Wu 🆔, *Member, IEEE*, and Yueming Ye

Abstract—Deep learning methods, especially convolutional neural networks, achieve state-of-the-art performance on seismic impedance inversion. Most of the methods are based on one-dimensional (1-D) convolution, tending to yield lateral discontinuities of impedance on field data applications. To alleviate this problem, we design a network equipped with 2-D convolutions and a coordinate attention (CA) block. The former can take the relationship between adjacent traces into consideration. The latter can capture the positional relationship of the geological structure, both horizontally and vertically. At the same time, we use a hybrid loss combined with an edge operator and mean square error to further improve the stability of the designed network. Comparison experiments on the synthetic SEAM model and field seismic data demonstrate the effectiveness of the adopted components, 2-D convolution, CA, and hybrid loss function in improving the lateral continuity of inverted impedance. For field seismic data, the impedance predicted by the proposed method shows improved lateral continuity and high resolution compared with the 1-D network and constrained sparse spike inversion method using commercial software (InverTrace Plus module in Jason).

Index Terms—Coordinate attention (CA), hybrid loss, seismic impedance inversion, two-dimensional (2-D).

I. INTRODUCTION

SEISMIC inversion is a crucial task in geological interpretation to speculate the physical parameter and spatial distribution of underground stratigraphic structures [1]. However, various noises and the unknown wavelet (usually unavailable and difficult to estimate) cause uncertainty in the inversion results. Predicting impedance from seismic is an ill-posed nonlinear problem, resulting in nonunique seismic impedance solutions and unstable results on real data applications [2]. Methods based on deep learning (DL) provide new problem-solving opportunities due to their powerful feature learning and computing capabilities [3]. Using nonlinear operators such as activation functions, DL maps the strong nonlinearity between impedance and seismic data in the network. Furthermore, the data-driven DL method, unlike the conventional model-driven method, directly learns from seismic data to impedance mapping without the assumption of an approximate forward model [4]. The amount of available seismic data increased exponentially, reaching 667.7 trillion bytes by September 2020 [5]. Dealing with Big Data with a huge number of parameters, DL has developed into a promising method to cope with different types of geophysical problems [5], [6], [7], [8], [9], [10]. In recent years, many DL networks have been proposed for seismic impedance inversion, such as convolutional neural network (CNN) [11], fully convolutional residual network [12], constitutional neural network [13], recurrent neural network [14], [15], generative adversarial networks (GANs) [16], [17], [18], and so on. All of these DL methods achieve outstanding results. However, for practical implementation, the results need to conform to a certain degree of geological priors, one of which is lateral continuity.

The convolution seismic data model states that a seismic trace is the convolution of the earth's reflectivity converted from the impedance with the source wavelet [19]. Poststack seismic data and impedance, serving as inputs and labels for data-driven methods, are corresponding trace by trace. Since a sequence of seismic data corresponds to an impedance sequence and there is a specific functional relationship between them, most DL methods train a model with one-dimensional (1-D) convolution [12], [16], [17]. To succeed in DL, it is important to provide more training examples than free parameters in deep networks with huge parameter space [20]. However, due to drilling costs, the number of available wells in a 3-D seismic survey is usually very limited. The 1-D algorithm using sequences with limited training data is a challenge to yield stable and accurate impedance estimation. However, the actual stratum is spatially continuous, and there is a strong correlation between seismic adjacent traces which are ignored by 1-D networks, such as lateral continuity. Many 1-D networks are also incapable of capturing sudden changes in the rock properties in the complex geological structure due to the limited training well logs [21]. The increasing number of publications [4], [22], [23] and the industry are keeping an eye on the 2-D network implementations for seismic impedance inversion which shows a positive trend in this field. In this article, we use 2-D convolution for training, which enhances the continuity with the help of stratum structure correlation.

The continuity of geological structure, both horizontal and vertical, means that the structures between adjacent traces are similar and related. Nevertheless, convolution operations can only capture local relations and fail to model long-range dependencies [7], [24]. To mitigate this problem, we introduce a

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

Yueming Ye is with the PetroChina Hangzhou Research Institute of Geology, Hangzhou 310023, China (e-mail: yeym_hz@163.com).
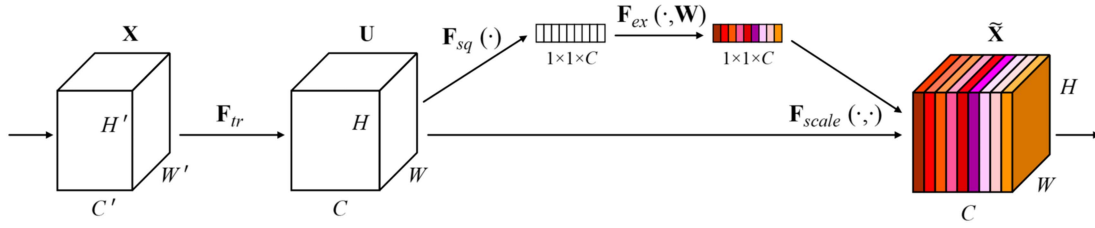
Fig. 1.   Squeeze-and-excitation block [38].

coordinate attention (CA) block to capture spatial correlation in seismic data. The CA block captures direction-specific information along each spatial direction. The network with a CA block can acquire not only cross channel but also direction-aware and position-sensitive information. This also improves network accuracy by locating and emphasizing the features of interest. Meanwhile, CA blocks can alleviate the loss of positional information in 2-D global pooling [7], [24].

Essential information exists on the edge of the image where local features of the image manifest discontinuity, namely, the structure changes violently. Edge detectors are a significant part of many computer vision systems to obtain useful structural information from image contours. The uncertainties such as potential ambient noise, acquisition limitations, and processing errors in the real seismic data make DL models difficult to describe the recorded seismic data properly and directly. In this article, we introduce an edge detection operator into loss function. By minimizing the feature distribution divergence, the structure feature distribution of predicted impedance can better match the true value feature. Functioning as a physical constraint, the obtained edge information can facilitate more reasonable inversion results, that is, better lateral continuity and less noise impact for practical application [25].

In summary, we propose a 2-D CNN equipped with a CA block using a hybrid loss of edge operator for seismic impedance inversion. In Section II, attention mechanisms including the CA block, the network structure, the hybrid loss function, and transfer learning are described in detail. In Section III, two ablation studies are conducted. We also validate the effectiveness of 2-D convolution, CA and hybrid loss function and the advantages of the proposed method on synthetic and field datasets. Discussions and conclusions are given in Sections IV and V, respectively.

II. METHODOLOGY

A. Attention Mechanism

Attention mechanisms [26], [27] have promoted various computer vision tasks, such as image classification [28], [29], [30], and image segmentation [31], [32], [33], over recent years. They are also beneficial in dealing with geophysical problems [34], [35], [36], [37]. Successful attention implementations include SENet [38], CBAM [29], GENet [30], AA [31], and self-attention [39]. Wu et al. [40] demonstrated the effectiveness of a multibranch attention block combining SENet [38] and SKNet [41] and designed a new attention block for seismic

impedance inversion. SE block and SK block are shown in Figs. 1 and 2, respectively. Tsotsos [26] models the channel relationship in the network and [41] captures the feature-map relationship with a multibranch nonlinear combination. The proposed ResANet [40] outperforms several comparable neural networks in accuracy and generalization ability while ensuring efficiency for seismic data impedance inversion. However, the 1-D algorithm is challenging to yield a stable impedance estimation. To enhance the lateral continuity, we utilize a 2-D convolution network that incorporates two-direction information and captures the positional information by introducing CA [24] on the 2-D model with SE attention and SK attention expanded from 1-D ResANet [40].

Different from the aforementioned attention approaches, CA captures positional information and channelwise relationships to strengthen the feature representations efficiently [24]. The two spatial directions can effectively make use of the geological structure and suppress unreasonable information, such as noise. In this way, CA can enhance the model's ability in tasks to deal with complex geological scenarios. When it comes to handling large-size datasets, DL models may suffer from memory consumption problems. CA improves the performance of various models with nearly no computational overhead.

