



# Parametric elastic full waveform inversion with convolutional neural network

Keting Guo<sup>1,2</sup> · Zhaoyun Zong<sup>1,2</sup> · Jidong Yang<sup>1,2</sup> · Yuanyuan Tan<sup>1</sup>

Received: 6 October 2022 / Accepted: 12 April 2023 / Published online: 9 June 2023

© The Author(s) under exclusive licence to Institute of Geophysics, Polish Academy of Sciences & Polish Academy of Sciences 2023

## Abstract

Elastic full waveform inversion (EFWI) is a powerful tool for estimating elastic models by reducing the misfit between multi-component seismic records and simulated data. However, when multiple parameters are updated simultaneously, the gradients of the loss function with respect to these parameters will be coupled together, the effect exacerbate the nonlinear problem. We propose a parametric EFWI method based on convolutional neural networks (CNN-EFWI). The parameters that need to be updated are the weights in the neural network rather than the elastic models. The convolutional kernel in the network can increase spatial correlations of elastic models, which can be regard as a regularization strategy to mitigate local minima issue. Furthermore, the representation also can mitigate the cross-talk between parameters due to the reconstruction of Frechét derivatives by neural networks. Both forward and backward processes are implemented using a time-domain finite-difference solver for elastic wave equation. Numerical examples on overthrust models, fluid saturated models and 2004 BP salt body models demonstrate that CNN-EFWI can partially mitigate the local minima problem and reduce the dependence of inversion on the initial models. Mini-batch configuration is used to speed up the update and achieve fast convergence. In addition, the inversion of noisy data further verifies the robustness of CNN-EFWI.

**Keywords** Full waveform inversion · Parametric · Machine learning · Convolutional neural network

## Introduction

Elastic full waveform inversion (EFWI) is a powerful tool for estimating the elastic parameters such as *P*-wave and *S*-wave velocity (Mora 1987; Tarantola 1986). EFWI problem is mathematically equivalent to partial differential

equation (PDE) constrained optimization problem. It solves elastic wave equation and minimizes the misfit between the multicomponent records and simulated seismic data. EFWI can be implemented in frequency domain (Brossier et al. 2009; Pratt et al. 1998), the time domain (Shipp and Singh 2002), or in a hybrid domain (Nihei and Li 2007; Sirgue et al. 2008). EFWI can utilize the long-offset reflection data to estimate the amplitude variation result with offset (AVO) (Borisov and Singh 2015; Chen et al. 2021, 2022). However, compared with the acoustic case, numerical simulation in elastic media may be computationally expensive 4–5 times. When multiple parameters are involved in the inversion, the nonlinearity increases, and the inversion result of EFWI may fall into a local minimum due to the cycle-skipping issue (Forgues and Lambaré, 1997; Innanen 2014; Prioux et al. 2013; Wang and Cheng 2017; Zhang et al. 2021), and leads to increasing ill-posedness of the inversion problem.

To mitigate these problems, many approaches have been proposed in recent years. For example, one can use the Hessian operator for mitigating the parameter cross-talk because that can be regarded as a decoupling operator on the gradients (Innanen 2014; Operto et al. 2013). The algorithm based

---

Edited by Prof. Junlun Li (ASSOCIATE EDITOR) / Prof. Gabriela Fernández Viejo (CO-EDITOR-IN-CHIEF).

✉ Zhaoyun Zong  
zongzhaoyun@upc.edu.cn

Keting Guo  
s20010079@s.upc.edu.cn

Jidong Yang  
jidong.yang@upc.edu.cn

Yuanyuan Tan  
tyy17865422842@163.com

<sup>1</sup> School of Geosciences, China University of Petroleum (East China), Qingdao, China

<sup>2</sup> Pilot National Laboratory for Marine Science and Technology, Qingdao, China

on Hessian operator has a quadratic convergence rate, but it needs a strong computational consumption. To avoid this problem, many scholars use the approximate substitutions of Hessian operator such as the pseudo-Hessian (Choi et al. 2007; Shin et al. 2001) or a quasi-Newton method (Brossier et al. 2009). Another popular solution to the cross-talk problem is based on subspace method, which considers local projection onto a subspace of model parameters (Baumstein 2014; Kennett et al. 1988). In addition, the hierarchically inversion (Freudenreich and Singh 2000; Ren and Liu 2016; Sears et al. 2008; Tarantola 1986) and mode decomposed (Wang and Cheng 2017) are demonstrated have the ability to reduce the cross-talk.

In recent study, there are many studies to apply the deep learning to FWI. In general, those deep learning-based methods for FWI can be categorized into two types: (1) to learn a direct map from seismic data to velocity models. For instance, Yang and Ma (2019) establish an inverse regression with convolutional neural network (CNN) from synthetic training data set. Fabien-Ouellet and Sarkar (2020) estimate seismic velocity from CMP gathers using recurrent neural network (RNN). These methods are purely data-driven methods which can perform inversion directly after training networks. However, massive training data set are needed and the accuracy and generalization cannot be guaranteed without the physical constraints. In order to address these issues, Sun et al. (2021) proposed a physics-guided training strategy, which incorporates physical laws in the training of neural network, and takes advantage of data-driven deep learning and conventional physics-driven methods. Wu and Lin (2019) used an encoding–decoding network structure to model the map from seismic data to velocity structures. (2) To regard deep learning as an effective signal processing tool and does not need any training data set. Richardson (2018) and Sun et al. (2020) implement FWI in the framework of recurrent neural network (RNN) and use automatic differentiation technology for the gradient calculating instead of adjoint state method. Zhu et al. (2021) uses reverse-mode automatic differentiation to invert P-wave velocity from acoustic FWI, and demonstrate the process of reverse mode automatic differentiation is same with conventional adjoint state method. Zhang et al. (2021) and Wang et al. (2021) extend the RNN-based FWI to isotropic-elastic medium and anisotropic elastic medium. The latter proves the minibatch configuration is faster and more accurate than full-batch EFWI. These approaches allow for efficient calculation of the derivatives of the residual through automatic differential backpropagation method. However, they need to store all intermediate variables for gradient back-propagation, which is demanding for large memory resources. Complex neural networks can be constrained by physical rules to reduce its degrees of freedom. For example, Wu and McMechan (2019) and Zhu et al. (2022) reparametrize the

velocity model by CNN, the unknown parameters become the weights and bias in CNN instead of physics models, this reparameterization has the ability to suppresses the local minima issue.

In this paper, we demonstrate that the representation of elastic models by CNNs can mitigate the cycle-skipping and the cross-talk effect in EFWI. We first generate the elastic models from neural network and fed to the PDE solvers to simulate the multicomponent seismic data. The  $L2$  norm between synthetic data and true data is defined as cost function and gradient-based optimization method such as Adam optimization method (Kingma and Ba 2014) is used to decrease the value of cost function. In contrast to the traditional EFWI, the parameters to be optimized in CNN-EFWI are the weight and bias in neural network. By decomposing and reconstructing the Frechét derivatives, CNNs can decouple the P-and S-wave modes, leads to restrain the cross-talk between parameters. In addition, structure of CNN framework can be regarded as an implicit regularization method, which is crucial to for solving the local minimum problem of EFWI.

In the follow section, we first review the base theory of EFWI, including the forward and inversion parts. Then, we illustrate the methodology and workflow of CNN-EFWI. Numerical examples on overthrust model, fluid model and 2004 BP model verify the reparameterization and regularization of the CNN-EFWI inversion scheme.

## Theory

In the time domain, the 2D wave equation of the velocity-stress form is:

$$\begin{cases} \partial_t v_x = 1/\rho (\partial_x \sigma_{xx} + \partial_z \sigma_{xz}) \\ \partial_t v_z = 1/\rho (\partial_x \sigma_{xz} + \partial_z \sigma_{zz}) \\ \partial_t \sigma_{xx} = (\lambda + 2\mu) \partial_x v_x + \lambda \partial_z v_z + f_x \\ \partial_t \sigma_{zz} = (\lambda + 2\mu) \partial_z v_z + \lambda \partial_x v_x + f_z \\ \partial_t \sigma_{xz} = \mu (\partial_z v_x + \partial_x v_z), \end{cases} \quad (1)$$

where,  $v_x$  and  $v_z$  are the particle velocity along the  $x$  and  $z$  axis, respectively.  $\sigma_{xx}$  and  $\sigma_{zz}$  are two component of normal stress tensor.  $\sigma_{xz}$  is shear stress tensor.  $\partial_t$ ,  $\partial_x$  and  $\partial_z$  are the derivatives.  $\rho$  is the density.  $\lambda$  and  $\mu$  are the Lamé parameters. We solve Eq. (1) using time domain finite-difference (FD), which allows for an accurate modelling of  $P$ - and  $S$ -waves in elastic models. At the process of forward propagate, we use convolution perfectly matched layers (CPML) (Komatitsch and Martin 2007) to reduce unrealistic boundary reflected waves.

Mathematically, the inversion process of EFWI can be expressed as a PDE constrained optimization problem, which consists of minimizing the misfit between observed data  $d_{obs}(x, t)$  and prediction data  $d_{syn}(x, t)$ :

$$\min J = \sum_{i=1}^{N_s} J_i = \sum_{i=1}^{N_s} \sum_{t=1}^{N_t} \left\| d_i^{obs} - d_i \right\|_2^2 \tag{2}$$

$$s.t. d_i(x, t) = S(u_i(x, t))$$

$$\partial_t u_i(x, t) = F(u_i(x, t), \lambda(x), \mu(x), \rho(x)),$$

where  $J$  is the scalar cost function.  $N_s$  and  $N_t$  is the number of shots and time steps,  $u_i(x, t)$  is the wavefield at the time step  $t$  and the position  $x$ , which could be partial velocity or stress.  $S$  denotes a sampling function that is only non-zeros at the position of receivers and  $F$  is the solution of Eq. (1).

