

# Deep-learning inversion: a next generation seismic velocity-model building method

Fangshu Yang<sup>1</sup>, Jianwei Ma<sup>1</sup>

<sup>1</sup>*Center of Geophysics, Department of Mathematics and Artificial Intelligence Laboratory,  
Harbin Institute of Technology, Harbin, China, E-mail: yfs2016@hit.edu.cn;  
jma@hit.edu.cn*

## ABSTRACT

Seismic velocity is one of the most important parameters used in seismic exploration. Accurate velocity models are key prerequisites for reverse-time migration and other high-resolution seismic imaging techniques. Such velocity information has traditionally been derived by tomography or full-waveform inversion (FWI), which are time consuming and computationally expensive, and they rely heavily on human interaction and quality control. We investigate a novel method based on the supervised deep fully convolutional neural network (FCN) for velocity-model building (VMB) directly from raw seismograms. Unlike the conventional inversion method based on physical models, the supervised deep-learning methods are based on big-data training rather than prior-knowledge assumptions. During the training stage, the network establishes a non-linear projection from the multi-shot seismic data to the corresponding velocity models. During the prediction stage, the trained network can be used to estimate the velocity models from the new input seismic data. One key characteristic of the deep-learning method is that it can automatically extract multi-layer useful features without the need for human-curated activities and initial velocity setup. The data-driven method usually requires more time during the training stage, and actual predictions take less time, with only seconds needed. Therefore, the computational time of geophysical inversions, including real-time inversions, can be dramatically reduced once a good generalized network is built. By using numerical experiments on synthetic models, the promising performances of our proposed method are shown in comparison with conventional FWI even when the input data are in more realistic scenarios. Discussions on the deep-learning methods, training dataset, lack of low frequencies, and advantages and disadvantages of the new method are also provided.

## INTRODUCTION

Currently, velocity-model building (VMB) is an essential step in seismic exploration because it is used during the entire course of seismic exploration including seismic data acquisition, processing, and interpretation. Accurate subsurface-image reconstruction from surface seismic wavefields requires precise knowledge of the local propagation velocities between the recording location and the image location at depth. Good velocity models are prerequisites for reverse-time migration (Baysal et al., 1983) and other seismic imaging techniques (Biondi, 2006). Estimated velocity models can also be used as initial models to recursively generate high-resolution velocity models with optimization algorithms (Tarantola, 2005). Many of the explored techniques such as migration velocity analysis (Al-Yahya and Kamal, 1989), tomography (Chiao and Kuo, 2001), and full-waveform inversion (FWI) (Tarantola, 1984; Mora, 1987; Virieux and Operto, 2009) share the same purpose of building more accurate velocity models.

Traditional tomography methods (Woodward et al., 2008), including reflection tomography, tuning-ray tomography, and diving-wave tomography (Stefani, 1995), are widely used for migration of seismic reflection data in building three-dimensional (3D) subsurface velocity models. Such methods have worked sufficiently well in most cases. Seismic inversion is performed by means of wave inversion of a simple prior model of the subsurface, and by using a back propagation loop to infer the subsurface geological structures (Tarantola, 1984). Typically, FWI is a data-fitting procedure used in reconstructing high-resolution velocity models of the subsurface, as well as other parameters that govern wave propagation, from the full information contained in seismic data (Virieux and Operto, 2009; Operto et al., 2013). In FWI, active seismic sources are used to generate seismic waves, and geophones are placed on the surface to record the measurements. An inverse problem is formulated to combine the measurements with the governing physics equations to obtain the model parameters. Numerical optimization techniques are utilized to solve for the velocity models. When a suitably accurate initial model is provided, FWI is highly effective for obtaining

a velocity structure through iterative updates. Efforts have been made recently to overcome the limitations in FWI. Even though these conventional methods have shown great success in many applications, they can be limited in some situations owing to a lack of low-frequency components as well as computational inefficiency, subjective human factors, and other issues. Additionally, iterative refinement is expensive when used in the entire workflow. Thus, a robust, efficient, and accurate velocity-estimation method is needed to address these problems.

Machine learning (ML) is a field of artificial intelligence that uses statistical techniques to give computer systems the ability to “learn” from big data. ML has shown its strength in many fields including image recognition, recommendation systems ([Bobadilla et al., 2013](#)), spam filters ([Androutsopoulos et al., 2000](#)), fraud alerts ([Ravisankar et al., 2011](#)), and other applications. Furthermore, ML has a long history of applications in geophysics. Nonlinear intelligent inverse technologies have been applied since the mid-1980s. [Röth and Tarantola \(1994\)](#) first presented an application of neural networks to invert from the time-domain seismic data to a depth profile of acoustic velocities. They used pairs of synthetic shot gathers (i.e., a set of seismograms obtained from a single source) and corresponding one-dimensional (1D) velocity models to train a multi-layer feed-forward neural network with the goal of predicting the velocities from new recorded data. They showed that the trained network can produce high-resolution approximations to the solutions of the inverse problem. In addition, their method can invert the geophysical parameters in the presence of white noise. [Nath et al. \(1999\)](#) used neural networks for cross-well travelttime tomography. After training the network with synthetic data, the velocities can be topographically estimated by the trained network with the new cross-well data. In recent years, most ML-based methods have focused mainly on pattern recognition in seismic attributes ([Zeng, 2004; Zhao et al., 2015](#)) and facies classifications in well logs ([Hall and Brendon, 2016](#)). In the work of [Guillen et al. \(2015\)](#), the authors proposed a novel workflow to detect salt bodies based on seismic attributes in a supervised learning method. An ML algorithm (i.e., Extremely Random Trees Ensemble) was used to train the mapping for automatically identifying salt regions. They

concluded that ML is a promising mechanism for classifying salt bodies when the selected training dataset has a sufficient capacity for describing the complex decision boundaries.

[Jia and Ma \(2017\)](#) used ML with supported vector regression ([Cortes and Vapnik, 1995](#)) for seismic data interpolation. Unlike the conventional methods, no assumptions are imposed on ML-based interpolation problems. On the basis of the above work, [Jia et al. \(2018\)](#) proposed a method based on the Monte Carlo method ([Yu et al., 2016](#)) for intelligent reduction of training sets. In that study, representative patches of seismic data were selected to train the method for efficient reconstructions.

Deep learning (DL) ([LeCun et al., 2015](#); [Goodfellow et al., 2016](#)), a new branch of machine learning, has drawn widespread interest by showing outstanding performance for recognition and classification ([Greenspan et al., 2016](#)) in image and speech processing. Recently, [Zhang et al. \(2014\)](#) proposed to use of a kernel regularized least-squares method ([Evgeniou et al., 2000](#)) for fault detection from seismic records. The authors used toy velocity models to generate seismic records and set the records and the velocity models as inputs and labels in the training set. The numerical experiments showed that this method obtained meaningful results. [Wang et al. \(2018a\)](#) developed a salt-detection technique from raw multi-shot gathers by utilizing a fully convolutional neural network (FCN). The testing performance showed that salt detections is much faster and efficient by this method than traditional migration and interpretation. [Lewis and Vigh \(2017\)](#) investigated a combination of DL and FWI to improve the performance for salt inversion. In that study, the network was trained to generate useful prior models for FWI by learning features relevant to earth model building from a seismic image. The authors tested this methodology by generating a probability map of salt bodies in the migrated image and incorporating it in the FWI objective function. The test results showed that this method is promising in enabling automated salt body reconstruction using FWI. In the work of [Araya-Polo et al. \(2017\)](#), a deep neural network (DNN)-based statistical model was used to automatically predict faults directly from synthetic two-dimensional (2D) seismic data. Inspired by this concept, [Araya-Polo et al. \(2018\)](#) proposed an approach for VMB. One key element of this

DL tomography is the use of a feature based on semblance that predigests the velocity information. Extracted features are obtained before the training process and are used as DNN inputs to train the network. [Mosser et al. \(2018\)](#) used a generative adversarial network ([Goodfellow et al., 2014](#)) with cycle-constraints ([Zhu et al., 2017](#)) to perform seismic inversion by formulating this problem as a domain-transfer problem. The mapping between the post-stack seismic traces and P-wave velocity models was approximated through this learning method. Before training the network, the seismic traces were transformed from the time domain to the depth domain based on the velocity models. Thus, the inputs and the outputs for training were in same domain. Most research has focused on identifying features and attributes in migrated images; few studies have discussed VMB or velocity inversion.