Fig. 3 shows the architecture of a CA block. First, CA aggregates the input (X) with two parallel 1-D global pooling operations to obtain separate information along two directions, written as

$$z_c^h\ (h) = \frac{1}{W} \sum_i x_c\ (h,\ i)\,, 0 \le i \le W \qquad (1)$$

$$z_c^w\ (w) = \frac{1}{H} \sum_j x_c\ (j,\ w)\,, 0 \le j \le H \qquad (2)$$

where z_c^h and z_c^w are the output of the cth channel at height h and width w, respectively. Second, the two feature maps are first concatenated and then activated using a 1×1 convolutional function to produce an intermediate feature map $(\hat{z})$. Split $\hat{z}$, two separate tensors $(\hat{z}^h$ and $\hat{z}^w)$, with direction-specific information are generated. To activate the two attention maps, $\widehat{X}_h$ and $\widehat{X}_w$ are calculated by

$$\widehat{X}_h = \sigma\left(G_h\left(\hat{z}^h\right)\right) \qquad (3)$$

$$\widehat{X}_w = \sigma\left(G_w\left(\hat{z}^w\right)\right) \qquad (4)$$

where σ is the sigmoid function and G denotes 1×1 convolutional transformation. Finally, both attention maps are utilized as attention weights and multiplied by the input feature map to
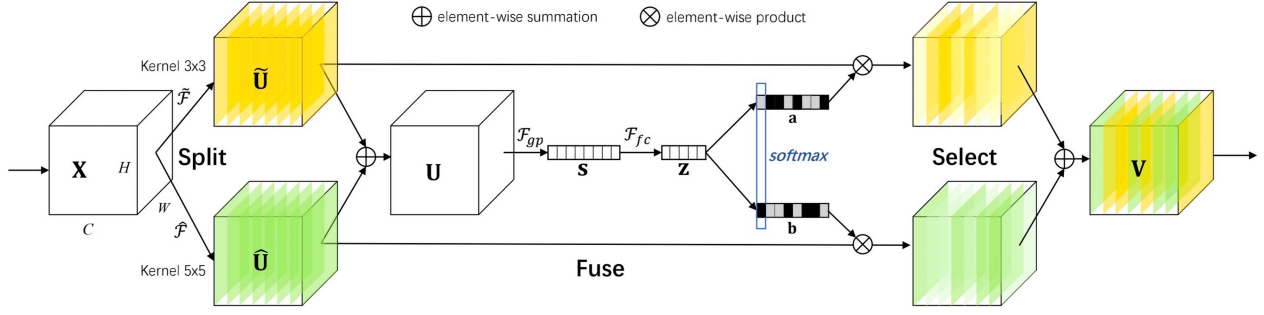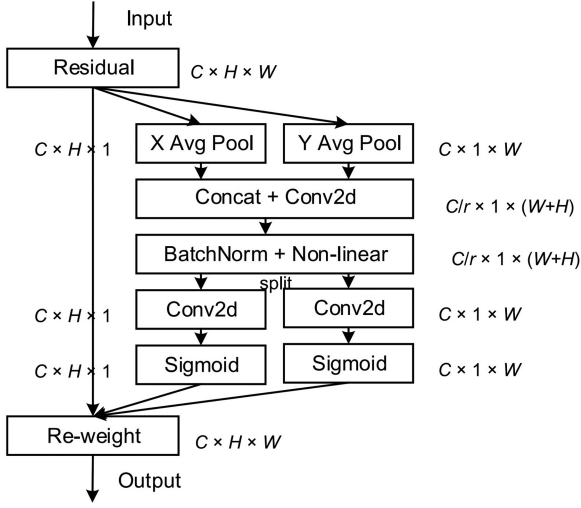
Fig. 2. Selective kernel convolution [41].

Fig. 3. CA block [24].

emphasize the feature expression, yielding

$$\widehat{X} = x_c\ (i,\ j) \times \widehat{X}_c^h\ (i) \times \widehat{X}_c^w\ (j) \tag{5}$$

where $\widehat{X}$ is the output of the CA block.

In the experiment part, we test appropriate positions of CA in the network on synthetic data, SEAM model. To demonstrate the advantages of CA, a series of experiments are conducted with/without CA modules on both synthetic and field data.

B. Network Architecture

In this article, we design a 2-D ResANet network with CA for impedance inversion, called 2-D CA-ResANet. The specific structure of the proposed model is shown in Fig. 4. The 2-D CA-ResANet consists of three parts: the input layer, the attention part, and the output layer. The attention part contains three stacked branch attention blocks (red dash-dot box in Fig. 4) and a CA block. The input data first go through a convolutional layer and a dropout layer with a 0.2 dropout probability. The former enables the model to capture the seismic data's low-level features. The latter is utilized to alleviate the problem of overfitting. Then, three branch blocks refine the information to form high-level information. Specifically, four convolution branches with (convolution kernel size, dilation) parameter pairs as (K×3,

1), (K×3, 2), (K/2×3, 1), and (K/2×3, 2) are applied to get the multiscale information. K is the width of the convolution kernel, which is related to the wavelength of the source wavelet [11]. Thus, we set different values for K on the synthetic and field model according to the length of a seismic trace after simple experiments. To extract the information, each convolutional branch is followed by SE operation. A convolution layer with batch normalization (BN) [42] and rectified linear unit (ReLU) [43] function nonlinearly aggregates information from multiple convolution kernels. Then, the global information is controlled by SE. After the stacked branch part, CA block is used to aggregate global information according to correlations. The last layer comprises a convolution layer and ReLU function for regression. The residual block is embedded into these attention modules to obtain stable deep networks. To magnify the generalization ability of the model and accelerate the network training, BN is applied after each convolutional layer except the last layer.

C. Network Training

Estimating an impedance sequence from a seismic trace using data-driven methods is a regression problem. Mean square error (MSE) loss function is used to measure the error of each pixel between the predicted data and the target data. Networks with a single MSE loss function are prone to be influenced by noises in the real seismic data, such as potential ambient noise, acquisition errors, and processing errors which can result in the irrationality of seismic data structure. To further enhance the lateral continuity of inversion results and antinoise performance, we add an edge detector operator based on MSE loss function. The hybrid loss function is introduced in detail as follows.

1) *MSE Loss Function*: The MSE loss function is defined as

$$L_{\mathrm{MSE}} =\ E\left[\|\ y - y'\ \|_F\right] \tag{6}$$

where E is the mathematical expectation, y denotes the ground truth, and y' represents the prediction result calculated from seismic data.

2) *Sobel operator*

The edge intensity is calculated by edge detection operators according to the gradient of the image. Specifically, edges correspond to a change of pixels' intensity and the value is acquired using the value of pixels $[i, j]$ and their neighbors. In this case, the intensity change in both directions, $I_h(X)$ (horizontal) and
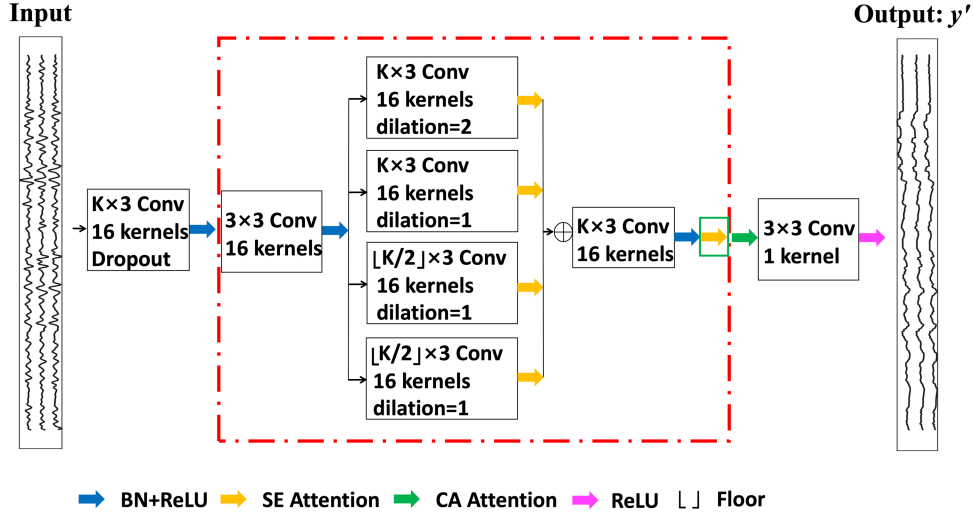
Fig. 4.   Architecture of the 2-D CA-ResANet for seismic impedance inversion. K is 599 in SEAM model and 29 in the field data experiments.

$I_v(X)$ (vertical), is computed as

$$I_h\left(X\right) = X\left[i, j+1\right] - X\left[i, j\right] \tag{7}$$

$$I_v\left(X\right) = X\left[i+1, j\right] - X\left[i, j\right] \tag{8}$$

where $X[i, j]$ denotes the value of the corresponding pixel $[i, j]$ of image X. Then, the magnitude G of the gradient is generated by

$$G\left(X\right) = \sqrt{I_h(X)^2 + I_v(X)^2}. \tag{9}$$

And the edge operator loss L_{EDGE} is obtained by measuring the error between the magnitude of the predicted data $G(y')$ and the target data $G(y)$ written as

$$L_{EDGE} = E\left[\| G\left(y\right) - G\left(y'\right) \|_F\right]. \tag{10}$$

Finally, the overall objective function in this article is formulated as

$$L_{Total} = \lambda_1 L_{EDGE} + \lambda_2 L_{MSE}, \tag{11}$$

where λ_1 and λ_2 are weights to balance the objective terms, and the sum of the weighting coefficients is 1.