The mini-batch strategy is used for speed-up the convergence, we split the shot gathers into small batches and update parameters from each batch. The minibatch setup is widely used in traditional deep learning field, and it has a faster convergence speed than using the entire data set as a batch (Bishop 2006). With mini-batch strategy, the cost function written as:

$$\min J = \sum_{i=1}^{N_s} J_i = \sum_{i=1}^{N_{bs}} \sum_{t=1}^{N_t} \left\| d_i^{obs} - d_i \right\|_2^2 \tag{3}$$

$$s.t. d_i(x, t) = S(u_i(x, t))$$

$$\partial_t u_i(x, t) = F(u_i(x, t), \lambda(x), \mu(x), \rho(x)),$$

where the  $N_{bs}$  is the number of batch size. In full-batch EFWI algorithms, the gradients of cost function with respect to elastic models are calculated from all source gathers. However, in minibatch strategy, source gathers are distributed into small batches and the parameters are updated after one batch. In other words, this strategy has more update times than the traditional full-batch EFWI (Wang et al. 2021).

Equation (3) gives a quantitative measure of the misfit. The gradient of misfit function with respect to elastic parameters is calculated by adjoint state method, whose details are presented in Appendix I. The gradients' formulas are:

$$\partial_\lambda J = - \sum_{t=1}^{N_t} (\tilde{\sigma}_{zz}^{t+1} + \tilde{\sigma}_{xx}^{t+1}) (\partial_z V_z^t + \partial_x V_x^t) \Delta t$$

$$\partial_\mu J = - \sum_{t=1}^{N_t} [\tilde{\sigma}_{xz}^{t+1} (\partial_x V_z^t + \partial_z V_x^t) + 2\tilde{\sigma}_{zz}^{t+1} \partial_z V_z^t + 2\tilde{\sigma}_{xx}^{t+1} \partial_x V_x^t] \Delta t$$

$$\partial_\rho J = \sum_{t=1}^{N_t} \rho^{-2} [\tilde{v}_x^{t+1} (\partial_z \sigma_{xz}^t + \partial_x \sigma_{xx}^t) + \tilde{v}_z^{t+1} (\partial_x \sigma_{xz}^t + \partial_z \sigma_{zz}^t)] \Delta t, \tag{4}$$

where  $v_x$  and  $v_z$  are the forward partial velocity vectors in Eq. (1).  $\tilde{v}_x^{t+1}$  and  $\tilde{v}_z^{t+1}$  are the adjoint velocity vectors at the time step  $t + 1$ ,  $\tilde{\sigma}_{xx}^{t+1}$  and  $\tilde{\sigma}_{zz}^{t+1}$  are adjoint normal stress

vectors.  $\tilde{\sigma}_{xz}^{t+1}$  is three adjoint shear stress vectors.  $t$  and  $T$  are the discrete time step and maximum recording time, respectively. It is well known that the model parameterization is important for multi-parameter EFWI (Tarantola 1986). We choose  $P$ -wave velocity ( $V_p$ ),  $S$ -wave velocity ( $V_s$ ) and density ( $\rho$ ) as the parameterization. The corresponding gradients are derived by chain rule in the manner of Mora (1987):

$$\partial_{V_p} J = 2\rho V_p \partial_\lambda J$$

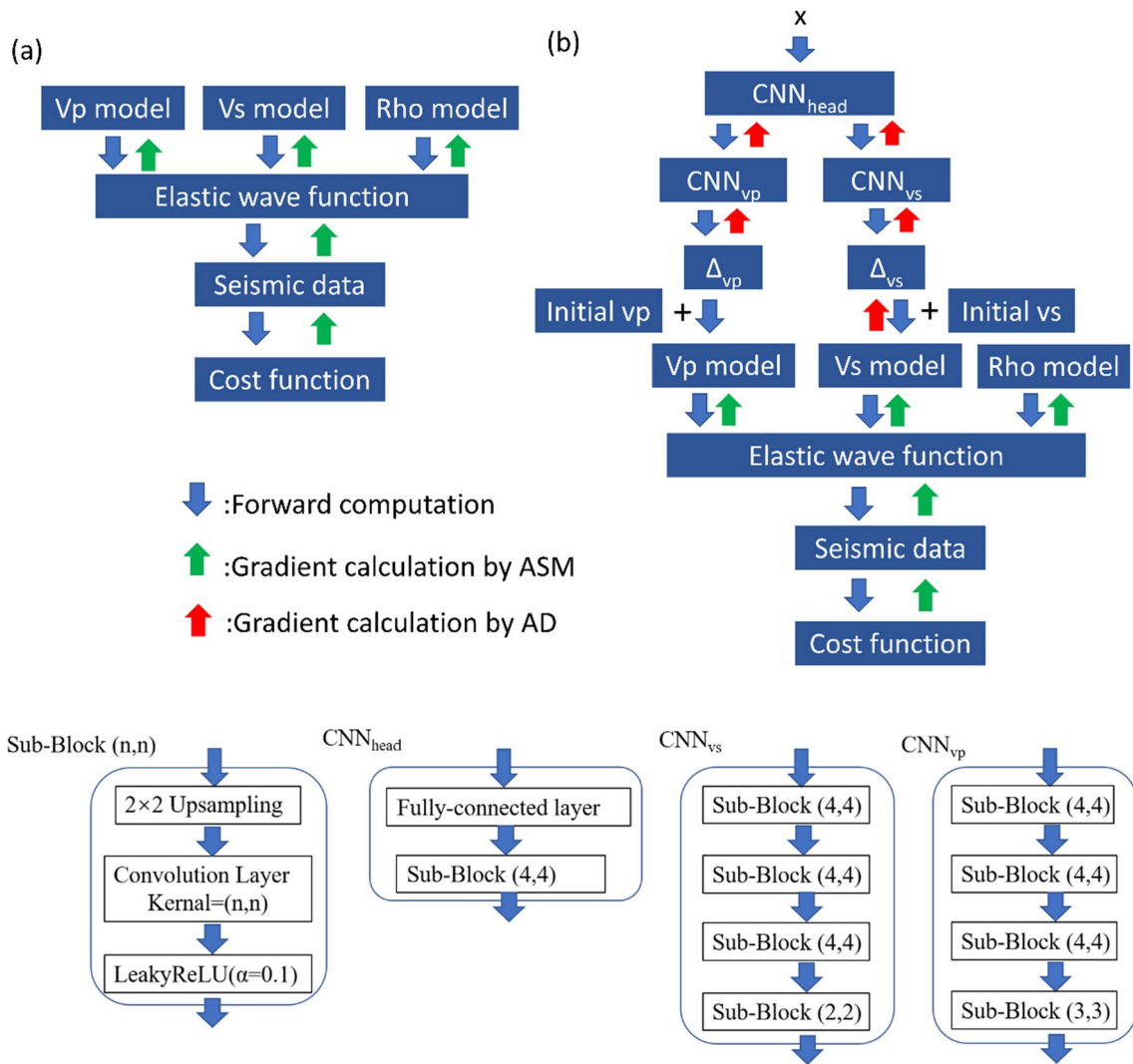
$$\partial_{V_s} J = -4\rho V_s \partial_\lambda J + 2\rho V_s \partial_\mu J, \tag{5}$$

Gradient-based inversion algorithm (such as L-BFGS, Adam) can be used to update velocity parameters. However, updating multiple parameters simultaneously is a challenging task because they can influence the seismic response together. The coupling of multiple elastic parameters can cause an update error in one parameter to affect other parameters. The various parameters have different sensitivities, which can exacerbate the crosstalk problem. Besides, due to the introduction of more degrees of freedom in the model space, the nonlinearity increases when multiple parameters are involved. In order to address these two issues, we reparameterize the velocity models using neural networks. The fundamental principle of neural network is the universal approximation theorem. In other words, neural networks have the ability to approximate any function using nonorthogonal basis functions. In the structure of convolutional neural networks (CNN), the deep image prior be used as regularization for tasks such as denoising (Lempitsky 2018). Therefore, the reparameterization using convolutional neural networks can introduce the regularization into EFWI, which is crucial to mitigate the local minimum problem in EFWI. In addition, CNN provide a sparse representation of the velocity models by introducing convolutional layers, so we can extract some specific features by constructing specific layers of the network. The diagram of traditional EFWI and proposed CNN-EFWI are shown in Fig. 1. The structure of CNN-EFWI is more complex than traditional EFWI. Velocity models  $V_p$  and  $V_s$  are reparametrized by generative neural network structure, the process is expressed as:

$$V_p = CNN_{vp}(x, w_{vp}) + V_{pinit}$$

$$V_s = CNN_{vs}(x, w_{vs}) + V_{sinit}, \tag{6}$$

where  $CNN_{vp}$  and  $CNN_{vs}$  are generative neural networks with respect to  $V_p$  and  $V_s$ , respectively.  $x$  is a random latent vector, which is transformed to a tensor by fully connected layer.  $w_{vp}$  and  $w_{vs}$  contain the weights and bias in neural networks. The velocity models generated by convolutional neural networks are input to the elastic wave Eq. (1) used in conventional EFWI.  $V_{pinit}$  and  $V_{sinit}$  are the initial  $P$ -wave and  $S$ -wave velocity models respectively. The initial model is directly combined with the output of the neural network,



**Fig. 1** The diagram of traditional EFWI (a) and proposed CNN-EFWI (b). Downward arrow denotes the forward computation. Red upward arrow denotes the gradient calculation by automatic differ-

entiation (AD) and green upward arrow denotes gradient calculation by adjoint state method (ASM). *SubBlock*(*n, n*) denotes the kernel of convolution layer in *SubBlock* is (*n* × *n*)

which means that the neural networks are trained to predict the residuals between the initial models and the real models. A detailed of the structure is shown in Fig. 1. Different with conventional inversion methods, which update  $V_{pinit}$  and  $V_{sinit}$  models directly, the parameters to be optimized in CNN-EFWI are the  $w_{vp}$  and  $w_{vs}$  in Eq. (6).