Multi-layer neural networks are computational learning architectures that propagate the input data across a sequence of linear operators and simple non-linearities. In this system, a deep convolutional neural network (CNN), proposed by [LeCun et al. \(2010\)](#), is implemented with linear convolutions followed by non-linear activation functions. A strong motivation to use FCN stems from the universal approximation theorem ([Hornik, 1991; Csaji, 2001](#)), which states that a feed-forward network with a single hidden layer containing a finite number of neurons can approximate any continuous function on compact subsets under a mild assumption on the activation functions. Additionally, FCN assumes that we learn representative features by convolutional kernels in a data-driven fashion to extract features automatically. Compared with DNN, FCN exhibits structures with fewer parameters to explain multi-layer perceptions while still providing good results ([Burger et al., 2012](#)).

In this study, we proposed the use of FCN to reconstruct subsurface parameters, i.e., P-wave velocity model, directly from raw seismic data, instead of performing a local-based inversion with respect to the subsurface represented through a grid. This method is an alternative formulation to conventional FWI and includes two processes. During the training process, multi-shot gathers are fed into the network together, and the network effectively approximates the non-linear mapping between the data and the corresponding velocity model. During the prediction process, the trained network can be saved to obtain unknown geo-

morphological structures only with new seismic data. Compared with traditional methods, less human intervention or no initial velocity models are involved throughout the process. Although the training process is expensive, the cost of the prediction stage by the network is negligible once the training is completed. Alternatively, our proposed method provides a possible method for velocity inversion when the seismic data are in more realistic cases such as in the presence of noise and with a lack of low frequencies. Moreover, numerical experiments are used to demonstrate the applicability and feasibilities of our method.

This paper is organized as follows. In the method section, we present a brief introduction to the basic inversion problem, the concepts of FCN, the mathematical framework, and the special architecture of the network. In the results section, we firstly show the dataset design. Two types of velocity models are discussed: a simulated dataset generated by the authors and an open experimental dataset of the Society of Exploration Geophysics (SEG). In addition, we compare the testing performance of our proposed method with that of conventional methods (i.e., FWI). In the discussion section, we present several open questions related to the utilization of our method for geophysical application. In the conclusion section, summaries of this study are presented, and future work is outlined. All of the acronyms used in this paper are listed in Table 1.

## FCN-BASED INVERSION METHOD

### **Basic inversion problem**

The constant density 2D acoustic wave equation is expressed as

$$\frac{1}{v^2(x,z)} \frac{\partial^2 u(x,z,t)}{\partial t^2} = \nabla^2 u(x,z,t) + s(x,z,t), \quad (1)$$

where  $(x,z)$  denotes the spatial location,  $t$  represents time,  $v(x,z)$  is the velocity of the longitudinal wave at the corresponding location,  $u(x,z,t)$  is the wave amplitude,  $\nabla^2(\cdot) = \frac{\partial^2(\cdot)}{\partial x^2} + \frac{\partial^2(\cdot)}{\partial z^2}$  represents the Laplace operator, and  $s(x,z,t)$  is the source signal.

Equation 1 is usually given by

$$u = H(v), \quad (2)$$

where the operator  $H(\cdot)$  maps  $v$  to  $u$ , and is usually nonlinear.

The classical inversion methods aim at minimizing the following objective function:

$$\bar{v} = \arg \min_v f(v) = \arg \min_v \frac{1}{2} \|H(v) - d\|_2^2, \quad (3)$$

where  $d$  denotes the measured seismic data,  $\|\cdot\|_2$  is the  $l_2$  norm, and  $f(\cdot)$  represents the data-fidelity residual.

In many applications, the engine for solving the above equation is to develop a fast and reasonably accurate inverse operator  $H^{-1}$ . An adjoint-state method (Plessix, 2006) is used to compute the gradient  $g(v) = \nabla f(v)$ , and iterative optimization algorithms are used to minimize the objective function. Owing to the nonlinear properties of the operator  $H$  and the imperfection of the surveys  $d$ , it is difficult to obtain precise subsurface models. Therefore, minimizing the above equation is generally an ill-posed problem, and the solutions are non-unique and unstable. If  $d$  contains full-waveform information, the above equation presents an FWI.

## A review of the FCN

Many DL algorithms are built with CNNs and provide state-of-the-art performance in challenging inverse problems such as image reconstruction (Schlemper et al., 2017), super-resolution (Dong et al., 2016), X-ray-computed tomography (Jin et al., 2017), and compressive sensing (Adler et al., 2017). They are also studied as neuro-physiological models of vision (Anselmi et al., 2016).

The FCN, proposed by Long et al. (2015) in the context of image and semantic segmentation, changes the fully connected layers of the CNN into convolutional layers to achieve

end-to-end learning. Figure 1 shows a sketch of a simple FCN. In this example, migrated seismic data are used as input, which is followed by a convolutional layer. Then, a pooling layer is inserted in the middle. After application of max-pooling, the sizes of the feature maps change to the previous one-half. Afterward, the transposed convolutional operation is applied to enlarge the size of the output to be the same as that of the input. Ultimately, we used a soft-max function to obtain the expected label, which is a label indicates which pixels belong to the salt structure in the migrated data. This FCN method can be described as follows:

$$y = Net(x; \Theta) = S(K_2 * (M(R(K_1 * x + b_1))) + b_2), \quad (4)$$

where  $Net(\cdot)$  denotes an FCN-based network and also indicates the nonlinear mapping of the network, and  $x, y$  denotes the inputs and outputs of the network, respectively.  $\Theta = \{K_1, K_2, b_1, b_2\}$  is the set of parameters to be learned, including the convolutional weights ( $K_1$  and  $K_2$ ) and the bias ( $b_1$  and  $b_2$ ).  $R(\cdot)$  introduces the nonlinear active function, such as the rectified linear unit (Dahl et al., 2013), sigmoid, or exponential linear unit (Clevert et al., 2015).  $M(\cdot)$  denotes the subsampling function (e.g., max-pooling, average pooling).  $*$  is the convolutional operation, and  $S(\cdot)$  represents the soft-max function.

## Mathematical framework

With the goal of estimating velocity models using seismic data as inputs directly, the network needs to project seismic data from the data domain  $(x, t)$  to the model domain  $(x, z)$ , as shown in Figure 2. The basic concept of the proposed method is to establish the map between inputs and outputs, which can be expressed as

$$\tilde{v} = Net(d; \Theta), \quad (5)$$

where  $d$  is the raw unmigrated seismic data, and  $\tilde{v}$  denotes the P-wave velocity model predicted by the network. Our method contains two stages: the training process and the prediction process, as shown in Figure 3. Before the training stage, many velocity models

are generated and are used as outputs. The supervised network needs pairs of datasets. Therefore, the acoustic wave equation is applied as a forward model to generate the synthetic seismic data, which are used as inputs. Following the initial computation, the input–output pairs, which are named  $\{d_n, v_n\}_{n=1}^N$ , are input to the network for learning the mapping.

During the training stage, the network learns to fit a nonlinear function from the input seismic data to the corresponding ground-truth velocity model. Therefore, the network learns by solving the optimization problem as

$$\widehat{\Theta} = \arg \min_{\Theta} \frac{1}{mN} \sum_{n=1}^N L(v_n, Net(d_n; \Theta)), \quad (6)$$

where  $m$  represents the total number of pixels in one velocity model, and  $L(\cdot)$  is a measure of the error between ground-truth values  $v_n$  and prediction values  $\tilde{v}_n$ . In our numerical experiments, the  $l_2$  norm is applied for measuring the discrepancy.