D. Transfer Learning

Transfer learning is a machine learning method, where the model developed for a task is reused as the starting point of a second task model. Specifically, transfer learning uses the related tasks learned in one setting to improve generalization in another setting through knowledge transfer [44]. It can be regarded as an optimization that allows rapid progress or improved performance when modeling the second task [45]. Transfer learning is popular in computer vision, natural language processing, geophysical practical applications, and other fields that need huge resources. Through transfer learning, the developed neural network models on these issues can save vast computing resources and turnaround time and also make a huge performance improvement on related tasks. Here, we use the pretraining method, which is commonly used in the field of DL. First, the source model is trained on basic datasets and tasks. Then, the input–output pair data available for the task of interest is used to train on the target dataset and task to repurpose the learned features or transfer them to a second target network [46]. In this article, we fine-tune the model trained on the synthetic dataset using interpolated data around the wells. The interpolated seismic data and impedance are more accurate than data far away from the wells and can be regarded as augmented data to train the neural networks.

III. EXPERIMENTS

A. Experiment on Synthetic Seismic Dataset

To quantify the effect of the CA block and edge detector, we first test on open SEAM dataset, which is widely used in DL inversion methods [47], [48]. The seismic data are generated by convolving reflectivity with a 30 Hz zero-phase Ricker wavelet. As shown in Fig. 5, the seismic data and impedance both have 1751 traces with 5001 time points and 4 ms time interval. To train the model, we choose 102 trace pairs with equal intervals from the impedance and seismic profile. Considering the lateral structural relationship, the left and right adjacent traces of each trace are also selected for 2-D network training. The (16, 3, 5001) are randomly selected as validation sets where 16 is the number of trace pairs, 3 denotes the adjacent 3 traces, and 5001 represents the time points. We use Adam optimization [49] with an initial learning rate of 0.001 and weight decay 1×10^{-7} as the optimization algorithm. Batch size is set to 10. Kaiming Initialization [50] is chosen to initialize network weights and train 1000 epochs.

Two brief ablation studies are first conducted to obtain appropriate weightings in loss function and network structure, respectively. Table I lists the results of predicting impedance, and the lowest MSE (highlighted in bold red) is obtained when λ_1 is 0.3. As mentioned above, CA can obtain spatial information and integrate global information. Therefore, we consider two
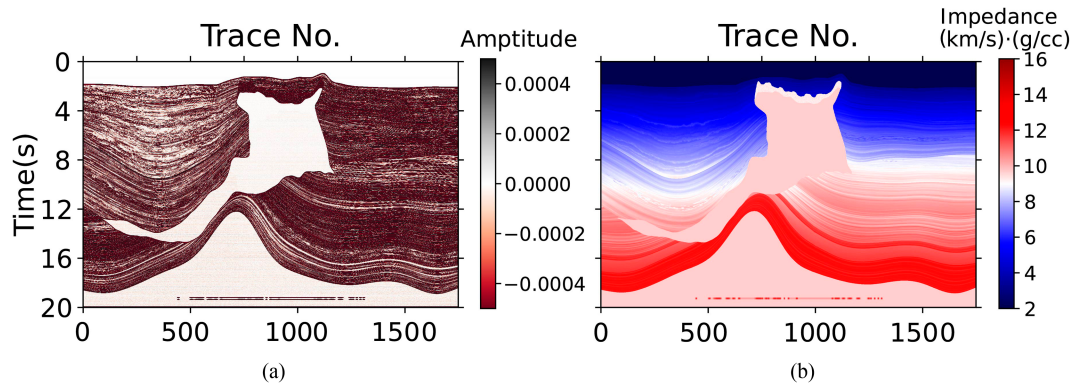
Fig. 5. SEAM model. (a) Synthetic seismic data. (b) Impedance.

TABLE I
SEAM MODEL PREDICTION RESULTS WITH DIFFERENT WEIGHTS IN LOSS FUNCTION

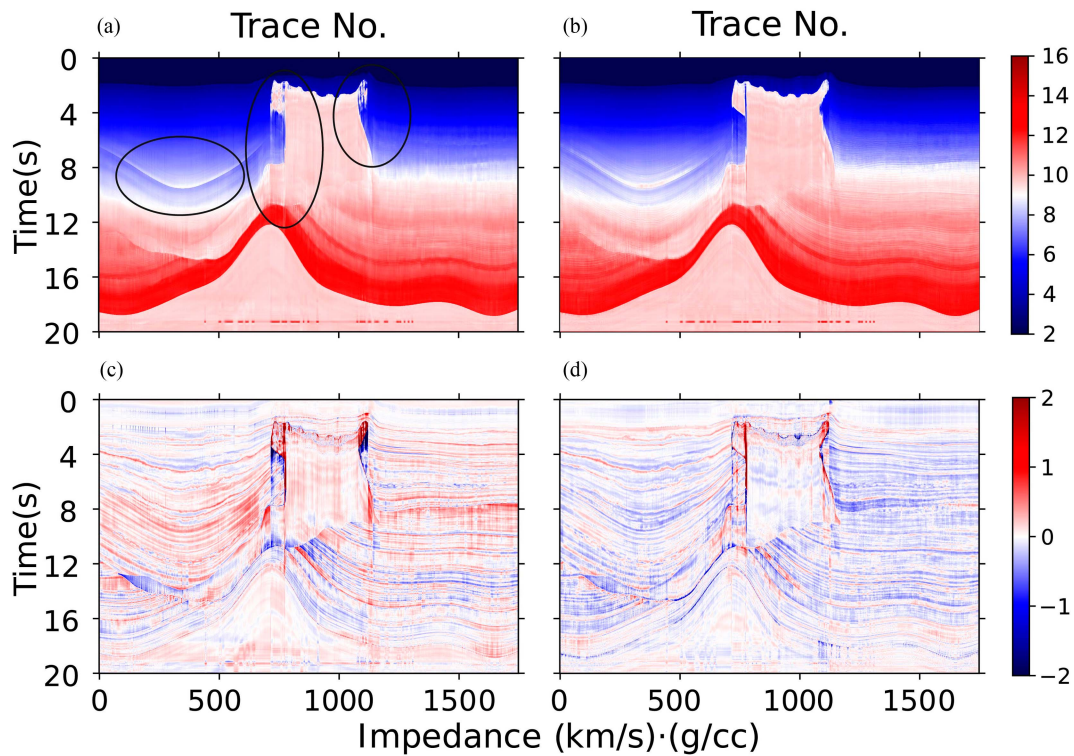| Weight coefficient (λ_1)	0	0.3	0.5	0.7	1
MSE	0.1036	**0.0672**	0.0682	0.0750	0.1010

Fig. 6. Impedance prediction results by networks with CA at different positions. (a) CA in the branch attention block. (b) CA after the branch attention block. (c), (d) Corresponding residual profiles between predictions and ground truth [see Fig. 5(b)].

positions to put CA. One is to replace SE operation at the end of the stacked branch attention block shown in the green box in Fig. 4. The other, as shown in Fig. 4, follows after the stacked branch attention block. Fig. 6 shows the prediction results of the networks with CA at the aforementioned positions, respectively. The first column is the result predicted by the network with CA in the branch attention block and the MSE of the prediction