The gradients of cost function with respect to parameters have been decomposed and reconstructed in the process of inversion. With reparameterization by CNNs, the cost function of CNN-EFWI is defined by:

$$\min J = \sum_{i=1}^{N_s} J_i = \sum_{i=1}^{N_s} \sum_{t=1}^{N_t} \left\| d_i^{obs} - d_i \right\|_2^2 \tag{7}$$

$$s.t. d_i(x, t) = S(u_i(x, t))$$

$$\partial_t u_i(x, t) = F(u_i(x, t), N(vp(x)), N(vs(x)), N(\rho(x)))$$

To minimize the Eq. (7) using the gradient-based method, the gradient of the cost function with respect to the learnable parameters  $w_{vp}$  and  $w_{vs}$  should be calculated. The gradient formulas as:

$$\begin{aligned} \partial_{w_{vp}} J &= \frac{\partial N_{vp}}{\partial w_{vp}} \partial_{vp} J \\ \partial_{w_{vs}} J &= \frac{\partial N_{vs}}{\partial w_{vs}} \partial_{vs} J, \end{aligned} \tag{8}$$

where  $\partial N_{vp} / \partial w_{vp}$ ,  $\partial N_{vs} / \partial w_{vs}$  are gradients of neural network with respect to weight  $w_{vp}$  and  $w_{vs}$  respectively, which are obtained by automatic differentiation back-propagation in the deep learning framework such as Tensorflow or PyTorch.  $\partial_{vp} J$  and  $\partial_{vs} J$  are the Frechét derivatives in the Eq. (5). Equation (8) is a process of reconstructing the Frechét derivatives, which can help mitigate the cross-talk problem as demonstrated in next section. We use Adam algorithm to minimize the cost function (Eq. 7) efficiently. Adam applied the concept of momentum and adaptively adjusted the gradient using the exponentially decaying average of the previous squared gradient. The update formula is as follows:

$$w_{k+1} = w_k - \alpha P_k, \tag{9}$$

where  $\alpha$  is learning rate,  $k$  denotes the update number,  $P_k$  denotes the update direction of the cost function, which is:

$$P_k = \partial_w J + \beta P_{k-1}, \tag{10}$$

where  $\beta$  denotes the momentum parameter and  $\partial_w J$  is the gradient of cost function with respect to neural network weights  $w$ .

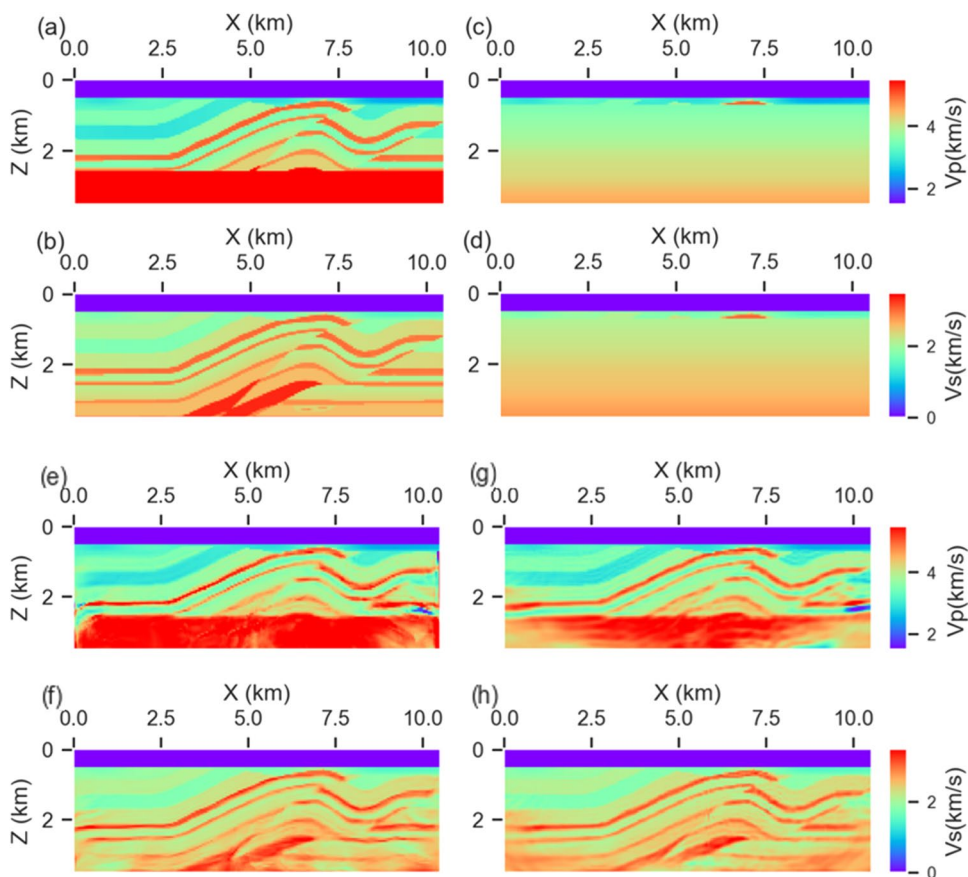
### Numerical experiments

In this section, we illustrate the reparameterization and regularization capabilities of CNN-EFWI with three comprehensive examples: SEG/EAGE overthrust models, fluid saturated models and 2004 BP salt body models. In these implementations, we assume that the source functions are well known and do not address this particular challenge of full waveform inversion.

#### SEG/EAGE over-thrust model

We take 2D overthrust models as the first example to evaluate the CNN-EFWI algorithm. The models are resampled with a grid size of  $68 \times 202$  and spatial sampling of  $\Delta x = \Delta z = 50m$ . Figure 2 shows the true  $P$ -wave and  $S$ -wave velocity models. To simulate offshore synthetic seismic data, we use 68 explosive Ricker wavelets with a domain frequency 3 Hz as the sources, and 199 receivers were fixed

**Fig. 2** SEG/EAGE overthrust  $P$ - and  $S$ -wave velocity models. (a)–(b) are the true models; (c)–(d) are the initial models; (e)–(f) is the inverted models by CNN-EFWI; (g)–(h) is the inverted models by conventional EFWI



on the seabed to simulate OBC survey. The total recording time is 5 s with an interval of 2.5 ms. In order to demonstrate the decoupling ability of proposed CNN-EFWI method, we insert a high velocity layer at the bottom of P-wave velocity

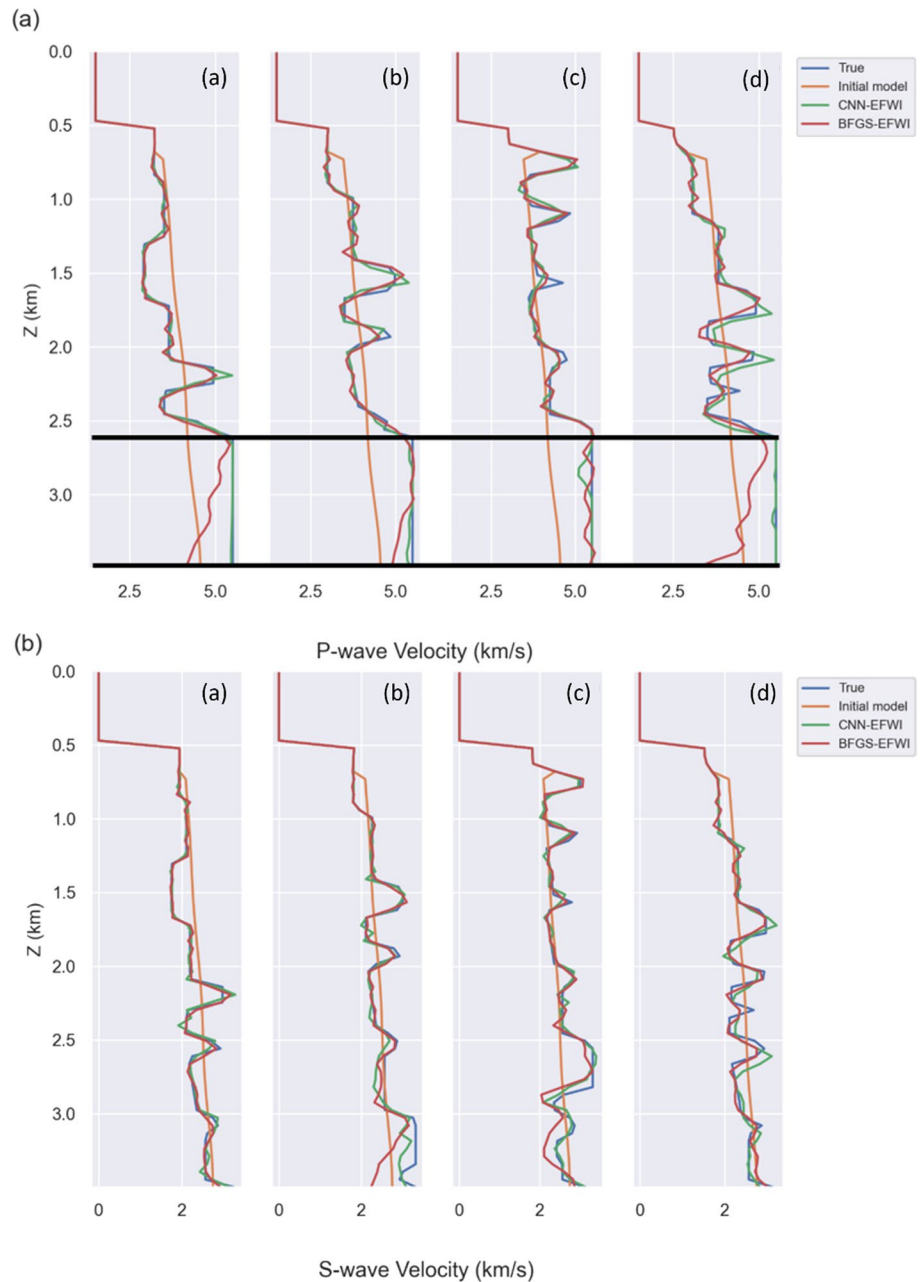
model. The initial models are the 1 – D increase models (Fig. 2). The same initial model and observed seismic data are used for conventional EFWI.

Table 1 lists the total computing times on GPUs with these two methods. We set the maximum iteration number as 40. We found that CNN-EFWI requires more computational time due to the additional neural network structure. But average time only increases about 5 percent. The inverted results are shown in Fig. 2. Both approaches have recovered *P*-wave velocity and *S*-wave velocity very well in the shallow part (above 2 km). The high-velocity thin layer in *P*-wave velocity has been recovered clearly in the results of CNN-EFWI. In contrast, conventional EFWI

**Table 1** The total computational costs of proposed CNN-EFWI and traditional EFWI

Method	Time spent (min)
CNN-EFWI	5.50
EFWI	5.26

**Fig. 3** Vertical profiles of overthrust *P*- wave velocity model (a) and *S*-wave velocity model (b). The area between two black lines in (a) is the high-velocity layer in *P*-wave velocity. The subplots (a): *X* = 2.02 km. (b): *X* = 4.04 km. (c): *X* = 6.06 km. (d): *X* = 8.08 km



losses some resolution for the layer at the left and right boundaries of the model. Due to the influence of cross-talk effect, the resolution at the bottom of *S*-wave wave velocity is low. The vertical profiles of the *P*-wave velocity and *S*-wave velocity (Fig. 3) further verify that the CNN-EFWI outperforms the conventional approach at the position of inserted layer.