For updating the learned parameters  $\Theta$ , the optimization problem can be solved by using back propagation and stochastic gradient-descent algorithms (SGD) ([Shamir and Zhang, 2013](#)). The number of training datasets is large, and the numerical computation of the gradient  $\nabla_{\Theta} L(d; \Theta)$  is not feasible based on our GPU memory. Therefore, to approximate the gradient, the mini-batch size  $h$  was applied for calculating  $L_h$ , i.e., the error between the prediction values and the corresponding ground-truth values of a small subset of the whole training dataset, in each iteration. This led to the following optimization problem:

$$\widehat{\Theta} = \arg \min_{\Theta} \frac{1}{mh} L_h = \arg \min_{\Theta} \frac{1}{mh} \sum_{n=1}^h \|v_n - Net(d_n; \Theta)\|_2^2, \quad (7)$$

Here, the ground-truth velocity models  $v_n$  are given during the training process but are unknown during testing. One epoch is defined when an entire training dataset is passed forward and backward through the neural network once. The training dataset is first shuffled into a random order and is subsequently chosen sequentially in mini-batches to ensure one pass. It should be noted that the loss function is different from that (equation 3) in FWI,

in which the loss measures the squared difference between the observed and simulated seismograms. In our case, we used the Adam algorithm ([Kingma and Ba, 2014](#)), i.e., a deformation of the conventional SGD algorithm. The parameters are iteratively updated as follows:

$$\Theta_{t+1} = \Theta_t - \delta g\left(\frac{1}{mh} \nabla_{\Theta} L_h(d_n; \Theta; v_n)\right), \quad (8)$$

where  $\delta$  is the positive step size, and  $g\left(\frac{1}{mh} \nabla_{\Theta} L_h(d_n; \Theta; v_n)\right)$  denotes a function. This algorithm is straightforward to implement, computationally efficient, and well suited for large problems in terms of data or parameters.

The network is built once the training process is completed. During the prediction stage, other unknown velocity models are obtained by the available learned network. In our work, the input seismic data for prediction is also synthetic seismic traces. In a real situation, however, the input is field data. The method can be calculated by algorithm [1](#).

## Architecture of the network

To achieve automatic seismic VMB from the raw seismic data, we adopted and modified the UNet ([Ronneberger et al., 2015](#)) architecture, which is a specific network built upon the concept of the FCN. Figure [4](#) shows the detailed architecture of the proposed network. It consists of a contracting path (left) used to capture the geological features and a symmetric shape of an expanding path (right) that enables precise localization. This symmetric form is an encoder–decoder structure and employing a contraction–expansion structure based on the max-pooling and the transposed convolution. The effective receptive field of the network increases as the input goes deeper into the network, when a fixed size convolutional kernel ( $3 * 3$  in our case) is given. The numbers of channels in the left path are 64, 128, 256, 512, and 1024, as the network depth increases. Skip layers are adopted to combine the local, shallow feature maps in the right path with the global, deep feature maps in the left path. We summarize the definitions of the different operations in Table [3](#), where  $K$  and  $\bar{K}$  denote the convolutional kernels. The mean and standard deviation in the batch normalization

---

**Algorithm 1** FCN-based inversion method

---

**Input:**  $\{d_n\}_{n=1}^N$ : seismic data,  $\{v_n\}_{n=1}^N$ : velocity models,  $T$ : epochs,  $lr$ : learning rate of network,  $h$ : batch size,  $num$  : number of training sets

**Given notation:**  $*$  : 2D convolution with channels including zero-padding,  $*\uparrow$  : 2D deconvolution (transposed convolution),  $R(\cdot)$  : rectified linear unit,  $B(\cdot)$  : batch normalization,  $M(\cdot)$  : max-pooling,  $C(\cdot)$  : copy and concatenate,  $\Theta = \{K, b\}$  : learnable parameters,  $L$ : loss function,  $Adam$ : SGD algorithm

**Initialize:**  $t = 1$ ,  $loss = 0.0$ ,  $y_0 = d$

### 1.Training process

1. Generate different velocity models that have similar geological structures.
2. Synthesize seismic data using the finite-difference scheme.
3. Input all data pairs into the network and use the Adam algorithm to update the parameters.

```

for t=1:1:T and (data, models) in traing set do
    for j=1:1:num/h do
        for i=1:1:l-1 do
             $y_i \leftarrow B(R(K_{(2i-1)} * y_{i-1} + b_{(2i-1)}))$ 
             $m_i \leftarrow B(R(K_{(2i)} * y_i + b_{(2i)}))$ 
             $y_i \leftarrow M(m_i)$ 
        end for
         $y_l \leftarrow B(R(K_{(2l-1)} * y_{l-1} + b_{(2l-1)}))$ 
         $y_l \leftarrow B(R(K_{(2l)} * y_l + b_{(2l)}))$ 
        for i=l-1:-1:1 do
             $y_i \leftarrow B(R(K_{(2l+3(l-i)-2)} * \uparrow y_{i+1} + b_{(2l+3(l-i)-2)}))$ 
             $m_i \leftarrow B(R(K_{((2l+3(l-i)-1))} * C(y_i, m_i) + b_{(2l+3(l-i)-1)}))$ 
             $y_i \leftarrow B(R(K_{(2l+3(l-i))} * m_i + b_{(2l+3(l-i))}))$ 
        end for
         $\tilde{v} \leftarrow B(R(K_{(5l-2)} * y_1 + b_{(5l-2)}))$ 
         $loss = L_h(\tilde{v}, v)$ 
    end for
     $\Theta_{j+1} \leftarrow Adam(\Theta_j; lr; loss)$ 
end for

```

### 2.Prediction process

1. Synthesize seismic data for different velocity models in the same way as that used for generating the training seismic data.
2. Input new seismic data into the learned network for prediction.

**Output:** Predicted velocity model  $v^*$

---

were calculated per dimension over the mini-batch.  $\varepsilon$  is a value added to the denominator for numerical stability,  $\gamma$  and  $\beta$  are also learnable parameters; however, they were not used in our method.

We made two main modifications to the original UNet to fit the seismic VMB. First, the original UNet, proposed in the image-processing community, reads input images in RGB color channels that represent the information from the input images. To process the seismic data, we assigned different shot gathers, generated at different source locations but from the same model as channels for the input. Therefore, the number of input channels is the same as the number of sources for each model. The multi-shot seismic data were fed into the network together to improve data redundancy. Second, in a usual UNet, the outputs and inputs are in the same (image) domain. However, for our goal, we expected the network to realize the domain projection, i.e., to transform the data from the  $(x, t)$  domain to the  $(x, z)$  domain and to build the velocity model simultaneously. To complete this, the size of feature maps obtained by the final  $3 * 3$  convolution was truncated to be the same size as the velocity model, and the channel of the output layer was modified to 1. This was done so that the neural network could train itself during the contracting and expanding processes to map the seismic data to the exact velocity model directly. The main body of the network is similar to that in the original UNet, and 23 convolutional layers in total are used in the network.

## NUMERICAL EXPERIMENTS AND RESULTS

In this section, the data preparation, including the model (output) design and data (input) design for training and testing datasets, is first presented. Subsequently, we use the simulated training dataset to train the network for velocity inversion, and we predict other unknown velocity models by the valuable learned network. Further, for SEG salt model training, the trained network for simulated models is regarded as the initialization; this pre-trained network is a common approach used in transfer learning ([Pan and Yang, 2010](#)).

The testing process on the SEG dataset is also performed. We compare the numerical results between our method and FWI. The numerical experiments are performed on an HP Z840 workstation with a Tesla K40 GPU, 32 Core Xeon CPU, 128 GB RAM, and an Ubuntu operating system that implements PyTorch (<http://pytorch.org>).

## Data preparation

To train an efficient network, a suitable large-scale training set, i.e., input–output pairs, is needed. In a typical FCN model, training outputs are provided by some of the labeled images. In this paper, 2D synthetic models are utilized for testing the data-driven method. Two types of velocity models are provided for numerical experiments: 2D simulated models and 2D SEG salt models extracted from a 3D salt model ([https://wiki.seg.org/wiki/SEG/EAGE\\_3D\\_modelling](https://wiki.seg.org/wiki/SEG/EAGE_3D_modelling)). Each velocity model is unique.