profile is 0.0621. The second column is the result using the network in Fig. 4. And the profile MSE is 0.0421, which is significantly lower than putting CA inside the attention block. Fig. 6 (especially the location inside the black ovals) and the MSEs illustrate that the network with CA in Fig. 4 achieves better performance. As a result, the network shown in Fig. 4 with hybrid loss setting weighting λ_1 as 0.3 is used in the following
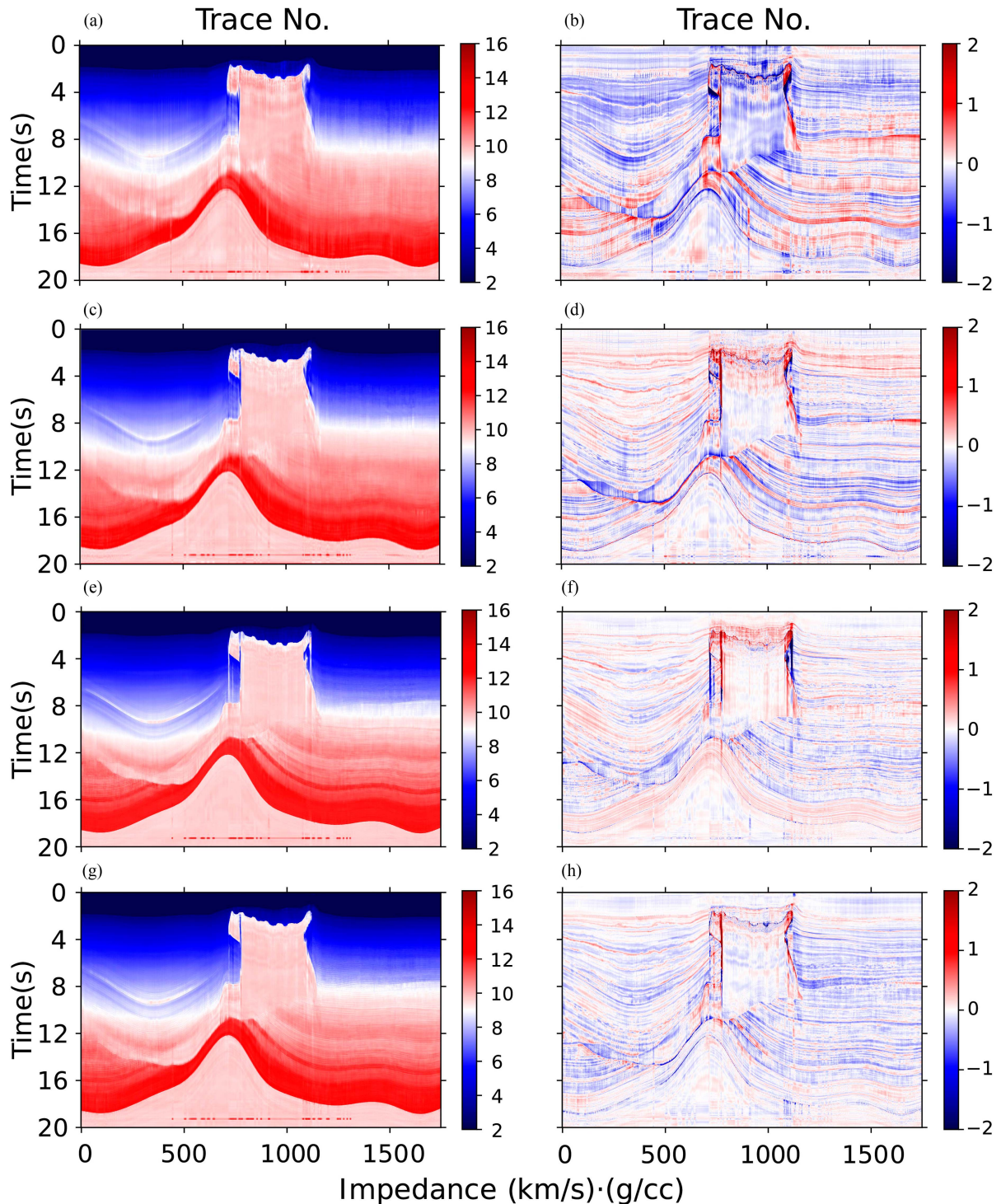
Fig. 7. Impedance prediction results of 2-D models. (a) ResANet ($\lambda_1 = 0$). (c) ResANet ($\lambda_1 = 0.3$). (e) CA-ResANet ($\lambda_1 = 0$). (g) CA-ResANet ($\lambda_1 = 0.3$). (b), (d), (f), (h) Corresponding residual profiles between predictions and ground truth [see Fig. 5(b)].

experiments.

To further verify the effect of the model and loss, we carry out a set of experiments on 2-D ResANet. "2-D" is omitted in the name of the four models in the following synthetic experiments to be concise. Fig. 7 and Table II list the results of the whole section with four experiments. We first predict the impedance using ResANet and the result is shown in Fig. 7(a). Then, λ_1

is adjusted to 0.3 [see Fig. 7(c), the third column of Table II] to test the hybrid loss effect compared with Fig. 7(a). Similarly, Fig. 7(e) and the fourth column of Table II list the result of CA-ResANet, which can testify CA effect. When only λ_1 is set to 0.3 or only CA is added, the overall prediction is optimized to varying degrees in comparison with 2-D ResANet [see Fig. 7(a), the second column of Table II]. Fig. 7(g) shows a better result in

TABLE II
2-D MODELS COMPARISON BY MSES ON THE WHOLE SEAM SECTION

Model (Weight coefficient)	ResANet ($\lambda_1 = 0$)	ResANet ($\lambda_1 = 0.3$)	CA-ResANet ($\lambda_1 = 0$)	CA-ResANet ($\lambda_1 = 0.3$)
MSE	0.1036	0.0672	0.0509	**0.0421**

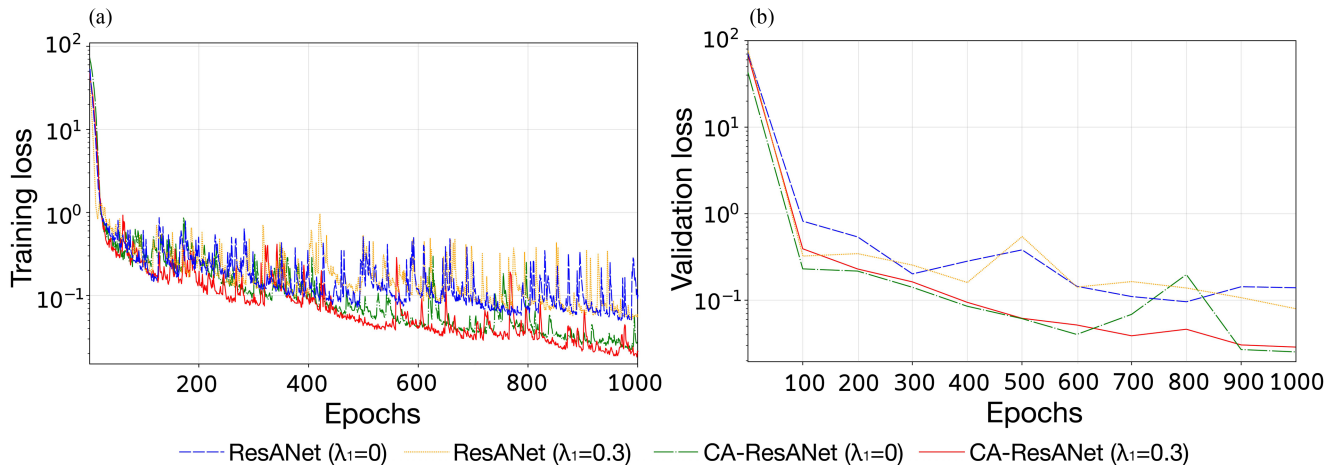
Fig. 8. Training and validation loss curves of ResANet ($\lambda_1 = 0$), ResANet ($\lambda_1 = 0.3$), CA-ResANet ($\lambda_1 = 0$), and CA-ResANet ($\lambda_1 = 0.3$). (a) Training. (b) Validation.

the middle of SEAM model by adding the edge operator in the loss function than Fig. 7(e), where the structure laterally changes violently. Through the comparison in Fig. 7 and Table II, it can be seen that CA-ResANet with hybrid loss [see Fig. 7(g), the last column of Table II] shows the most accurate prediction result.

Fig. 8 shows the training and validation curves (verified every 100 epochs to save training time) for the SEAM synthetic model test. CA-ResANet with hybrid loss has the best training performance shown in Fig. 8(a). The curves of the other three methods are all above the proposed network (red), which presents inferior performance. Fig. 8(b) shows that the validation curves of the three methods, ResANet with or without hybrid loss and CA-ResANet with MSE loss, tend to rise suddenly. In contrast, the validation loss curve of CA-ResANet with hybrid loss shows a stable decline. From Fig. 8(b), we can see that both of CA-ResANet perform better than the other models. Between them, although the model with only MSE loss has the lowest MSE with a small margin, the model with hybrid loss predicts more precisely on the whole section given in Table II, which may be caused by overfitting for using solely the MSE loss function.

We further conduct experiments on seismic data contaminated by five different levels of Gaussian white noise to demonstrate the robustness of the proposed method. Specifically, the model is trained with seismic data, which is noise free, and then we predict impedance using data with different signal-to-noise ratios (SNR). The SNRs in the synthetic data of SEAM model are 5, 15, 25, 35, and 45 dB, respectively, as shown in Fig. 9(a)–(e). To avoid accidental errors, the entire experimental procedure is repeated ten times. We predict the impedance profile by the proposed method when the inputs are noisy seismic data from the first row in Fig. 9 and average the ten MSEs calculated using

predictions and true impedances. Fig. 9(f)–(j) presents one of the impedance profiles predicted by the proposed method when the inputs are noisy seismic data in the first row. The third row in Fig. 9 is the residual profile between the prediction of different SNR seismic data and the ground truth [see Fig. 5(b)]. The average MSEs are provided in Table III. Apart from the result predicted from seismic data at an SNR of 5 dB, we can see that there is no significant difference between these prediction results, which are slightly worse than the prediction impedance on data without noise (see Fig. 7(g) and the fifth column of Table II). This presents that the robustness of the model tends to be stable after SNR is higher than 15 dB. Even on seismic data with 5 dB SNR shown in Fig. 9(a), we can see that our method shows satisfying prediction in Fig. 9(b). These illustrate that the model has the ability to deal with field data, which always contains noise to some degree and shows strong robustness against random noise when SNR reaches 15 dB.