In order to quantitatively compare the accuracy of these inverted results, we calculate the mean square error (*MSE*) and structural similarity index measure (*SSIM*) between the inverted models and the true models. They are defined as:

$$MSE(x, \hat{x}) = \sum_{i=1}^N [x(i) - \hat{x}(i)]^2 \tag{11}$$

$$SSIM(x, \hat{x}) = \frac{(2\mu_x\mu_{\hat{x}} + c_1)(2\sigma_{x\hat{x}} + c_2)}{(\mu_x^2 + \mu_{\hat{x}}^2 + c_1)(\sigma_x^2 + \sigma_{\hat{x}}^2 + c_2)}, \tag{12}$$

where *x* is the true model and  $\hat{x}$  is the estimated model,  $\mu_x$  and  $\mu_{\hat{x}}$  are the meaning of true model and inverted model, respectively.  $\sigma_x$  and  $\sigma_{\hat{x}}$  are variance of true model and inverted model, respectively.  $\sigma_{x\hat{x}}$  denotes the covariance between true and estimated model. *N* denotes the total number of points in the model. *MSE* is measure of *L2* norm distance between true and inverted models. *SSIM* evaluates the similarity of two images from local statistics, which close to

one indicate high quality inversion results. In Table 1, CNN-EFWI achieves more accurate inversion than conventional EFWI (Table 2).

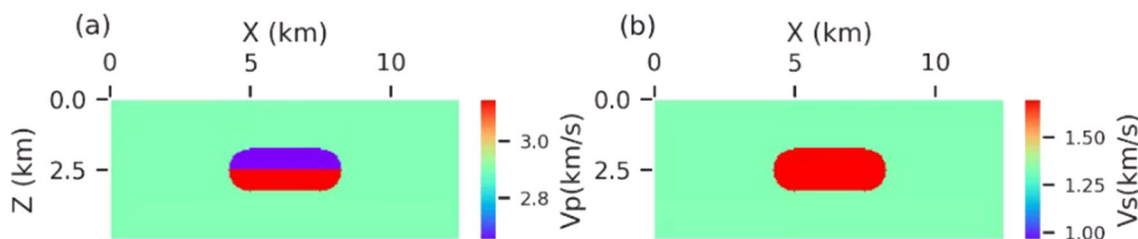
### Fluid saturated lens model

we consider a simple example of fluid saturated lens model. In many cases, FWI suffers from insufficient data to fully constrain inversion results. For example, inaccurate initial models can trap the optimization in the local minima and raise cycle skipping problem. As shown in Fig. 4, a square homogeneous background is perturbed two lens inclusions. We assume that the upper lens is gas saturated ( $V_p=2.65 \text{ km s}^{-1}, V_s=1.66 \text{ km s}^{-1}$ ), and lower lens is water-saturated ( $V_p=3.0 \text{ km s}^{-1}, V_s=1.66 \text{ km s}^{-1}$ ). In order to evaluate the regularization effect of CNN-EFWI, two initial velocity models are used. One is the smooth version of true velocity models (Fig. 5a,d) and another is set to the background homogeneous models (Fig. 5g,j). The grid size of the model is  $100 \times 250$  and the spatial sample is  $\Delta x = \Delta z = 15m$ . 32 cells are added to the boundary of models for CPML conditions. On the surface of the model, 50 shots are triggered with a horizontal interval of 30 m and receivers are deployed with an interval of 15 m. The sources are Ricker wavelets with a dominant frequency of 4.5 Hz. The seismic simulation time step is 2.5 ms, and the whole simulation time is 2.5 s.

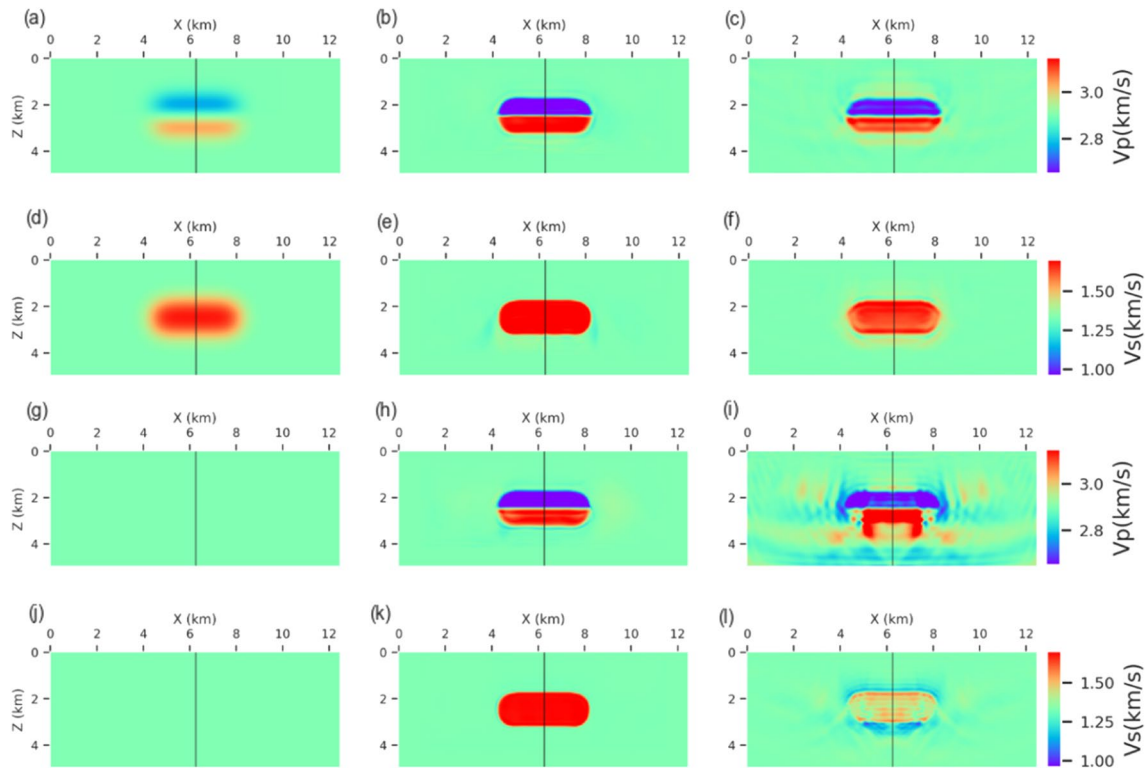
At the case of using smooth initial model, Fig. 5b,e show the *P*-wave velocity models inverted by CNN-EFWI and conventional EFWI, respectively. Figure 5c,f show the inversion results of corresponding *S*-wave velocity models. We observe that both *P*- and *S*-wave velocity are recovered very well by the CNN-EFWI, the sharpness of recovered lens outlines and the boundary is encouraging. In contrast, the inverted lens using conventional method are not as accurate as that obtained by CNN-EFWI especially at the low boundaries of the lens. Figure 6a,b,c,d show the difference between the true and inverted models. The figure indicate that the proposed approach estimates accurately for most parts of the models. The vertical profiles at the middle part of the models are shown in Fig. 7. We can see that the boundary of lens in CNN-EFWI results is sharper than results from conventional

**Table 2** *MSE* and *SSIM* between true and inverted overthrust models using CNN-EFWI and conventional EFWI

<i>P</i> -wave velocity		<i>S</i> -wave velocity	
CNN-EFWI	Conventional EFWI	CCN-EFWI	Conventional EFWI
<i>MSE</i>			
294.54	459.4	154.71	180.34
<i>SSIM</i>			
0.9754	0.9411	0.9862	0.981



**Fig. 4** The true fluid-saturated lens models. (a) is the true *P*-wave velocity model. (b) is the true *S*-wave velocity model



**Fig. 5** The initial fluid-saturated lens models and corresponding inverted results. (a)–(c) are smooth initial  $P$ -wave velocity model, inverted result with CNN-EFWI and conventional EFWI, respectively. (d)–(f) are corresponding  $S$ -wave velocity models. (g)–(i) are

constant initial  $P$ -wave velocity model, inverted result with CNN-EFWI and inverted result with conventional EFWI, respectively. (j)–(l) are corresponding  $S$ -wave velocity models

EFWI, which indicates that CNN-EFWI is effective and can provide higher resolution models compared with traditional methods. Quantitatively comparison is shown in Table 3. From this table one could get the same conclusion with above comparison.