### *Training dataset*

*Model (output) design:* To explore and prove the available capabilities of DL for seismic waveform inversion, we first generated random velocity models with smooth interface curvatures and increased the velocity values with depth. For the sake of simplicity, we assumed that each model had 5 to 12 layers as the background velocity and that the velocity values of each layer ranged arbitrarily from 2000 m/s to 4000 m/s. A salt body with an arbitrary shape and position was embedded into each model, each having a constant velocity value of 4500 m/s. The size of each velocity model used  $x \times z = 201 \times 301$  grid points with a spatial interval  $\Delta x = \Delta z = 10$  m. Figure 5 shows 12 models from the simulated training dataset, and Figure 7(a) shows 6 examples of the testing dataset. In our work, the simulated training dataset contained 1600 velocity samples. To better apply our new method for inversion, a 3D salt velocity model from the SEG reference website was utilized for obtaining the 2D salt models. This type of model had the same size as the simulated models, and the values ranged from 1500 m/s to 4482 m/s. Figure 6 shows the 12 representative examples of the SEG salt models from the training dataset, and Figure 7(b) shows the 6 models from the

testing dataset. Owing to the limited extraction, 130 velocity models were included in SEG training dataset.

*Data (input) design:* To solve the acoustic wave equation, we used the time-domain stagger-grid finite-difference scheme that adopts a second-order time direction and eighth-order space direction (Ozdenvar and McMechan, 1997; Hardi and Sanny, 2016). For each velocity model, 29 sources were evenly placed, and shot gathers were simulated sequentially. The recording geometry consisted of 301 receivers evenly placed at a uniform spatial interval. The detailed parameters for forward modeling are shown in Table 2. The perfectly matched layer (PML) (Komatitsch and Tromp, 2003) absorbing boundary condition was adopted to reduce unphysical reflection on the left, right, and bottom edges. Additionally, to verify the stability of our method, we added Gaussian noise, with zero mean and standard derivation of 5%, to each testing seismic data. Moreover, we made the amplitude of the seismic data two times higher. The noisy or magnified data were also used as inputs and were fed into the network to invert the velocity values.

#### *Testing dataset*

The ground-truth velocity models of the testing dataset had geological structures similar to those of the training dataset owing to the usage of supervised learning method. All of the velocity models for prediction were not included in the training dataset and were unknown in the prediction process. The input seismic data for prediction were also obtained by using the same method as that used for generating the inputs for the training dataset. For simulated models and SEG salt models, the testing dataset was composed of 100 and 10 velocity samples, respectively.

### **Inversion for simulated dataset**

The first inversion case was performed for 2D simulated velocity models. During the training stage, the training batch for each epoch was constructed by randomly choosing 10 samples of velocity-model dimension  $201 \times 301$  from the training dataset. In each pair data,

the dimension of one-shot seismic data was downsampled to  $400 \times 301$ . The network deemed to work better was selected when the hyper-parameters were set as shown in Table 4 based on the training dataset and experimental guidance (Bengio, 2012). The mean squared error between the prediction velocity values and ground-truth velocity values is shown in Figure 9(a). Figure 10(j)–Figure 10(l) show three exemplified results of the proposed method. Visually, a generally good match was achieved between the predictions and the corresponding ground-truth.

In this case study, a comparison between our method and FWI was performed. We used the same parameter setting as that used to generate the training seismic data for the time-domain forward modeling. Multi-scale frequency-domain inversion strategy (Sirgue et al., 2008) was adopted. The selected inversion frequencies were 2.5 Hz, 5 Hz, 10 Hz, 15 Hz, and 21 Hz, based on the research of Sirgue and Pratt (2004). An adjoint-state based gradient descending method was adopted in this experiment (Plessix, 2006). The observed data of FWI was same as the seismic data we used for prediction. In addition, the true velocity model smoothed by the Gaussian smooth function was taken as the initial velocity model, as shown in Figure 10(d)–Figure 10(f). The numerical experiments of FWI were performed on a computer cluster with four Tesla K80 GPU units, and a central operating system. Figure 10(g)–Figure 10(i) shows the results of FWI. All subfigures have the same colorbar, and the velocity value ranges from 2000 m/s to 4500 m/s. In this scenario, the FCN-based inversion method showed comparable results and preserved most of the geological structures.

To quantitatively analyze the accuracy of the predictions, we chose two horizontal positions,  $x = 900$  m and  $x = 2000$  m, and we plotted the prediction (blue), FWI (red), and ground-truth (green) velocity values in the velocity versus depth profiles shown in Figure 11. Most prediction values matched well with the ground-truth values. Moreover, in Figure 12, a comparison of the shot records of the 15th receiver is displayed, including the observed data using the ground-truth velocity model, reconstructed data obtained by simulating the inversion result of FWI, and reconstructed data obtained by forward modeling the prediction velocity model of our method. The reconstructed data with predictions obtained by

the proposed method also matched well with the observed data.

The process of FWI for one simulated model inversion incurred a GPU time of 37 min. In contrast, after training the FCN-based inversion network with 1078 min, the GPU time per prediction in our simulated models was only 2 s on the lower-equipment machine, which is more than 1000 times less than that for FWI.

To further validate the capability of the proposed method, more test results under more realistic conditions should be performed. Here, we present experiments in which seismic data are contaminated with random noise or the seismic amplitude is doubled.. The noisy data were generated by adding zero-mean Gaussian noise with a standard deviation of 5%. Figure 13(a)–Figure 13(c) show the prediction results using the proposed method with noisy inputs; examples are shown in second column of Figure 14. A comparison of the predictions (i.e., the results shown in Figure 10(j)–Figure 10(l)) and the clean inputs (e.g., examples shown in first column of Figure 14) revealed that our method still provides acceptable results. However, compared with the ground truth, some parts of the predictions are not close to the true values, particularly the superficial background layers. This may have been caused by perturbations. In future research, the sensitivity to other type noise will be considered such as coherent noise and multiples.

Similarly, to test the sensitivity to amplitude, another test was performed in which the amplitude of the testing seismic data was doubled; examples are shown in third column of Figure 14. In this test, the processed data were applied as the input for prediction; the performance comparison is displayed in Figure 13(d)–Figure 13(f). The prediction velocities using the processed inputs with higher amplitudes were consistent with the predictions using the original inputs. This is in compliance with the theoretical analysis and indicates that our proposed method achieves velocity inversion adaptively and stably.

## Inversion for SEG salt dataset

To further show the outstanding ability of the proposed method, the trained network for the simulated dataset was utilized as the initialized network, which is one approach of transfer learning to train the SEG salt dataset. In the training process, the number of epochs was set to 50, and each epoch had 10 training samples (i.e., training mini-batch size). The other hyper-parameters for learning were same as those used for simulated model inversion. Figure 9(b) shows the mean-squared error of the SEG training dataset. The loss converged to zero when only 130 models were used for training. Similar to the test above, a comparison was performed between the proposed method and FWI with the same algorithm as that used in the experiments for simulated models. In this FWI experiment, the selected inversion frequency was 2.5 Hz, and the other three values ranged from 5 Hz to 15 Hz with a uniform frequency interval of 5 Hz. The initial velocity models were also obtained by using the Gaussian smooth function shown in Figure 15(d)–Figure 15(f). Figure 15 describes all performances of the numerical experiments, in which all subfigures have the same colorbar, and the value is from 1500 m/s to 4500 m/s. The comparison results of velocity values in the velocity versus depth profiles are displayed in Figure 16. In this test, compared with FWI method, the proposed method yielded a slightly lower performance, which could be attributed to the small number of training datasets. However, the predictions using the pre-trained initialized network (i.e., transfer learning in our study) were better than those obtained using the random initialized network. Moreover, the results of our method could serve as the initial models for FWI or travel time tomography. They could also be used for on-site quality control during seismic survey development.