B. Experiment on Field Data

After determining the model and loss through the synthetic data test, we validate the performance of 2-D CA-ResANet with the hybrid loss on a 3-D field dataset shown in Fig. 10. The 3-D poststack field seismic volume is from the northern Gulf of Mexico off the southern coast of Louisiana with a turbidite sedimentary target layer. The seismic data have 1501 time points with 2 ms time interval, and the target layer is between 2244 and 2494 ms. Fig. 11 shows a time slice of the seismic data and the locations of six wells. To demonstrate the performance of the proposed method, we predict the impedance of the cross-well section indicated by the black solid line on a
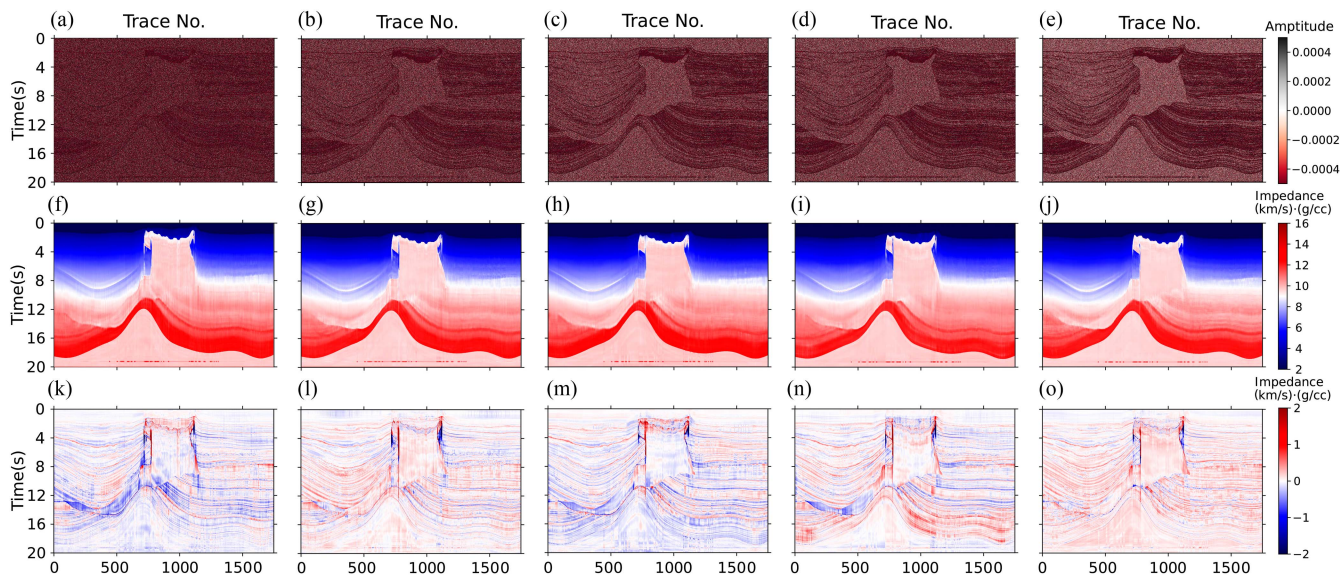
Fig. 9. Seismic data with Gaussian white noise and the impedance prediction results on SEAM model. Seismic data with different SNR. (a) 5 dB. (b) 15 dB. (c) 25 dB. (d) 35 dB. (e) 45 dB. (f)–(j) Corresponding impedance prediction results by 2-D CA-ResANet ($\lambda_1 = 0.3$). (k)–(o) Corresponding residual profiles between predictions and ground truth [see Fig. 5(b)].

TABLE III
2-D MODEL COMPARISONS BY MSEs OF THE WHOLE SEAM SECTION

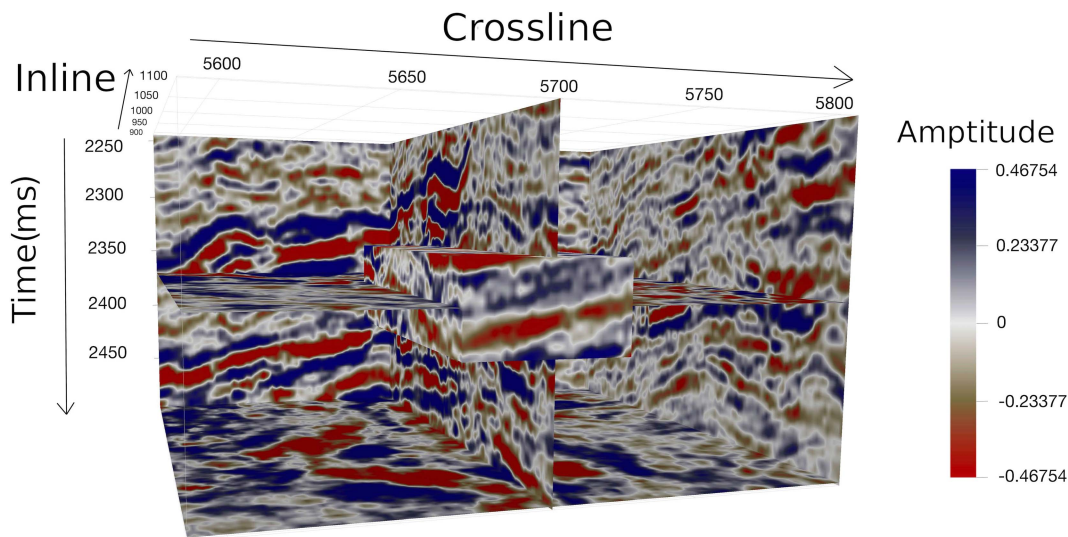| SNR	5 dB	15 dB	25 dB	35 dB	45 dB
MSE	0.09811	0.05416	0.05384	0.05376	0.05461

Fig. 10. 3-D field seismic volume.

time slice in Fig. 11. Fig. 11 shows the selected trace locations (black crossings) for fine-tuning on the interpolated seismic data and impedance profile. W1 and the selected traces around W1 are locally enlarged in the lower-left corner shown in the red box for detail inspection. Specifically, we select 48 interpolated trace pairs centered on each well and combine them with three adjacent traces, as shown in the red oval for fine-tuning.

Fig. 12 shows the impedance and seismic data on the cross-well section we used in this article. Since we have no intention to improve seismic resolution, the well logs were processed to facilitate model training. A low-pass filter is used to cut off the high-frequency components for better comparison with the inversion results of the seismic frequency band. We first train the models on the interpolated impedance [see Fig. 12(a)] and
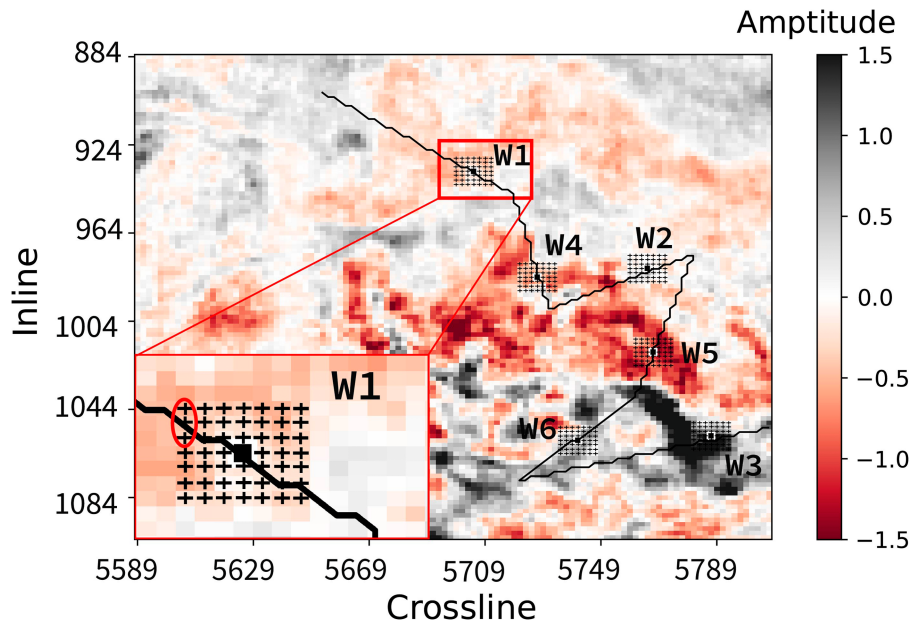
Fig. 11. Time slice of the field seismic data. The black squares are the well locations from W1 to W6. The solid black line is the cross-well profile used to demonstrate the impedance prediction test on this dataset. The red box region in Fig. 11 is locally enlarged in the lower-left corner for detail inspection.
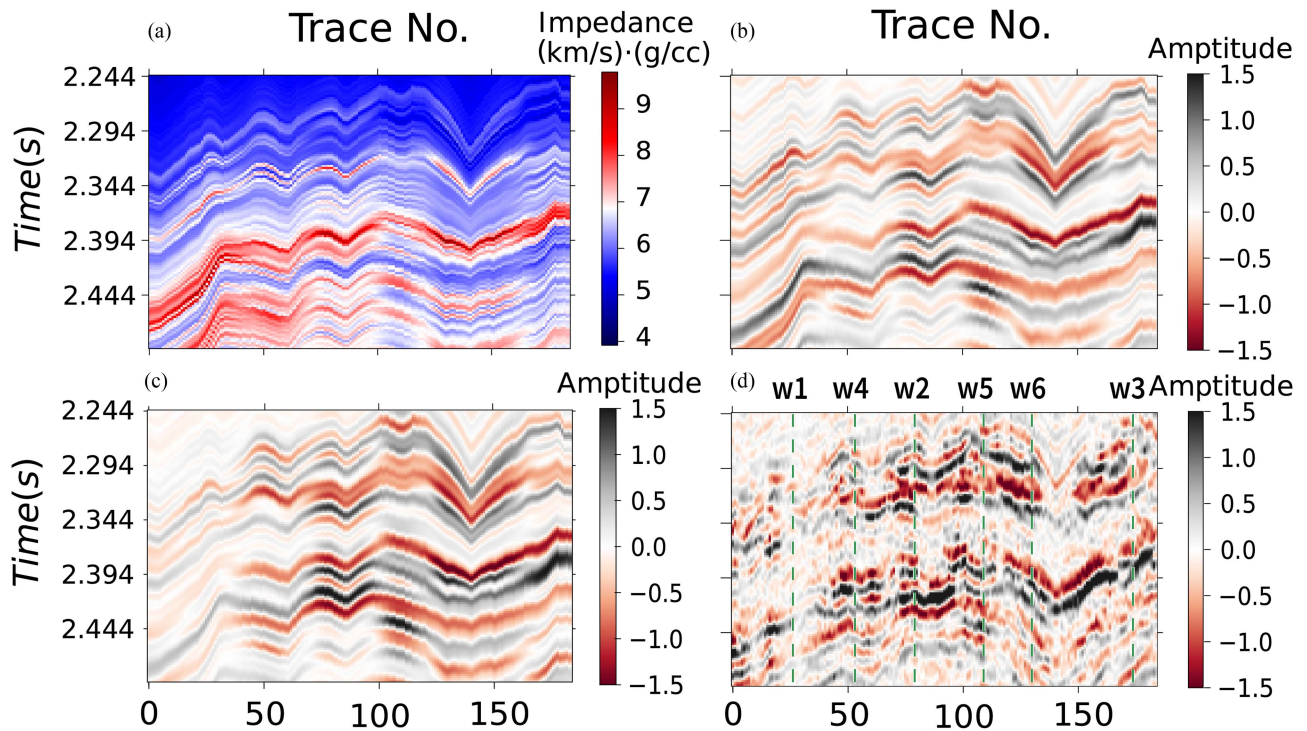