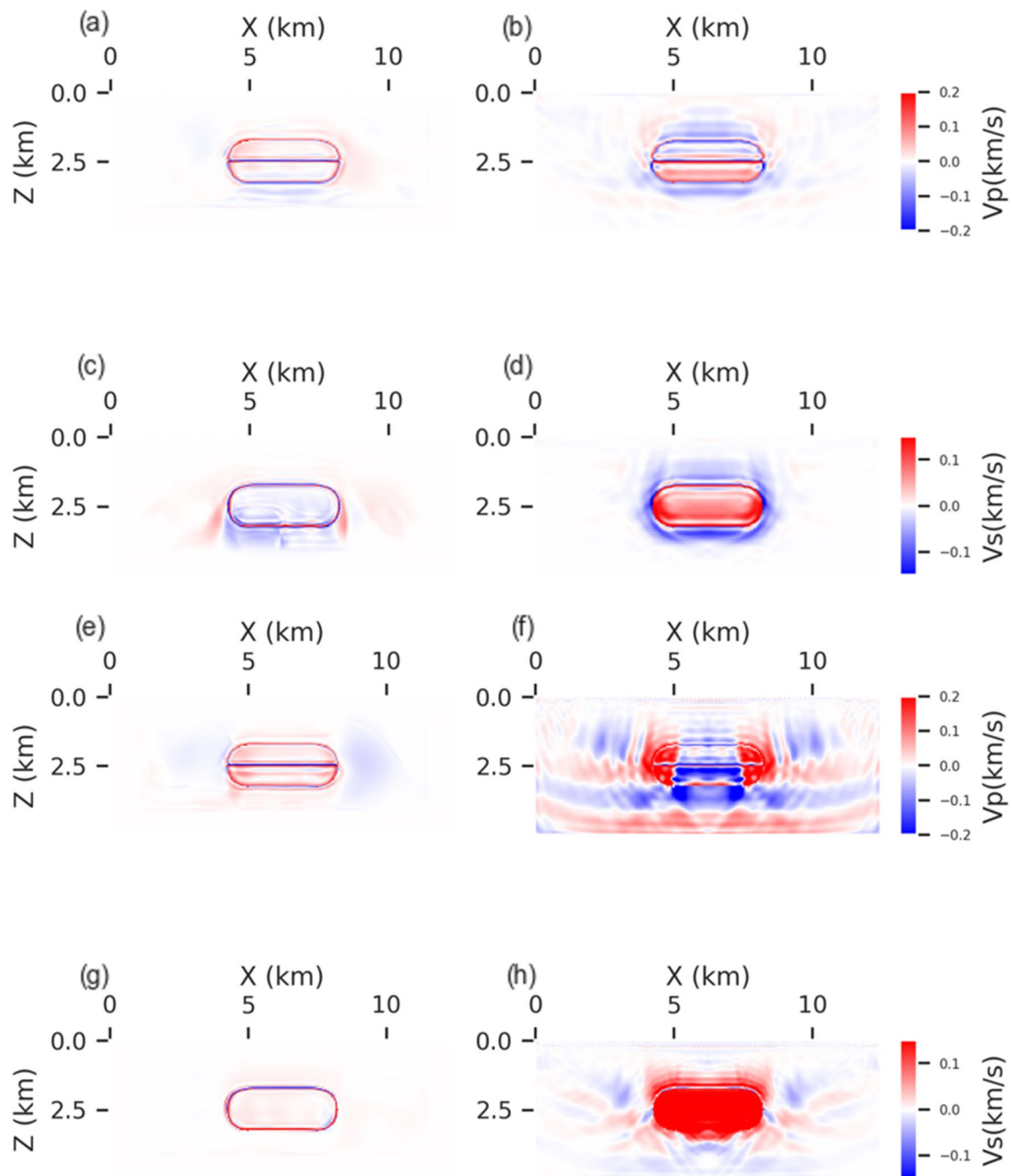
The inverted models using constant initial models is shown in Fig. 5. We can see that the inversion results using conventional method hardly image the structure of the lens in  $S$ -wave velocity model because of inadequate initial models. The cross-talk between  $P$ - and  $S$ -wave velocity lead to incorrect imaging lens in the  $S$ -wave velocity model. In addition, the lower boundaries of the lens inverted by conventional method is fuzzy. The lack of low-frequency content will accentuate the cycle-skipping problem, the optimization is trapped at a local minimum. In the early stage of optimization, due to the influence of parameter crosstalk, both  $P$ - and  $S$ -wave velocity cannot be updated correctly. Thus, the inversion eventually inadequate converges due to limited iterations. However, CNN-EFWI successfully mitigated the crosstalk and recovered sharper both  $P$ - and  $S$ -wave velocity models, especially at the upper and lower boundaries of the lens. Residual between true and inverted models are shown in Fig. 6. It is interesting to note that inverted  $S$ -wave velocity using constant initial models is more accurate than results

of using smooth initial models. Vertical profiles across the lens (Fig. 7) further verify that the CNN-EFWI has the ability to image fluid saturated lens even using constant initial models. The  $MSE$  and  $SSIM$  between models and seismogram are presented in Table 4. These comparisons verify the regularization effect and reparameterization ability from the CNN.

### BP salt body model

2004 BP models (Billette and Brandsberg-Dahl 2005) are difficult to image for EFWI because their complex salt bodies, which is based on a geological cross section through the Western Gulf of Mexico. The sharp contrast of large salt bodies leads to serious cycle-skipping and amplitude discrepancy. We slice the origin models and extract the left parts for tests. The  $S$ -wave velocity (Fig. 8c) was derived from  $P$ -wave velocity model (Fig. 8a) using an empirical formula (Mavko et al. 2020). The field covers the area of  $9.4 \text{ km} \times 3.4 \text{ km}$  ( $X$ - and  $Z$ -directions, respectively), the size of each cell is 50 m for both dimensions, representing a uniform mesh of  $68 \times 188$  grid points. C-PML boundary condition is also used in the test with 20 grid points. 63 sources wavelets with a 1.2 Hz domain frequency is used to generate



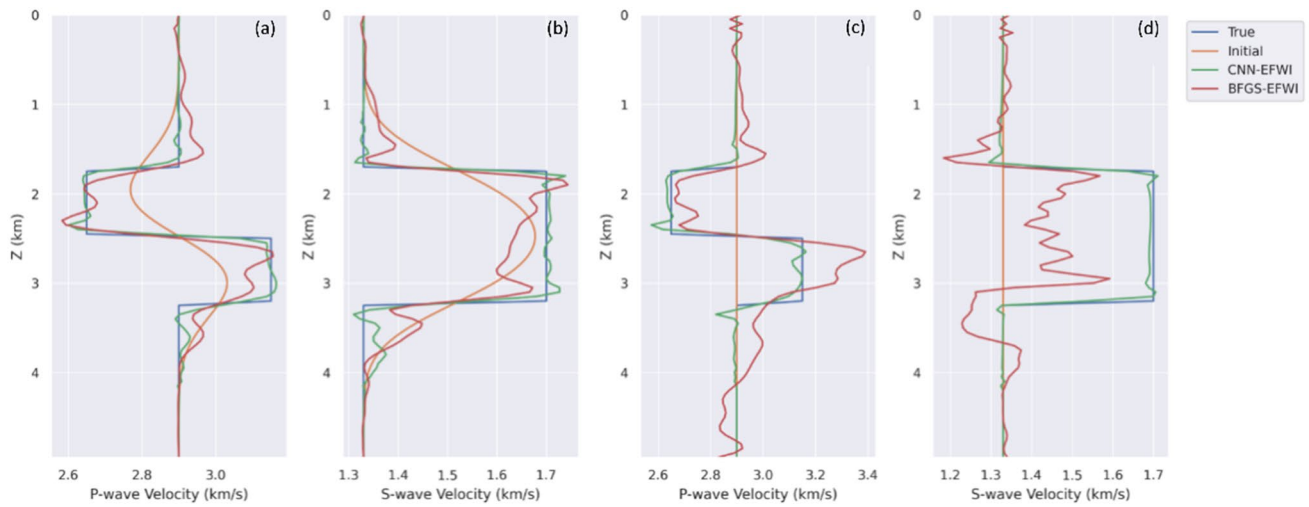


**Fig. 6** The residuals of true models minus inverted models. (a)–(b) are true  $P$ -wave velocity model minus inverted  $P$ -wave velocity models using smooth initial models with CNN-EFWI and conventional EFWI, respectively. (c)–(d) are corresponding  $S$ -wave velocity mod-

els. (e)–(f) are  $P$ -wave velocity models using constant initial models with CNN-EFWI and conventional EFWI, respectively. (g)–(h) are corresponding  $S$ -wave velocity models

multicomponent seismic waveform data. 185 receivers are deployed with an interval of 15 m. The time step is 9.2 ms, which is small enough to avoid dispersion. It is assumed that we know nothing of the salt bodies and the initial models are set to constant value:  $P$ -wave velocity =  $1.486 \text{ km s}^{-1}$  and  $S$ -wave velocity =  $0.891 \text{ km s}^{-1}$ , respectively.

Because the salt body is very complex, the learning rate setting is a trial-and-error process. If the learning rate is too small, the convergence speed of network parameters will be very slow. However, a high learning rate leads to divergence of the loss function. To permit a cross-talk between the speed of convergence and stability, we implement a multiple



**Fig. 7** The vertical profiles at the middle of the models. (a) and (b) are inverted *P*- and *S*-wave velocity results with smooth initial models. (c) and (d) are inverted *P*- and *S*-wave velocity results with constant initial models

**Table 3** MSE and SSIM between true and inverted fluid-saturated models using CNN-EFWI and conventional EFWI at the case of smooth initial models

P-wave velocity		S-wave velocity	
CNN-EFWI	Conventional EFWI	CCN-EFWI	Conventional EFWI
<i>MSE</i>			
17.9	27.39	16.24	33.56
<i>SSIM</i>			
0.9716	0.9255	0.9884	0.9397

**Table 4** MSE and SSIM between true and inverted fluid-saturated models using CNN-EFWI and conventional EFWI at the case of constant initial models

P-wave velocity		S-wave velocity	
CNN-EFWI	Conventional EFWI	CCN-EFWI	Conventional EFWI
<i>MSE</i>			
22.94	66.06	14.7	95.78
<i>SSIM</i>			
0.9522	0.7368	0.9901	0.3878

learning rate strategy within the process of optimization. The learning rate become smaller when loss function decrease.