The additional experiments using SEG salt datasets under more realistic conditions are also performed. When the seismic data were contaminated with noise, as shown in the second column of Figure 18, most prediction values obtained by the inversion network shown in Figure 17(a)–Figure 17(c) were close to ground-truth velocities, but were slightly lower than the predictions obtained using the clean data. However, the prediction results

shown in Figure 17(d)–Figure 17(f) using the inputs with higher amplitudes, as shown in third column of Figure 18, made no differences. In this test, because the training dataset of the SEG salt models was less than that for the simulated models, the performance of the proposed method for the SEG dataset was not outstanding. Therefore, in future work, we will augment the diversity of the training set gradually and take advantage of the transfer learning to apply this novel method on other complex samples.

For one SEG salt velocity-model inversion, the process of FWI incurred a GPU time of 25 min. In comparison, the training time of our method for all model inversion was 43 min; the GPU time per prediction in the SEG salt dataset was 2 s on the lower-equipment machine. A comparison of the time consumed for training and prediction process is shown in Table 5.

In summary, the numerical experiments provide promising evidence for the feasibility of our proposed method for velocity inversion from the raw input of seismic shot gathers directly without the need for initial velocity models. This indicates that the neural network can effectively approximate non-linear mapping even when the inputs have perturbation. Compared with conventional FWI, the computational time of the proposed method is fast because it does not involve the iterations to search for optimal solutions. The main computational costs are incurred mostly during the training stage, which is only once during the model setup; this can be handled off-line in advance. After training, the prediction costs are negligible. Thus, the FCN-based inversion method makes the overall computational time a fraction of that needed for traditional physical-based inversion techniques.

## DISCUSSION

From an experimental perspective, the numerical results demonstrate that our proposed method presents promising capabilities of DL for velocity inversion. The objective of the research is to apply the latest breakthroughs in data science, particularly in DL techniques. Although the indications of our method are inspiring, many factors can affect its perfor-

mance, including the choice of training dataset; the selection of hyper-parameters such as learning rate, batch size, and training epoch; and the architecture of the neural network. For our purposes, we focused on the profound understanding of DL applied for seismic inversion. Therefore a discussion on more impressive results and the advantages and disadvantages of the new method is provided in this section.

*(a) How does the training dataset affect the network?*

The limitation of our approach is that the capability of the network relies on the dataset. In general, the models to be trained should involve structures or characteristics similar to those contained in the predictions. That is, the supervised learned network for prediction is limited to the choice of the training dataset, and the amount of data required for training depends on many factors. In most cases, a large amount of large-scale and diverse training samples results in a more powerful network. Moreover, the time consumed for training process is longer. One representative test of our proposed method for SEG salt models was conducted. In our experiments, predictions without salt are also presented. A comparison of the results between our method and FWI are shown in Figure 19. Our method yielded a lower performance than FWI. In particular, the sediment was vague because only 10 training samples without salt were utilized for training. Thus, the capability of the network to learn these models is lower than that for other salt models.

In addition, according to the simple geological structures contained in the simulated dataset, such as several smooth interfaces, increasing background velocity, and constant-velocity salt, relatively high similarity was noted between the training and testing datasets. According to experimental guidance ([Bengio, 2012](#)) and the other similar research ([Wang et al., 2018b; Wu et al., 2018](#)), the number of simulated training datasets was set to 1600. The effect of the of training dataset number on the network will be investigated in the future.

In training stage of the DL method, one needs a training dataset that including lots of training pairs. The observed data could be a 1-shot gather, or other number shot gathers. In our method, the shot numbers were fixed and were specific to the network such that 29

available shot seismic datasets were used for training the 29-shot network; in the prediction step, the same was considered for the 29 shot gathers. For further exploration, a comparison of performances with only 1-, 13-, 21-, 27-, and 29-shot training data is shown in Figure 20 owing to the constraint of the GPU memory. In Figure 20(a), the comparison mean loss revealed that all of the training losses in different cases converged to zero along the epoch number. This indicates that our proposed method can be applied with arbitrary training shots and may be an advantage over the traditional FWI. Figure 20(b)–Figure 20(d) show the testing performance for the mean loss, mean peak signal-to-noise ratio (PSNR), and mean structural similarity (SSIM). The 1-shot case displayed a little bit unstable. The quantitative results may be misleading, however, because all testing evaluations are obtained for 10 selected network during training stage and take their average. In our next work, we will investigate the effects of training shots to apply the novel method in a more realistic scenario.

(b) *How can we apply the network when a lack of low frequencies exists in the testing data?*

The lack of low frequencies in field data is a main problem for practical application of FWI. However in the ML or DL methods, it is possible to learn the “low frequency” from simulation data or prior-information data. Two other numerical experiments are provided to show the performance of our method. As shown in Figure 21, all of the training datasets have low-frequency information, which is same as the original information used for prediction. However, the low frequency (i.e., 0–1/10 normalized Fourier spectrum) of the testing seismic data is removed by Fourier transform and the Butterworth high-pass filter. Then, the reconstructed seismic data, shown in Figure 21(d), were used for prediction, the results of which are shown in Figure 21(b). In this case, the proposed method predicted most parts of the velocity model. A comparison of the prediction shown in Figure 21(a) with complete data shown in Figure 21(c) revealed that the structure boundaries are less clear, and the background velocity layers are somewhat vague, which could be attributed to the low-frequency information.

In addition, the performance of the supervised learning method relies on a training set. Therefore, the new training seismic dataset missing the low frequencies, which was processed by using the same approach as that used for the testing data (Figure 21(d)), and the corresponding ground-truth velocity models could be utilized to train the network. The prediction results are displayed in Figure 22. Visually, the predictions were slightly better than that shown in Figure 21 but were still lower than those with complete data.

*(c) Is the learned network robust and stable for any prediction?*

A general question often asked when learning is applied to some problems is whether the method can be generalized to other problems, e.g., whether a method trained on a specific dataset can be applied to another dataset without re-training. Thus far, it has been difficult to test complex (e.g., SEG salt models) or real models by using the trained network directly because the performance of our proposed method relies on the datasets, and similar distribution is relatively weak between the two types different velocity models. In our work, transfer learning (Pan and Yang, 2010), i.e., a research problem in ML that focuses on storing knowledge gained while solving one problem and applying it to a different but related problem, was applied when the new training models are similar to the simulated models. The goal of using the pre-trained network as an initialization is to more effectively show the nonlinear mapping between the inputs and outputs rather than just allowing the machine to remember the characteristics of the dataset. A comparison of the training loss versus the number of epochs between random initial networks (i.e., the same as parameter initialization in UNet) and a pre-trained initial network (i.e., trained network for a simulated dataset) is shown in Figure 23. The network learned better with the pre-trained initialization in the same computational time.

## CONCLUSION

In this study, we proposed a supervised end-to-end DL method in a new fashion for velocity inversion that presents an alternative to “conventional” FWI formulation. In the

proposed formulation, rather than performing local-based inversion with respect to subsurface parameters, we used a FCN to reconstruct these parameters. After a training process, the network is able to propose a subsurface model from only seismic data. The numerical experiments showed impressive results in the potential of the DL in seismic model building and clearly demonstrated that a neural network can effectively approximate the inverse of a non-linear operator that is very difficult to resolve. The learned network still computes satisfactory velocity profiles when the seismic data are under more realistic conditions. Compared with FWI, once the network training is completed, the reconstruction costs are negligible. Moreover, little human intervention is needed, and no initial velocity setup is involved. The loss function is measured in the model domain, and no seismograms are generated when using the network for prediction. In addition, no cycle-skipping problem exists.

The large-scale diverse training set plays an important role in the supervised learning method. Inspired by the success of transfer learning and generative adversarial learning in computer vision, and the combination of traditional methods and neural networks. We propose two possible directions for future work. The first is to generate more complex and realistic velocity models using a generative adversarial network, which is a type of semi-supervised learning network, based on the limited open dataset. Then, we can train the network with these complex datasets and apply the trained network to field data by transfer learning. The second is to uncover the potential relationship between conventional approaches for inversion and specific networks. This approach enables to develop novel network designs that can reveal the hidden wave-equation model and invert more complex geological structures based on the physical systems. Further studies are required to adopt these methods to large problems, field data, and other applications.