Fig. 12. Cross-well section of impedance and seismic data. (a) Interpolated impedance. (b) Synthetic seismic data. (c) Interpolated seismic data. (d) Field seismic data. The vertical solid green line in (d) stands for the position of the six wells (W1–W6).

synthetic seismic data [see Fig. 12(b)] generated from the interpolated impedance to learn the theoretical mapping relationship. Then, the pretrained models are fine-tuned with the interpolated seismic data [see Fig. 12(c)] and the interpolated impedance around the wells to fit the actual geological features. Finally, we predict the impedance from cross-well profile of the field

seismic [see Fig. 12(d)] to verify the effect of the proposed method.

To demonstrate the effectiveness of the proposed method, we conduct tests by adjusting the convolution kernel dimension, loss, and CA. The results are shown in Fig. 13. In total, 144 seismic and impedance pairs around the 6 wells (24 wells each) from
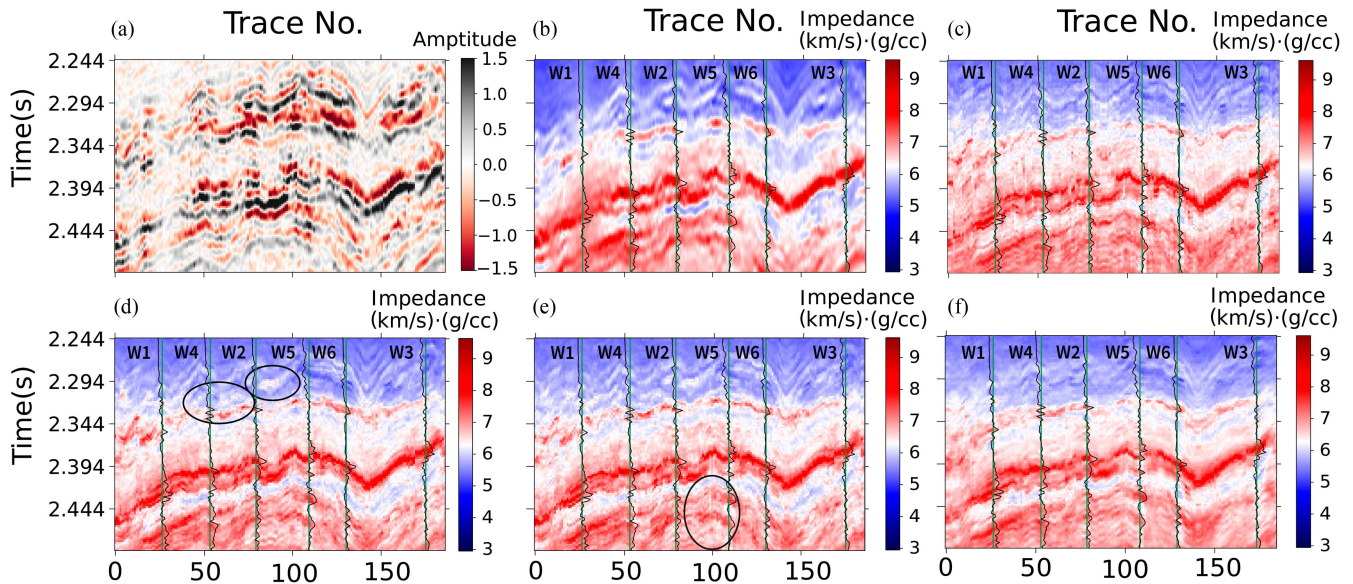
Fig. 13.   Impedance prediction results using field seismic data (a), (b) CSSI. (c) 1-D ResANet. (d) 2-D ResANet ($\lambda_1 = 0$). (e) 2-D ResANet ($\lambda_1 = 0.3$). (f) 2-D CA-ResANet ($\lambda_1 = 0.3$). The vertical solid green line and black curve stand for the position and the impedance of the six wells (W1–W6), respectively.

interpolated data are used to fine-tune the 1-D ResANet. These 2-D models use 102 sets of traces with 125 time points for fine-tuning. Fig. 13(d) shows that the overall visual lateral continuity of the 2-D ResANet has been improved compared with the 1-D ResANet [see Fig. 13(c)] containing some fragmentation layers (shown in the black oval). After adding CA [see Fig. 13(e)], those anomalies are suppressed. Fig. 13(f) shows that not only the fragments but also the vertical strips [the black oval in Fig. 13(e)] are further reduced. Fig. 13 illustrates that CA-ResANet with the hybrid loss can predict impedance with better lateral continuity and fewer fragments correspondingly. In addition, the result of the constrained sparse spike inversion (CSSI) [51] method [see Fig. 13(b)] from commercial software (InverTrace Plus module in Jason) is also used as a criterion when these predictions are compared. For CSSI, the resolution of inverted P-impedance is a tradeoff on many factors. The improvement of vertical resolution depends not merely on the input seismic wavelet and seismic data but on the inversion parameters. For example, seismic SNR and sparsity uncertainty have the most direct influence on vertical resolution. The determination of these two parameters can be understood as an optimization problem. From a practical point of view, the consistency between the invert P-impedance and P-impedance at the well is a more important economic criterion than the improvement of vertical resolution. Thus, CSSI sacrifices vertical resolution for a stable P-impedance result. As a result, we can see that the proposed method is closest to the CSSI result with high lateral continuity and little small-scale anomalies but has a higher vertical resolution.

To compare the overall lateral continuity improvement, we compare the proposed method with the 1-D ResANet using the interpolated wells around W1, W3, W4, and W6 as fine-tuning data while W2 and W5 are used as blind wells. As a result, we select totally 96 seismic and impedance pairs for 1-D ResANet and 68 traces with two traces on the left and right for the proposed

method to fine-tune the networks. Fig. 14 and Table IV list the prediction results of the field seismic data at W2 and W5. Although the PCC result of both blind wells of 1-D ResANet and 2-D CA-ResANet are very close, 2-D CA-ResANet has smaller MSE with fewer outliers (best results highlighted in bold red in Table IV). From the green oval box in Fig. 14, we can see that some offsets are reduced. On top of that, the prediction comparison profile on the cross-well section is shown in Fig. 15. The vertical solid green line and black curve represent the position and the impedance of the six wells (W1–W6), respectively. Fig. 15, especially the black circles, shows that the proposed method matches the ground truth better. These results further prove that 2-D CA-ResANet with hybrid loss tends to predict impedance with better lateral continuity and less noise.