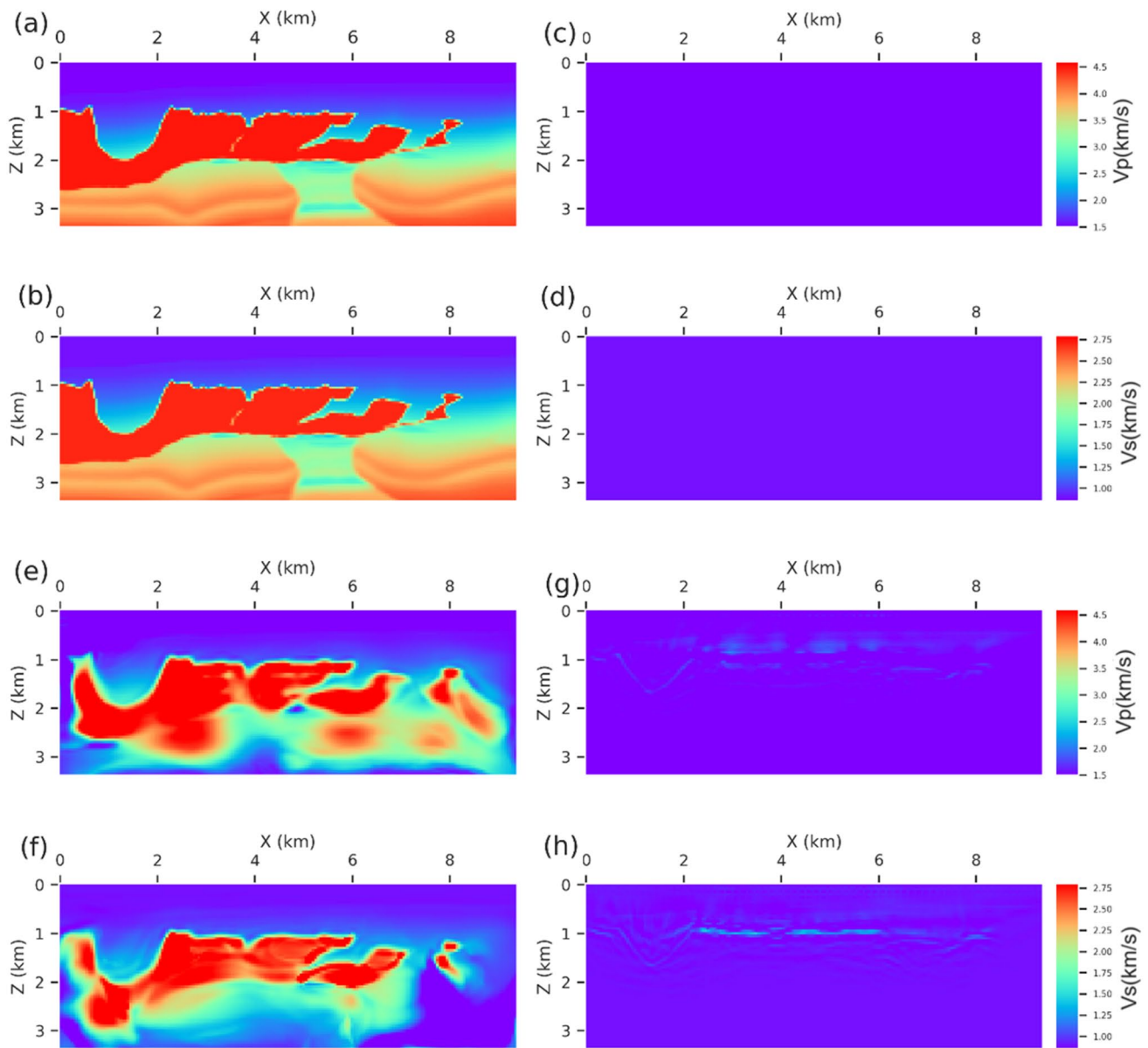
A comparison of the inverted results from different approaches is shown in Fig. 8. As we can see, Convolutional method cannot recover the shape of salt body of both *P*- and *S*- wave velocity models. Only the upper boundary of salt body is recovered. There are no updates for both velocity

models in the case of conventional EFWI. In contrast, CNN-EFWI correctly images the salt body, especially the lower boundary of salt body and the left U-shaped target. It is demonstrated that CNN-EFWI can image the *P*- and *S*- velocity of complex salt body using constant initial models. Figure 9 shows the features learned from one of the CNNs of the 50 iterations and 100 iterations. Noted that with the increasing of iterations, the features learned from the network are closer to the shape of salt body.

Next, we added two different levels of noise to the CMP gathers to verify the robust of CNN-EFWI. As shown in Fig. 10. The SNR (signal/noise ratio) are 25 and 20 dB, respectively. Figure 11 shows the final inverted *P*- and *S*-wave velocity models using different noisy data. We observe that a significant error in the lower right corner of *P*-wave velocity results with the increase of noise. However, the position of main salt body is located successfully, even the U-shape target is correctly imaged. The example demonstrates that spatial regularization from convolutional neural networks could reduce the influence of noise on inversion results.

### Discussion

Proposed method provides a framework for combining the convolutional neural networks and EFWI applications. Different with purely data-driven method, which rely on a large number of training data set, including data pairs of velocity models and corresponding seismic data. CNN-EFWI can introduce elastic physical information into neural networks and does not need any extra data set.



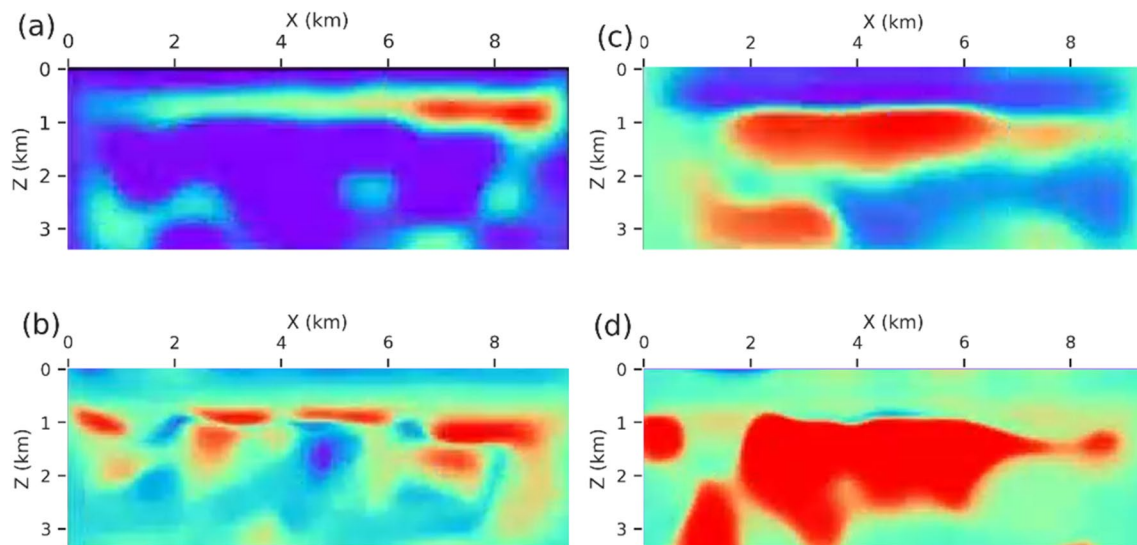
**Fig. 8** 2004 BP salt body models. (a) and (b) The true  $P$ - and  $S$ -wave velocity models. (c) and (d) are initial  $P$ - and  $S$ -wave velocity models. (e) and (f) are inverted  $P$ - and  $S$ -wave velocity models using CNN-

EFWI. (g) and (h) inverted  $P$ - and  $S$ -wave velocity models using conventional EFWI

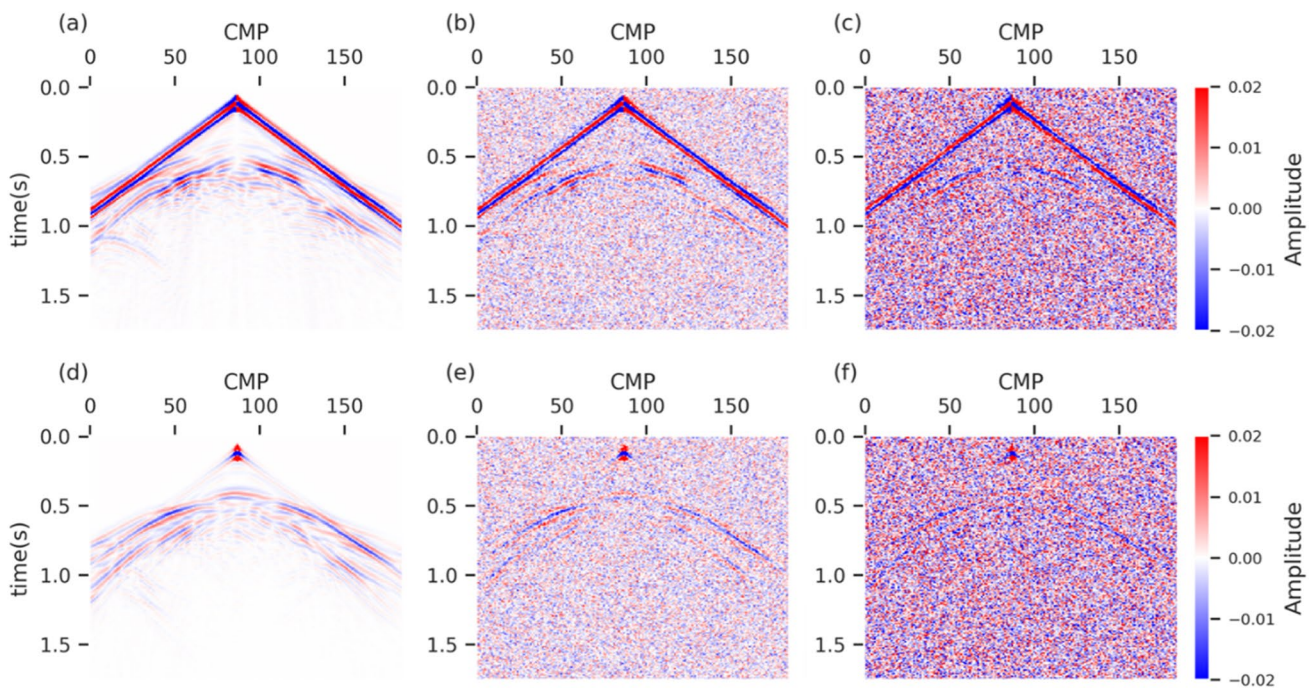
The EFWI also could be applied in the frequency domain, which is mathematically equivalent to the time domain method (Pratt et al. 1998). However, the implementation of the frequency domain approach is memory consuming. In addition, another crucial problem is that the inversion of a set of sparse frequencies is susceptible to the Gibbs phenomenon (Brenders et al. 2012). Therefore, the time domain approach is more suitable for inverting complex elastic models. In addition, in the time domain method, the number of shots and the number of time steps are the key factors of parallel computation, which makes it easy to use GPU to accelerate the computation.

Compare to traditional EFWI, CNN-EFWI can provide more stable inversion results because of the convolutional structure in neural network, which increase spatial correlations of elastic models and can be regard as a regularization strategy to mitigate local minima issue, even if the initial models are not accurate enough. But it deeply relies on the low frequency data in seismic data, CNN-EFWI can't invert reasonable results without it.