## ACKNOWLEDGEMENTS

The authors would like to thank the editors and reviewers for offering useful comments to improve this manuscript. Thanks are extended to Dr. Wenlong Wang for providing primary CNN code compiled with PyTorch and the suggestion of feeding multi-shot gathers into the network together to improve data redundancy. This work is supported in part by the National Key Research and Development Program of China under Grant 2017YFB0202902, NSFC under Grant 41625017 and Grant 91730306, and the China Scholarship Council.

## REFERENCES

- Adler, A., D. Boublil, and M. Zibulevsky, 2017, Block-based compressed sensing of images via deep learning: IEEE International Workshop on Multimedia Signal Processing (MMSP), 1–6.
- Al-Yahya, and Kamal, 1989, Velocity analysis by iterative profile migration: *Geophysics*, **54**, 718–729.
- Androuloutsopoulos, I., G. Paliouras, V. Karkaletsis, G. Sakkis, C. D. Spyropoulos, and P. Stamatopoulos, 2000, Learning to filter spam e-mail: A comparison of a naive bayesian and a memory-based approach: arXiv preprint cs/0009009.
- Anselmi, F., J. Z. Leibo, L. Rosasco, J. Mutch, A. Tacchetti, and T. Poggio, 2016, Unsupervised learning of invariant representations: *Theoretical Computer Science*, **633**, 112–121.
- Araya-Polo, M., T. Dahlke, C. Frogner, C. Zhang, T. Poggio, and D. Hohl, 2017, Automated fault detection without seismic processing: *The Leading Edge*, **36**, 208–214.
- Araya-Polo, M., J. Jennings, A. Adler, and T. Dahlke, 2018, Deep-learning tomography: *The Leading Edge*, **37**, 58–66.
- Baysal, E., D. Kosloff, and W. Sherwood, 1983, Reverse time migration: *Geophysics*, **48**, 1514–1524.
- Bengio, Y., 2012, Practical recommendations for gradient-based training of deep architectures: *437–478*.
- Biondi, B., 2006, 3D seismic imaging: Society of Exploration Geophysicists.
- Bobadilla, J., F. Ortega, A. Hernando, and A. Gutiérrez, 2013, Recommender systems survey: *Knowledge-Based Systems*, **46**, 109–132.
- Burger, H. C., C. J. Schuler, and S. Harmeling, 2012, Image denoising: Can plain neural networks compete with BM3D?: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2392–2399.
- Chiao, L., and B. Kuo, 2001, Multiscale seismic tomography: *Geophysical Journal International*, **145**, 517–527.
- Clevert, D.-A., T. Unterthiner, and S. Hochreiter, 2015, Fast and accurate deep network

- learning by exponential linear units (ELUs): arXiv preprint arXiv:1511.07289.
- Cortes, C., and V. Vapnik, 1995, Support vector networks: Machine Learning, **20**, 273–297.
- Csaji, B., 2001, Approximation with artificial neural networks: Faculty of Sciences, Etvs Lornd University, **24**.
- Dahl, G., T. Sainath, and G. Hinton, 2013, Improving deep neural netwrks for LVCSR using recitified linear units and dropout: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 8609–8613.
- Dong, C., C. C. Loy, K. He, and X. Tang, 2016, Image super-resolution using deep convolutional networks: IEEE Transactions on Pattern Analysis and Machine Intelligence, **38**, 295–307.
- Evgeniou, T., M. Pontil, and T. Poggio, 2000, Regularization networks and support vector machines: Advances in Computational Mathematics, **13**, 1–50.
- Goodfellow, I., Y. Bengio, and A. Courville, 2016, Deep Learning: MIT Press.
- Goodfellow, I., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, 2014, Generative adversarial networks: Advances in Neural Information Processing Systems, 2672–2680.
- Greenspan, H., B. van Ginneken, and R. M. Summers, 2016, Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique: IEEE Transactions on Medical Imaging, **35**, 1153–1159.
- Guillen, P., G. Larrazabal\*, G. González, D. Boumber, and R. Vilalta, 2015, Supervised learning to detect salt body: 85th Annual International Meeting, SEG, Expanded Abstracts, 1826–1829.
- Hall, and Brendon, 2016, Facies classification using machine learning: The Leading Edge, **35**, 906–909.
- Hardi, B. I., and T. A. Sanny, 2016, Numerical modeling : Seismic wave propagation in elastic media using finite-difference and staggered-grid scheme: Presented at the 41th HAGI Annual Convention and Exhibition.
- Hornik, K., 1991, Approximation capabilities of multilayer feedforward networks: Neural

- Networks, **4**, 251–257.
- Jia, Y., and J. Ma, 2017, What can machine learning do for seismic data processing? An interpolation application: *Geophysics*, **82**, V163–V177.
- Jia, Y., S. Yu, and J. Ma, 2018, Intelligent interpolation by Monte Carlo machine learning: *Geophysics*, **83**, V83–V97.
- Jin, K. H., M. T. McCann, E. Froustey, and M. Unser, 2017, Deep convolutional neural network for inverse problems in imaging: *IEEE Transactions on Image Processing*, **26**, 4509–4522.
- Kingma, D. P., and J. Ba, 2014, Adam: A method for stochastic optimization: arXiv preprint arXiv:1412.6980.
- Komatitsch, D., and J. Tromp, 2003, A perfectly matched layer absorbing boundary condition for the second-order seismic wave equation: *Geophysical Journal International*, **154**, 146–153.
- LeCun, Y., Y. Bengio, and G. Hinton, 2015, Deep learning: *Nature*, **521**, 436–444.
- LeCun, Y., K. Kavukcuoglu, and C. Farabet, 2010, Convolutional networks and applications in vision: *IEEE International Symposium on Circuits and Systems: Nano-Bio Circuit Fabrics and Systems, ISCAS*, 253–256.
- Lewis, W., and D. Vigh, 2017, Deep learning prior models from seismic images for full-waveform inversion: *SEG International Exposition and Annual Meeting, Society of Exploration Geophysicists*, 1512–1517.
- Long, J., E. Shelhamer, and T. Darrell, 2015, Fully convolutional networks for semantic segmentation: *IEEE conference on computer vision and pattern recognition*, 3431–3440.
- Mora, P., 1987, Nonlinear two-dimensional elastic inversion of multioffset seismic data: *Geophysics*, **52**, 1211–1228.
- Mosser, L., W. Kimman, J. Dramsch, S. Purves, A. De la Fuente, and G. Ganssle, 2018, Rapid seismic domain transfer: Seismic velocity inversion and modeling using deep generative neural networks: *80th EAGE Conference and Exhibition*.
- Nath, S. K., S. Chakraborty, S. K. Singh, and N. Ganguly, 1999, Velocity inversion in

- cross-hole seismic tomography by counter-propagation neural network, genetic algorithm and evolutionary programming techniques: *Geophysical Journal International*, **138**, 108–124.
- Operto, S., Y. Gholami, V. Prieux, A. Ribodetti, R. Brossier, L. Metivier, and Virieux, 2013, A guided tour of multiparameter full-waveform inversion with multicomponent data: From theory to practice: *The Leading Edge*, **32**, 1040–1054.
- Ozdenvar, T., and G. A. McMechan, 1997, Algorithms for staggered-grid computations for poroelastic, elastic, acoustic, and scalar wave equations: *Geophysical Prospecting*, **45**, 403–420.
- Pan, S. J., and Q. Yang, 2010, A survey on transfer learning: *IEEE Transactions on knowledge and data engineering*, **22**, 1345–1359.
- Plessix, R.-E., 2006, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications: *Geophysical Journal International*, **167**, 495–503.
- Ravisankar, P., V. Ravi, G. Raghava Rao, and I. Bose, 2011, Detection of financial statement fraud and feature selection using data mining techniques: *Decision Support Systems*, **50**, 491–500.
- Ronneberger, O., P. Fischer, and T. Brox, 2015, U-Net: Convolutional networks for biomedical image segmentation: *International Conference on Medical Image Computing and Computer-assisted Intervention*, 234–241.
- Röth, G., and A. Tarantola, 1994, Neural networks and inversion of seismic data: *Journal of Geophysical Research: Solid Earth*, **99**, 6753–6768.
- Schlemper, J., J. Caballero, J. V. Hajnal, A. Price, and D. Rueckert, 2017, A deep cascade of convolutional neural networks for MR image reconstruction: *International Conference on Information Processing in Medical Imaging*, 647–658.
- Shamir, O., and T. Zhang, 2013, Stochastic gradient descent for non-smooth optimization: Convergence results and optimal averaging schemes: *International Conference on Machine Learning*, 71–79.