IV. DISCUSSION

Table V compares the computational cost between 1-D ResANet and 2-D CA-ResANet. The number of parameters of 2-D convolution is much larger than that of 1-D convolution. For example, the number of parameters of the first convolution layer of 2-D CA-ResANet is 28 752 and that of 1-D ResANet is 464. We can see that the number of total parameters of 2-D CA-ResANet, 2 454 023, is about six times as large as that of 1-D ResANet (416 687). Thus, the training time for 2-D CA-ResANet (43 min 40 s) is longer than 1-D ResANet (15 min 38 s). The difference in the test efficiency of the two networks is not that large. The test time of the profile prediction of 2-D CA-ResANet (371 ms) is 102 ms longer (37%) than the time for 1-D ResANet.

From visual inspection of the cross-well section, the lateral continuity of the inverted impedance is an important criterion for practical implementation. We implement three tricks on this regard. First, 2-D convolutions can learn structural features
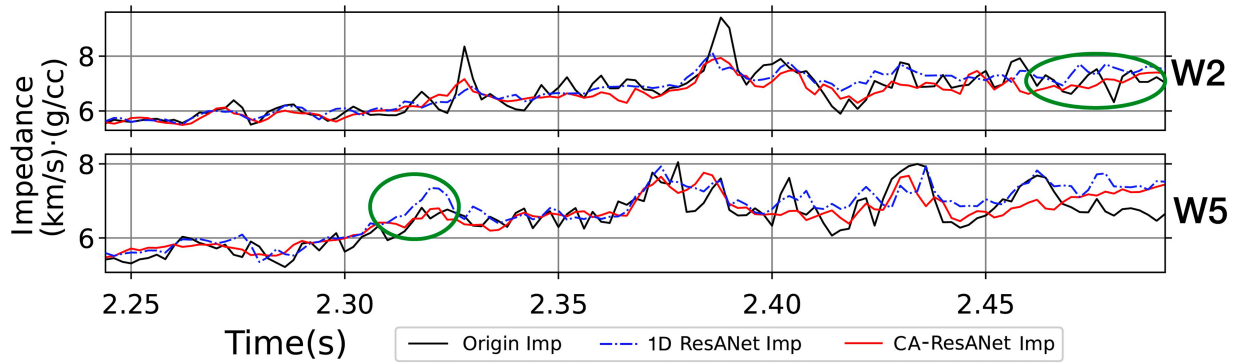
Fig. 14.   Prediction results for the field seismic data at W2 and W5 used as blind well test.

TABLE IV
QUANTITATIVE COMPARISON ON BLIND WELLS BETWEEN 1-D RESANET AND THE PROPOSED METHOD

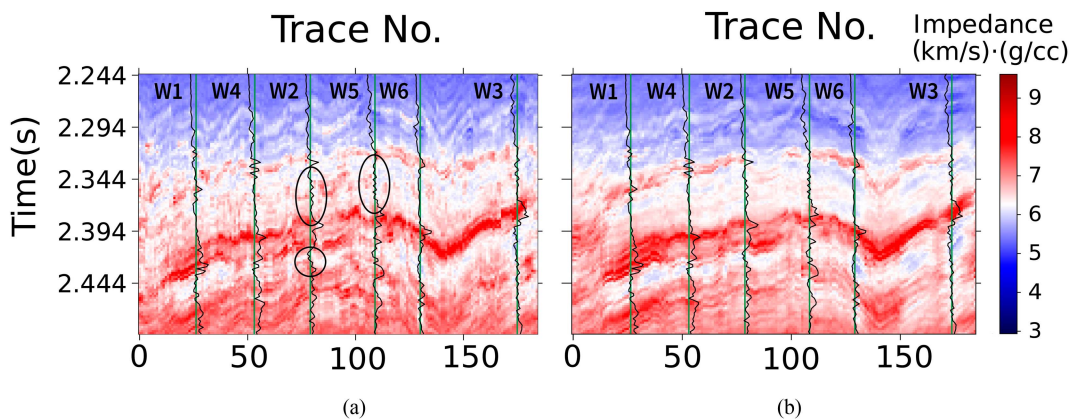| Evaluation metrics	Well	1D ResANet	2D CA-ResANet ($\lambda_1 = 0.3$)
MSE	W2	0.0018	**0.0017**
	W5	0.0016	**0.0014**
PCC	W2	0.8255	**0.8592**
	W5	**0.8583**	0.8306

(a)   (b)

Fig. 15.   Impedance prediction results taken W2 and W5 as the blind wells. (a) 1-D ResANet ($\lambda_1 = 0$). (b) 2-D CA-ResANet ($\lambda_1 = 0.3$).

TABLE V
COMPARISON BETWEEN DIFFERENT NETWORKS ON FIELD DATA

Model	Training time	Test time of profile prediction	Total parameters
1D ResANet	28min 02s	269ms	416687
2D CA-ResANet	43min 40s	371ms	2454023

and relationships between adjacent traces to predict impedance more stable. Second, a CA block can capture the stratum direction-aware and position-sensitive information in both time and horizontal structure dimensions. At the same time, CA can improve the DL-model performance without significant computation overhead. Third, the edge operator used in the hybrid loss can be regarded as a physical constraint to match the predicted structure with the true distribution. Under the constraint of structure information, the model can obtain more realistic

prediction results with fewer vertical strips and unreasonable noise and better lateral continuity, thus further enhancing the stability of the designed network. In this way, the proposed method can predict complex geological scenarios with better lateral continuity and stability. Comparison between Fig. 13(d) and Fig. 13(c), Fig. 13(e) and Fig. 13(c), Fig. 13(f) and Fig. 13(d) verifies the effectiveness of 2-D convolutions, the hybrid loss, and CA block, respectively. First and foremost, we recommend the use of 2-D networks for impedance inversion. Then CA block

can be inserted in any 2-D network to improve lateral continuity performance on the task of impedance inversion from seismic data, such as U-Net, GANs, temporal convolutional networks (TCNs), and so on. Finally, the loss function with an edge operator can be a good choice for facilitating model stability and reasonable results.

In addition, we can improve accuracy by adding other prior constraints. Biswas et al. [52] introduce physical laws, that is, wave-propagation physics, into the training process for seismic impedance inversion. Moreover, a low-frequency model is added to the generated impedance. Alfarraj et al. [15] utilized geophysical constraints, and seismic forward modeling, for the impedance inversion. Zhang et al. [53] conducted impedance inversion by a semisupervised framework with low-frequency extrapolated data. Mustafa et al. [54] performed impedance inversion based on a TCN. Our proposed method focuses more on geological structure, which can also be improved with prior constraints especially by adding low-frequency data. Lateral continuity occupies an essential position as an evaluation criterion for impedance inversion in practical application. It is promising that the proposed hybrid loss and the attention module are useful in DL-based impedance inversion methods. This offers a potential solution for lateral continuity improvement to the impedance inversion on field data or other related geological imaging tasks.

V. CONCLUSION

In this article, a 2-D CA-ResANet with a hybrid loss function is designed for impedance inversion. In synthetic SEAM model experiments, we use two ablation studies to determine our network architecture and the best hybrid loss. The quantitative tests are conducted to demonstrate the effectiveness of CA and the proposed loss. Furthermore, we use field data to illustrate the effectiveness of the proposed method. The field data experiments with the interpolated wells around all available wells indicate that the 2-D convolution, CA, and hybrid loss all can capture geological structure information. Overall, our method has a great improvement in lateral continuity, stability, and robustness against noise compared with the 1-D method and CSSI method using commercial software (InverTrace Plus module in Jason). A comparison between 1-D ResANet and the proposed method with two blind wells on the field data also justifies the superiority of our method.

ACKNOWLEDGMENT

REFERENCES

[1] Y. Wang, "Seismic impedance inversion using l 1-norm regularization and gradient descent methods," *J. Inverse Ill-Posed Probl.*, vol. 18, no. 7, pp. 823–838, Dec. 2010. [Online]. Available: https://doi.org/10.1515/jiip.2011.005

[2] H. Bertete-Aguirre, E. Cherkaev, and M. Oristaglio, "Non-smooth gravity problem with total variation penalization functional," *Geophys. J. Int.*, vol. 149, no. 2, pp. 499–507, May 2002.

[3] S. Dargan et al., "A survey of deep learning and its applications: A new paradigm to machine learning," *Arch. Comput. Methods Eng.*, vol. 27, no. 4, pp. 1071–1092, Sep. 2020.

[4] X. Wu et al., "Deep learning for multidimensional seismic impedance inversion," *Geophysics*, vol. 86, no. 5, pp. R735–R745, Sep. 2021.