Models parameterized using Neural networks increases the flexibility of inversion. However, more flexibility makes the method less robust. To mitigate the problem, the maximum and minimum of models should be defined to constrain



**Fig. 9** Features learned by neural networks. (a) and (b) are features of one layer in  $CNN_{vp}$  after 50 and 100 iterations, respectively. (c) and (d) are features of one layer in  $CNN_{vs}$  after 50 and 100 iterations, respectively



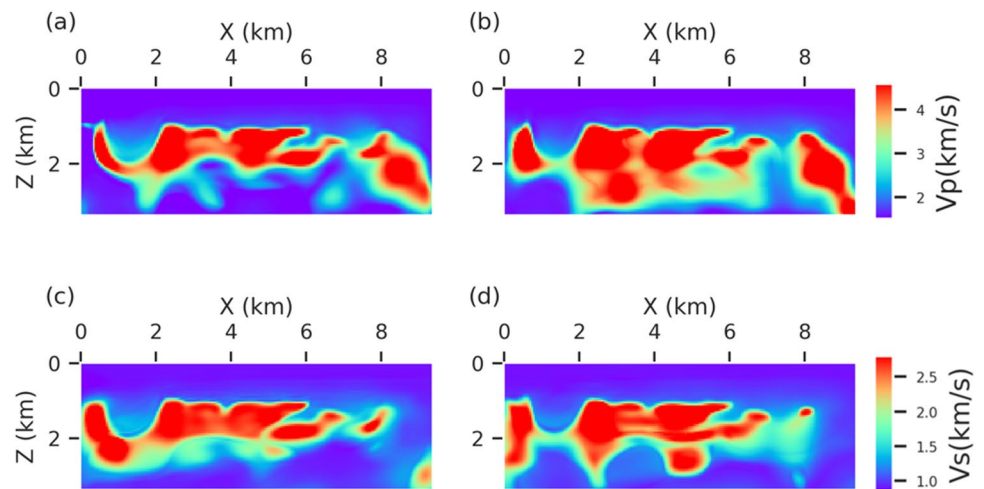
**Fig. 10** The seismogram generated from true models at the position of  $(x, z) = (5.0, 0.0)$  km. (a) and (d) are noise free horizontal and vertical particle-velocity common-source gathers, respectively. (b) and (e) are horizontal and vertical particle-velocity common-

source gathers with 10 dB noise, respectively. (c) and (f) are horizontal and vertical particle-velocity common-source gathers with 20 dB noise, respectively

the optimization in the process of inversion. Otherwise, the models will update in the wrong direction and be trapped in local minima. Typically, the minimum value of the models is set to zero, while the maximum value of models is set as:

$$V_{p_{\max}} = \frac{0.6 \min(dx, dz)}{\sqrt{2}dt} \quad (13)$$

**Fig. 11** Inverted models with noise data using CNN-EFWI. (a) and (c) are inverted  $P$ - and  $S$ -wave velocity models with 25 dB noise. (b) and (d) are inverted  $P$ - and  $S$ -wave velocity models with 20 dB noise



$$V_{s_{\max}} = V_{p_{\max}} / \sqrt{2}, \quad (14)$$

where the  $V_{p_{\max}}$  and  $V_{s_{\max}}$  denote the maximum value of  $P$ - and  $S$ -wave velocity models respectively.  $\min(dx, dz)$  is the minimum between  $dx$  and  $dz$ . Formula (12) is the stability condition of finite difference.

Noted that density is not inverted in our tests, which is difficult to reconstruct because its perturbations have hardly any effect on the phases or travel times of  $P$ - and  $S$ -waves. For simplicity, many studies assume that models have the constant density. (Brossier et al. 2009; Sears et al. 2008; Shipp and Singh 2002) or calculate by the empirical relationship with  $P$ -wave velocity (Borisov and Singh 2015). However, there also are some scholars consider density as an inversion parameter. For example, Xu and McMechan (2014) use a multistep-length EFWI approach to mitigate the cross-talk between  $P$ -wave velocity,  $S$ -wave velocity and density. Recently, Zhang et al. (2021) invert the density in the framework of RNN. Extension to density inversion is one direction of future works.

CNN-EFWI is not limited to the specific configuration used in this paper. The tune of hyperparameters of neural networks is a crucial challenge. A variety of types of hyperparameters in CNN-EFWI should be considered, including the layers of neural networks, the learning rate, the scale parameters, the activation function and the optimization algorithm. The hyperparameters in our implementation are not unique, because the turning of these hyperparameters is a trial-and-error process. It should be noted that the resolution of  $P$ -wave velocity is less than  $S$ -wave velocity because the spatial wavelengths of  $P$ -wave propagating in medium

are greater than wavelengths of  $S$ -wave. Therefore, the convolutional kernels in CNNs with respect to  $P$ -wave velocity should be greater than that of  $S$ -wave velocity.

## Conclusion

In this work, we introduce a method CNN-EFWI to mitigate the cycle-skipping problem in EFWI by combining the convolutional neural network and PDEs. The weights and bias in CNNs are updated by connecting the gradient from adjoint state method and automatic differentiation. The convolutional kernel in CNN increases the spatial correlation of models, which can be regarded as an implicit regularization, which is significant for EFWI. In addition, CNNs also have the ability to decouple the  $P$ -wave and  $S$ -wave modes by reconstructing the Fréchet derivatives. Numerical example on overthrust models demonstrate the reparameterization effect of proposed CNN-EFWI. Then, we design a fluid saturated lens model, inversion results show that proposed method outperforms than conventional method. Even the initial model is a constant model, the proposed method can image the lens clearly. Finally, the example of 2004 BP salt body model further verify the regularization ability of neural network. In addition, the features learned by neural network automatically filter out noise, which improves the robustness of inversion. CNN-EFWI can be directly applied to the same datasets as conventional EFWI to improve the inversion performance.

## Appendix I

In this appendix, we derive the gradient of cost function with respect to Lamé parameters ( $\lambda$  and  $\mu$ ) and density ( $\rho$ ). For simplicity, we ignore the non-reflecting PML regions. The sum of squared errors is used as the cost function:

$$\min J = \sum_{i=1}^{N_s} J_i = \sum_{i=1}^{N_{bs}} \sum_{t=1}^{N_t} \left\| d_i^{obs} - d_i \right\|_2^2, \quad (15)$$

where  $N_t$  is the number of time points. In the adjoint state method algorithms, the gradient for elastic model parameters  $\partial J/\partial\lambda$ ,  $\partial J/\partial\mu$  and  $\partial J/\partial\rho$  depend on the adjoint wavefields, which is calculated by chain rule. We obtain the adjoint equations:

$$\begin{aligned} \tilde{\sigma}_{xx}(r, t - \Delta t) &= \tilde{\sigma}_{xx}(r, t) + \frac{\partial \tilde{v}_x(r, t)}{\partial_{-x}} \Delta t \frac{1}{\rho} + S_x(t) \\ \tilde{\sigma}_{zz}(r, t - \Delta t) &= \tilde{\sigma}_{zz}(r, t) + \frac{\partial \tilde{v}_z(r, t)}{\partial_{-z}} \Delta t \frac{1}{\rho} + S_z(t) \\ \tilde{\sigma}_{xz}(r, t - \Delta t) &= \tilde{\sigma}_{xz}(r, t) + \left( \frac{\partial \tilde{v}_x(r, t)}{\partial_{+z}} + \frac{\partial \tilde{v}_z(r, t)}{\partial_{+x}} \right) \Delta t \frac{1}{\rho} \\ \tilde{v}_x(r, t - \Delta t) &= \tilde{v}_x(r, t) + \left[ \frac{\partial \tilde{\sigma}_{xx}(r, t)}{\partial_{+x}} (\lambda + 2\mu) + \frac{\partial \tilde{\sigma}_{zz}(r, t)}{\partial_{+x}} \lambda + \frac{\partial \tilde{\sigma}_{xz}(r, t)}{\partial_{-z}} \mu \right] \Delta t \\ \tilde{v}_z(r, t - \Delta t) &= \tilde{v}_z(r, t) + \left[ \frac{\partial \tilde{\sigma}_{zz}(r, t)}{\partial_{+z}} (\lambda + 2\mu) + \frac{\partial \tilde{\sigma}_{xx}(r, t)}{\partial_{+z}} \lambda + \frac{\partial \tilde{\sigma}_{xz}(r, t)}{\partial_{-x}} \mu \right] \Delta t, \end{aligned} \quad (16)$$

where the wavefields with a tilde denote the adjoint wavefield. In other words, they are the partial derivatives of  $J$  with respect to corresponding wavefields. For example,  $\tilde{v}_z$  indicates  $\partial J/\partial v_z$ .  $\Delta t$  is discrete time interval.  $S_x(t)$  and  $S_z(t)$  are adjoint source.

As the adjoint wave fields propagate backward in time, the derivatives of the objective function with respect to the elastic parameters are calculated as:

$$\begin{aligned} \frac{\partial \sigma_{xx}}{\partial \lambda} &= \frac{\partial \sigma_{zz}}{\partial \lambda} = \left( \frac{\partial v_x}{\partial x} + \frac{\partial v_z}{\partial z} \right) \Delta t \\ \frac{\partial v_x}{\partial \rho} &= -\rho^{-2} \left( \frac{\partial \sigma_{xx}}{\partial x} + \frac{\partial \sigma_{xz}}{\partial z} \right) \Delta t \\ \frac{\partial v_z}{\partial \rho} &= -\rho^{-2} \left( \frac{\partial \sigma_{zz}}{\partial z} + \frac{\partial \sigma_{xz}}{\partial x} \right) \Delta t \\ \frac{\partial \sigma_{xx}}{\partial \mu} &= 2 \frac{\partial v_x}{\partial x} \Delta t, \quad \frac{\partial \sigma_{zz}}{\partial \mu} = 2 \frac{\partial v_z}{\partial z} \Delta t \\ \frac{\partial \sigma_{xz}}{\partial \mu} &= \left( \frac{\partial v_x}{\partial z} + \frac{\partial v_z}{\partial x} \right) \Delta t. \end{aligned} \quad (17)$$

Note that the system is solved backwards in time. Connect Eqs. (3) and (2), the gradients of the objective function  $J$  with respect to elastic models  $\lambda$ ,  $\mu$  and  $\rho$  are:

$$\begin{aligned} \frac{\partial J}{\partial \lambda} &= - \sum_{t=1}^{N_t} (\tilde{\sigma}_{zz}^{t+1} + \tilde{\sigma}_{xx}^{t+1}) \left( \frac{\partial V_z^t}{\partial z} + \frac{\partial V_x^t}{\partial x} \right) \Delta t \\ \frac{\partial J}{\partial \mu} &= - \sum_{t=1}^{N_t} \left[ \tilde{\sigma}_{xz}^{t+1} \left( \frac{\partial V_z^t}{\partial x} + \frac{\partial V_x^t}{\partial z} \right) + 2\tilde{\sigma}_{zz}^{t+1} \frac{\partial V_z^t}{\partial z} + 2\tilde{\sigma}_{xx}^{t+1} \frac{\partial V_x^t}{\partial x} \right] \Delta t \\ \frac{\partial J}{\partial \rho} &= \sum_{t=1}^{N_t} \rho^{-2} \left[ \tilde{v}_x^{t+1} \left( \frac{\partial \sigma_{xz}^t}{\partial z} + \frac{\partial \sigma_{xx}^t}{\partial x} \right) + \tilde{v}_z^{t+1} \left( \frac{\partial \sigma_{xz}^t}{\partial x} + \frac{\partial \sigma_{zz}^t}{\partial z} \right) \right] \Delta t. \end{aligned} \quad (18)$$

**Acknowledgements** We would like to acknowledge the sponsorship of National Natural Science Foundation of China (42174139, 41974119, 42030103) and Science Foundation from Innovation and Technology Support Program for Young Scientists in Colleges of Shandong province and Ministry of Science and Technology of China (2019RA2136) and support by the Marine S&T Fund of Shandong Province for Pilot National Laboratory for Marine Science and Technology (Qingdao) (Grant No. 2021QNLM020001-6). (Forgues and Lambaré, 1997).