- Sirgue, L., J. Etgen, and U. Albert, 2008, 3D frequency domain waveform inversion using time domain finite difference methods: Presented at the 70th European Association of Geoscientists and Engineers Conference and Exhibition, Expanded Abstracts.
- Sirgue, L., and R. G. Pratt, 2004, Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies: *Geophysics*, **69**, 231–248.
- Stefani, J., 1995, Turning ray tomography: *Geophysics*, **60**, 1917–1929.
- Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: *Geophysics*, **49**, 1259–1266.
- , 2005, Inverse problem theory and methods for model parameter estimation: siam, **89**.
- Virieux, J., and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: *Geophysics*, **74**, WCC1–WCC26.
- Wang, W., F. Yang, and J. Ma, 2018a, Automatic salt detection with machine learning: 80th European Association of Geoscientists and Engineers Conference and Exhibition, Extended Abstracts, 9–12.
- , 2018b, Velocity model building with a modified fully convolutional network: Society of Exploration Geophysicists, Technical Program Expanded Abstracts 2018, 2086–2090.
- Woodward, M. J., D. Nichols, O. Zdraveva, P. Whitfield, and T. Johns, 2008, A decade of tomography: *Geophysics*, **73**, VE5–VE11.
- Wu, Y., Y. Lin, and Z. Zhou, 2018, InversionNet: Accurate and efficient seismic waveform inversion with convolutional neural networks: Society of Exploration Geophysicists, Technical Program Expanded Abstracts 2018, 2096–2100.
- Yu, S., J. Ma, and S. Osher, 2016, Monte Carlo data-driven tight frame for seismic data recovery: *Geophysics*, **81**, V327–V340.
- Zeng, H., 2004, Seismic geomorphology-based facies classification: *The Leading Edge*, **23**, 644–688.
- Zhang, C., C. Frogner, M. Araya-Polo, and D. Hohl, 2014, Machine-learning based automated fault detection in seismic traces: 76th European Association of Geoscientists and

- Engineers Conference and Exhibition, Extended Abstracts, 807–811.
- Zhao, T., V. Jayaram, A. Roy, and K. J. Marfurt, 2015, A comparison of classification techniques for seismic facies recognition: Interpretation, **3**, SAE29–SAE58.
- Zhu, J., T. Park, P. Isola, and A. A. Efros, 2017, Unpaired image-to-image translation using cycle-consistent adversarial networks: IEEE International Conference on Computer Vision (ICCV), 2242–2251.

## LIST OF FIGURES

- 1 Sketch of a simple fully convolutional neural network (FCN) with a convolutional layer, a pooling layer and a transposed convolutional layer. Migrated data were adopted as the input, and the pixel-wise output includes salt and non-salt parts.
- 2 Schematic diagram depicting the velocity-model prediction from recorded seismic data by the fully convolutional neural network.
- 3 Flow chart of the FCN-based inversion process.
- 4 Architecture of the network used for seismic velocity inversion. Each blue and green cube corresponds to a multi-channel feature map. The number of channels is shown on bottom of the cube. The x-z size is provided at the lower left edge of the cube (example shown for  $25 \times 19$  in lower resolution). The arrows denote the different operations, and the size of the corresponding parameter set is defined in each box. The abbreviations shown in the explanatory frame, i.e., conv, max-pooling, BN, Relu, deconv and skip connection + concatenation, are defined in Table 3.
- 5 Twelve representative samples from 1600 simulated training velocity models.
- 6 Twelve representative samples from 130 SEG-salt training velocity models.
- 7 Typical samples from testing dataset for velocity inversion: (a) six velocity models of the simulated dataset; (b) six velocity models of the Society of Exploration Geophysics (SEG) dataset.
- 8 Six shots of the seismic data generated by the finite-difference scheme. The corresponding velocity model is the first model shown in Figure 5.
- 9 Loss decreases during the training process: (a) mean-squared error for the simulated velocity inversion; (b) mean-squared error for the SEG velocity inversion.
- 10 Comparisons of the velocity inversion (simulated models): (a)–(c) ground truth; (d)–(f) initial velocity model of FWI; (g)–(i) results of FWI; (j)–(l): prediction of our method.
- 11 Vertical velocity profiles of our method and FWI. For the three test samples given in Figure 10, the prediction, FWI, and ground-truth velocities in the velocity versus depth

profiles at two horizontal positions ( $x = 900$  m,  $x = 2000$  m) are presented in each row.

12 Shot records of the 15th receiver. Given from left to right in each row are the observed data according to the ground-truth velocity model (shown in Figure 10(a)–Figure 10(c)), reconstructed data by obtained by forward modeling of the FWI inverted velocity model (shown in Figure 10(g)–Figure 10(i)), and reconstructed data obtained by forward modeling the prediction of the FCN-based inversion method (shown in Figure 10(j)–Figure 10(l)).

13 Sensitivity of the proposed method to noise and amplitude (simulated models): (a)–(c) prediction results with the noisy seismic data; (d)–(f) prediction results with magnified seismic amplitude. Our method showed acceptable results when the input data were perturbed.

14 Comparison of records of simulated seismic data. Given in each row from left to right are original data, noisy data (with added Gaussian noise), and magnified data (to twice as large). The corresponding velocity models of each row are the three models shown in Figure 10(a)–Figure 10(c), respectively.

15 Comparisons of the velocity inversion (SEG salt models): (a)–(c) ground truth; (d)–(f) initial velocity model of FWI; (g)–(i) results of FWI; (j)–(l) prediction of our method.

16 Vertical velocity profiles of our method and FWI. The prediction, FWI, and the ground-truth velocities in the velocity versus depth profiles at two horizontal positions ( $x = 800$  m,  $x = 1700$  m) of the three test samples in Figure 15 are shown in each row.

17 Sensitivity of the proposed method to noise and amplitude (SEG models): (a)–(c) prediction results with the noisy seismic data; (d)–(f) prediction results with magnified seismic amplitude. For the open dataset, our method yielded acceptable predictions when the input data were perturbed.

18 Comparison of records of SEG seismic data. Given in each row from left to right are original data, noisy data (with added Gaussian noise), and magnified data (to twice as larger). The corresponding velocity models of each row are the three models shown in Figure 15(a)–Figure 15(c), respectively.

19 Inversion of velocity model without salt dome: (a) ground-truth velocity model; (b) initial velocity model of FWI; (c) result of FWI; (d) result of the proposed method. Our method showed a slightly lower performance than FW because only 10 training models without the salt body were utilized. More training data are required to obtain corresponding improvement.

20 Comparison of performance versus number of epoch between different numbers of training shots: (a) mean-square error during training stage; (b) mean square error during testing stage; (c) mean peak signal-to-noise ratio (PSNR) during the testing stage; (d) mean structural similarity (SSIM) during the testing stage. All testing evaluation was obtained for 100 testing models.