[5] A. Adler, M. Araya-Polo, and T. Poggio, "Deep learning for seismic inverse problems: Toward the acceleration of geophysical analysis workflows," *IEEE Signal Process. Mag.*, vol. 38, no. 2, pp. 89–119, Mar. 2021.

[6] A. Pochet, P. H. B. Diniz, H. Lopes, and M. Gattass, "Seismic fault detection using convolutional neural networks trained on synthetic post-stacked amplitude maps," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 3, pp. 352–356, Mar. 2019.

[7] X. Li, B. Wu, X. Zhu, and H. Yang, "Consecutively missing seismic data interpolation based on coordinate attention UNet," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2021, Art. no. 3005005.

[8] X. Chai, G. Tang, S. Wang, K. Lin, and R. Peng, "Deep learning for irregularly and regularly missing 3-D data reconstruction," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 6244–6265, Jul. 2021.

[9] C. Qiu, B. Wu, N. N. Liu, X. Zhu, and H. Ren, "Deep learning prior model for unsupervised seismic data random noise attenuation," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art no. 7502005.

[10] D. Liu, W. Wang, X. Wang, C. Wang, J. Pei, and W. Chen, "Poststack seismic data denoising based on 3-D convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 3, pp. 1598–1629, Mar. 2020.

[11] V. Das, A. Pollack, U. Wollner, and T. Mukerji, "Convolutional neural network for seismic impedance inversion," *Geophysics*, vol. 84, no. 6, pp. R869–R880, Nov. 2019.

[12] B. Wu, D. Meng, L. Wang, N. 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.

[13] H. Di, X. Chen, H. Maniar, and A. Abubakar, "Seismic acoustic impedance estimation by learning from sparse wells via deep neural networks," in *Proc. 82nd EAGE Annu. Conf. Exhib.*, 2020, vol. 2020, pp. 1–5.

[14] A. Mustafa, M. Alfarraj, and G. AlRegib, "Estimation of acoustic impedance from seismic data using temporal convolutional network," in *Proc. SEG Tech. Prog. Expanded Abstr. Soc. Exploration Geophys.*, 2019, pp. 2554–2558.

[15] M. Alfarraj and G. AlRegib, "Semisupervised sequence modeling for elastic impedance inversion," *Interpretation*, vol. 7, no. 3, pp. SE237–SE249, Aug. 2019.

[16] B. Wu, D. Meng, and H. Zhao, "Semi-supervised learning for seismic impedance inversion using generative adversarial networks," *Remote Sens.*, vol. 13, no. 5, Feb. 2021, Art. no. 909.

[17] D. Meng, B. Wu, Z. Wang, and Z. Zhu, "Seismic impedance inversion using conditional generative adversarial network," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 7503905.

[18] A. Cai, H. 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. Soc. Exploration Geophys.*, 2020, pp. 1274–1278.

[19] J. A. Coffeen, *Seismic Exploration Fundamentals.* Tulsa, OK, USA: PennWell, 1978.

[20] C. Zhang and M. van der Baan, "Complete and representative training of neural networks: A generalization study using double noise injection and natural images," *Geophysics*, vol. 86, no. 3, pp. V197–V206, May 2021.

[21] X. Wu et al., "Deep learning for multidimensional seismic impedance inversion," *Geophysics*, vol. 86, no. 5, pp. R735–R745, Sep. 2021.

[22] Q. Wang, Y. Wang, Y. Ao, and W. Lu, "Seismic inversion based on 2D-CNNs and domain adaption," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5921512.

[23] M. Ge et al., "Semi-supervised impedance inversion by Bayesian neural network based on 2-D CNN pre-training," in *Proc. SEG Workshop, 4th Int. Workshop Math. Geophys., Traditional Learn., Virtual, 17–19 December 2021 Soc. Exploration Geophys.*, 2022, pp. 129–133.

[24] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 13708–13717.

[25] A. K. Jain, *Fundamentals of Digital Image Processing.* Upper Saddle River, NJ, USA: Prentice-Hall, 1989.

[26] J. Tsotsos, *A Computational Perspective On Visual Attention.* Cambridge, MA, USA: MIT Press, 2011.

[27] J. Tsotsos, "Analyzing vision at the complexity level," *Behav. Brain Sci.*, vol. 13, no. 3, pp. 423–469, Sep. 1990.

[28] J. Hu et al., "Gather-excite: Exploiting feature context in convolutional neural networks," in *Proc. Int. Conf. Neural Inf. Process Syst.*, 2018, pp. 9401–9411.

[29] S. Woo, J. Park, J. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 3–19.

[30] I. Bello et al., "Attention augmented convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2019, pp. 3285–3294.

[31] Q. Hou, L. Zhang, M. Cheng, and J. Feng, "Strip pooling: Rethinking spatial pooling for scene parsing," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 4003–4012.

[32] Z. Huang et al., "CCNet: Crisscross attention for semantic segmentation," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2019, pp. 603–612.

[33] J. Fu et al., "Dual attention network for scene segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 3146–3154.

[34] J. Yu and B. Wu, "Attention and hybrid loss guided deep learning for consecutively missing seismic data reconstruction," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5902108.

[35] K. Gao, L. Huang, Y. Zheng, R. Lin, H. Hu, and T. Cladohous, "Automatic fault detection on seismic images using a multiscale attention convolutional neural network," *Geophysics*, vol. 87, no. 1, pp. N13–N29, Jan. 2022.

[36] Y. Dou, K. Li, J. Zhu, X. Li, and Y. Xi, "Attention-based 3-D seismic fault segmentation training by a few 2-D slice labels," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5906715, doi: 10.1109/TGRS.2021.3113676.

[37] O. M. Saad, M. Bai, and Y. Chen, "Uncovering the microseismic signals from noisy data for high-fidelity 3D source-location imaging using deep learning," *Geophysics*, vol. 86, no. 6, pp. KS161–KS173, Oct. 2021.

[38] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141.

[39] A. Vaswani et al., "Attention is all you need," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.

[40] B. Wu, Q. Xie, and B. Wu, "Seismic impedance inversion based on residual attention network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4511117.

[41] X. Li, W. Wang, X. Hu, and J. Yang, "Selective Kernel networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 510–519.

[42] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.

[43] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn.*, 2010, pp. 807–814.

[44] G. Ian, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT press, 2016.

[45] E. S. Olivas et al., *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques: Algorithms, Methods, and Techniques.* Hershey, PA, USA: IGI Global, 2009.

[46] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?," *Neural Inf. Process. Syst.*, pp. 27, 2014.

[47] A. Mustafa and G. AlRegib, "Joint learning for seismic inversion: An acoustic impedance estimation case study," in *Proc. SEG Tech. Prog. Expanded Abstr. Soc. Exploration Geophysicists*, 2020, pp. 1686–1690.

[48] A. Mustafa, M. Alfarraj, and G. AlRegib, "Spatiotemporal modeling of seismic images for acoustic impedance estimation," in *Proc. SEG Tech. Prog. Expanded Abstr. Soc. Exploration Geophysicists*, 2020, pp. 1735–1739.

[49] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Representations*, 2015.

[50] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 1026–1034.

[51] H. W. J. Debeye and P. Van Riel, "Lp-NORM DECONVOLUTION," *Geophysical Prospecting*, vol. 38, no. 4, pp. 381–403, Jun. 1990.

[52] R. Biswas, M. K. Sen, V. Das, and T. Mukerji, "Prestack and poststack inversion using a physics-guided convolutional neural network," *Interpretation*, vol. 7, no. 3, pp. SE161–SE174, Aug. 2019.

[53] H. Zhang, P. Yang, Y. Liu, Y. Luo, and J. Xu, "Deep learning-based low-frequency extrapolation and impedance inversion of seismic data," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 7505905.

[54] A. Mustafa, M. Alfarraj, and G. AlRegib, "Joint learning for spatial context-based seismic inversion of multiple data sets for improved generalizability and robustness," *Geophysics*, vol. 86, pp. O37–O48, Jul. 2021.

Qiao Xie received the bachelor's degree in food science and engineering from Wuhan Polytechnic University, Wuhan, China, in 2015. She is currently working toward the master's degree in applied statistics with Xi'an Jiaotong University, Xi'an, China.

Her research interests include machine learning and deep learning for seismic data processing.

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 with Statoil (Beijing) Technology Service Company, Ltd., Beijing, China, from 2012 to 2015. He is 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.

Yueming Ye received the Ph.D. degree in geophysics from the China University of Petroleum (East China), Dongying, China, in 2010.

From 2008 to 2009, he was a visiting scholar with the University of California at Santa Cruz, Santa Cruz, CA, USA. He is a Senior Geophysical Engineer with PetroChina Hangzhou Research Institute of Geology, Hangzhou, China, and in charge of the Department of Computer Technology Application. His research interests include seismic wave imaging and multiples attenuation of marine seismic data.