**Data availability** Data associated with this research are available and can be accessed via the following URL: <https://github.com/guoketing/CNN-FWI>.

## Declarations

**Conflict of interest** The authors declare that they have no competing interests. No infringement of others' intellectual property rights.

## References

- Baumstein A (2014) Extended subspace method for attenuation of crosstalk in multi-parameter full waveform inversion In: Proceedings 2014 SEG annual meeting 2014, All days: SEG-2014-0546.
- Billette FJ, and Brandsberg-Dahl S (2005) The 2004 BP velocity benchmark In: Extended abstracts <https://doi.org/10.3997/2214-4609-pdb.1.B035>.
- Bishop CM (2006) Pattern recognition and machine learning (information science and statistics). Springer, New York
- Borisov D, Singh SC (2015) Three-dimensional elastic full waveform inversion in a marine environment using multicomponent ocean-bottom cables: a synthetic study. *Geophys J Int* 201(3):1215–1234. <https://doi.org/10.1093/gji/ggv048>
- Brenders AJ, Albertin U and Mika J (2012) Comparison of 3-D time- and frequency-domain waveform inversion: Benefits and insights of a broadband, discrete-frequency strategy In: Proceedings 2012 SEG annual meeting, OnePetro.
- Brossier R, Operto S, Virieux J (2009) Seismic imaging of complex onshore structures by 2D elastic frequency-domain full-waveform inversion. *Geophysics* 74(6):105–118. <https://doi.org/10.1190/1.3215771>
- Chen F, Zong Z, Jiang M (2021) Seismic reflectivity and transmissivity parametrization with the effect of normal in situ stress. *Geophys J Int* 226(3):1599–1614. <https://doi.org/10.1093/gji/ggab179>
- Chen F, Zong Z, Yang Y, Gu X (2022) Amplitude-variation-with-offset inversion using P- to S-wave velocity ratio and P-wave velocity. *Geophysics* 87(4):N63–N74. <https://doi.org/10.1190/geo2021-0623.1>
- Choi Y, Shin C, Min D-J (2007) Frequency-domain elastic full-waveform inversion using the new pseudo-Hessian matrix: elastic

- Marmousi-2 synthetic test, SEG technical program expanded abstracts 2007. Soc Explor Geophys 65:1908–1912
- Fabien-Ouellet G, Sarkar R (2020) Seismic velocity estimation: a deep recurrent neural-network approach. Geophysics 85(1):U21–U29. <https://doi.org/10.1190/geo2018-0786.1>
- Forgues E, Lambaré G (1997) Parameterization study for acoustic and elastic ray plus born inversion. J Seism Explor 6(2–3):253–277
- Freudenreich Y, Singh S (2000). Full waveform inversion for seismic data—Frequency versus time domain. <https://doi.org/10.3997/2214-4609-pdb.28.C54>
- Innanen KA (2014) Seismic AVO and the inverse Hessian in precritical reflection full waveform inversion. Geophys J Int 199(2):717–734. <https://doi.org/10.1093/gji/ggu291>
- Kennett BLN, Sambridge MS, Williamson PR (1988) Subspace methods for large inverse problems with multiple parameter classes. Geophys J Int 94(2):237–247. <https://doi.org/10.1111/j.1365-246X.1988.tb05898.x>
- Kingma DP and Ba JJ (2014) Adam: a method for stochastic optimization. *arXiv preprint arXiv:1412.6980*
- Komatitsch D, Martin R (2007) An unsplit convolutional perfectly matched layer improved at grazing incidence for the seismic wave equation. Geophysics 72(5):SM155–SM167. <https://doi.org/10.1190/1.2757586>
- Lempitsky, V., A. Vedaldi, and D. Ulyanov. 2018, Deep image prior. paper read at 2018 IEEE/CVF conference on computer vision and pattern recognition. pp. 18–23.
- Mavko G, Mukerji T, Dvorkin J (2020) The rock physics handbook. Cambridge University Press
- Mora P (1987) Nonlinear two-dimensional elastic inversion of multi-offset seismic data 52(9):1211–1228. <https://doi.org/10.1190/1.1442384>
- Nihei KT, Li X (2007) Frequency response modelling of seismic waves using finite difference time domain with phase sensitive detection (TD—PSD). Geophys J Int 169(3):1069–1078. <https://doi.org/10.1111/j.1365-246X.2006.03262>
- Operto S, Gholami Y, Prieux V, Ribodetti A, Brossier R, Metivier L, Virieux J (2013) A guided tour of multiparameter full-waveform inversion with multicomponent data: from theory to practice. Lead Edge 32(9):1040–1054. <https://doi.org/10.1190/tle32091040.1>
- Pratt RG, Shin C, Hick GJ (1998) Gauss-Newton and full Newton methods in frequency–space seismic waveform inversion. Geophys J Int 133(2):341–362. <https://doi.org/10.1046/j.1365-246X.1998.00498.x>
- Prieux V, Brossier R, Operto S, Virieux J (2013) Multiparameter full waveform inversion of multicomponent ocean-bottom-cable data from the Valhall field Part 2: imaging compressive-wave and shear-wave velocities. Geophys J Int 194(3):1665–1681. <https://doi.org/10.1093/gji/ggt178>
- Ren Z, Liu Y (2016) A hierarchical elastic full-waveform inversion scheme based on wavefield separation and the multistep-length approach. Geophysics 81(3):R99–R123. <https://doi.org/10.1190/geo2015-0431.1>
- Richardson A (2018) Seismic full-waveform inversion using deep learning tools and techniques. *arXiv preprint arXiv:1807.07232*.
- Sears TJ, Singh SC, Barton PJ (2008) Elastic full waveform inversion of multi-component OBC seismic data. Geophys. Prospect 56(6):843–862. <https://doi.org/10.1111/j.1365-2478.2008.00692.x>
- Shin C, Jang S, Min DJ (2001) Improved amplitude preservation for prestack depth migration by inverse scattering theory. Geophys Prospect 49(5):592–606. <https://doi.org/10.1046/j.1365-2478.2001.00279.x>
- Shipp RM, Singh SC (2002) Two-dimensional full wavefield inversion of wide-aperture marine seismic streamer data. Geophys J Int 151(2):325–344. <https://doi.org/10.1046/j.1365-246X.2002.01645.x>
- Sirgue L, Etgen JT, Albertin U (2008) 3D frequency domain waveform inversion using time domain finite difference methods. Google Pat. <https://doi.org/10.3997/2214-4609.20147683>
- Sun J, Niu Z, Innanen KA, Li J, Trad DO (2020) A theory-guided deep-learning formulation and optimization of seismic waveform inversion Theory-guided DL and seismic inversion. Geophysics 85(2):R87–R99
- Sun J, Innanen KA, Huang C (2021) Physics-guided deep learning for seismic inversion with hybrid training and uncertainty analysis. Geophysics 86(3):R303–R317. <https://doi.org/10.1190/geo2020-0312.1>
- Tarantola A (1986) A strategy for nonlinear elastic inversion of seismic reflection data. Geophysics 51(10):1893–1903. <https://doi.org/10.1190/1.1442046>
- Wang TF, Cheng JB (2017) Elastic full waveform inversion based on mode decomposition: the approach and mechanism. Geophys J Int 209(2):606–622. <https://doi.org/10.1093/gji/ggx038>
- Wang W, McMechan GA, Ma J (2021) Elastic isotropic and anisotropic full-waveform inversions using automatic differentiation for gradient calculations in a framework of recurrent neural networks. Geophysics 86(6):795–810. <https://doi.org/10.1190/geo2020-0542.1>
- Wu Y, Lin Y (2019) InversionNet: an efficient and accurate data-driven full waveform inversion. IEEE Trans Comput Imaging 6:419–433
- Wu Y, McMechan GAJG (2019) Parametric convolutional neural network-domain full-waveform inversion. Geophysics 84(6):881–896
- Xu K, McMechan GAJG (2014) 2D frequency-domain elastic full-waveform inversion using time-domain modeling and a multistep-length gradient approach. Geophysics 79(2):41–53
- Yang F, Ma J (2019) Deep-learning inversion: a next-generation seismic velocity model building method. Geophysics 84(4):R583–R599. <https://doi.org/10.1190/geo2018-0249.1>
- Zhang T, Sun J, Innanen KA and Trad D (2021) Numerical analysis of a deep learning formulation of elastic full waveform inversion with high order total variation regularization in different parameterization. *arXiv preprint arXiv:2101.08924*
- Zhu W, Xu K, Darve E, Beroza G (2021) A general approach to seismic inversion with automatic differentiation: Computers. Comput Eosci 151:104751
- Zhu W, Xu K, Darve E, Biondi B, Beroza GCJG (2022) Integrating deep neural networks with full-waveform inversion: Reparameterization, regularization, and uncertainty quantification. Geophysics 87(1):93–109

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.