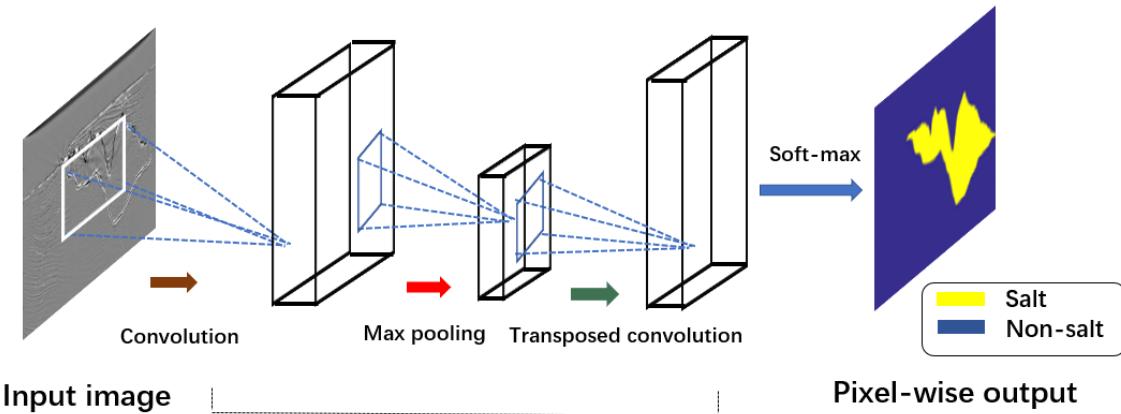
21 Typical results obtained by using our method when seismic data are lacking in low-frequency components: (a) ground truth; (b) prediction with data lacking low frequencies; (c) original seismic data (15th shot); (d) reconstructed data lacking 1/10 normalized Fourier spectrum.

22 Results of the velocity inversion obtained when the training data lack low frequencies: (a)–(c) ground truth of simulated models; (d)–(f) prediction results; (g)–(i) ground truth of SEG salt models; (j)–(l): prediction results.

23 Comparison of training loss versus number of epochs. The red line denotes training with a random initial network. The blue line represents training with a pre-trained initial network (i.e., the trained network for the simulated dataset).

**LIST OF TABLES**

- 1 All acronyms used in this paper and their definitions.
- 2 Parameters of forward modeling.
- 3 Definitions of the different operations for our proposed network.
- 4 Parameters of training process in our proposed network.
- 5 Time consumed for the training and testing processes. The three columns of each method from left to right indicate the GPU time for the simulated velocity-model inversion, and SEG salt-model inversion. The training time is the total time required for all training sets; the testing time is for only one model. N/A indicates that FWI had no training time.



### Fully convolutional layers

Figure 1: Sketch of a simple fully convolutional neural network (FCN) with a convolutional layer, a pooling layer and a transposed convolutional layer. Migrated data were adopted as the input, and the pixel-wise output includes salt and non-salt parts.

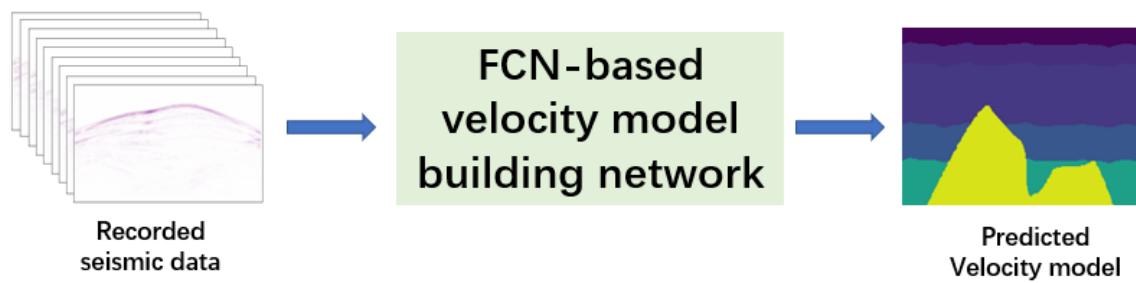


Figure 2: Schematic diagram depicting the velocity-model prediction from recorded seismic data by the fully convolutional neural network.

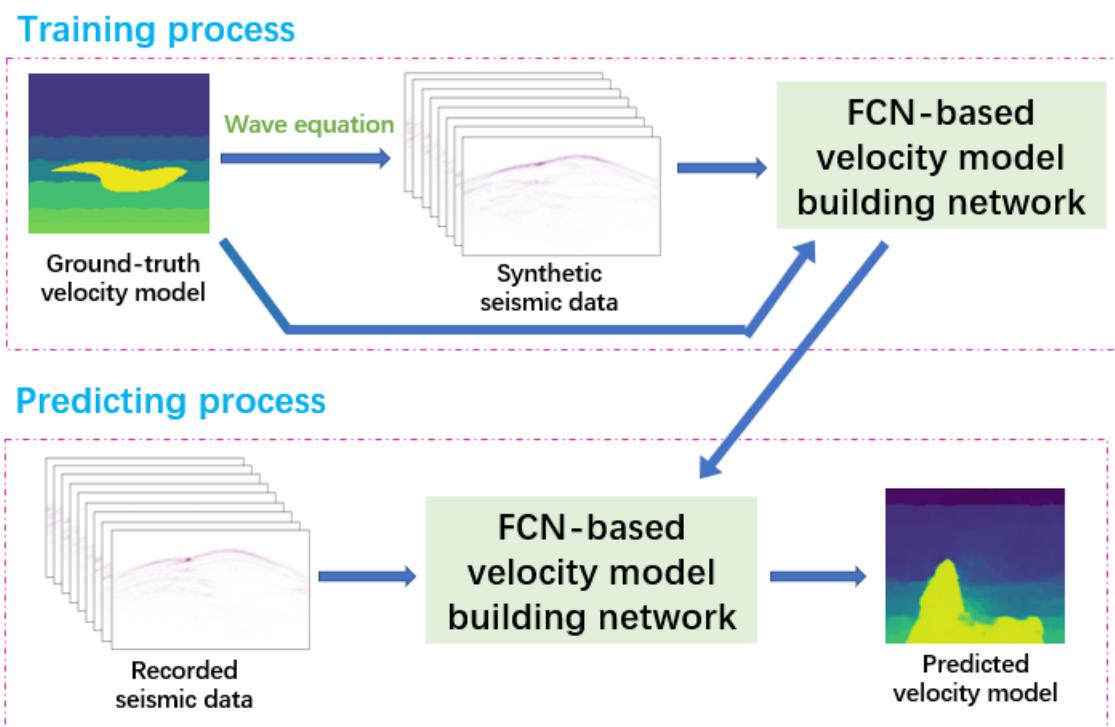


Figure 3: Flow chart of the FCN-based inversion process.

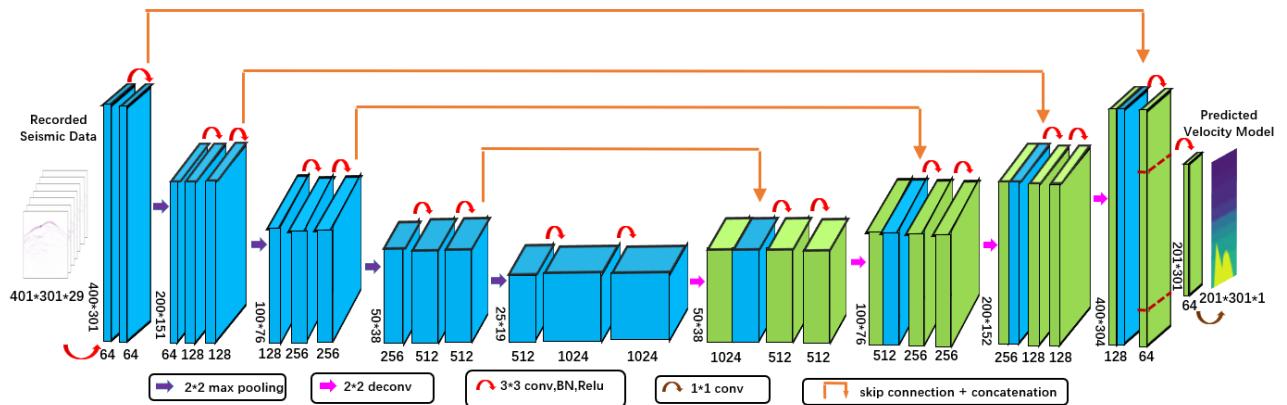


Figure 4: Architecture of the network used for seismic velocity inversion. Each blue and green cube corresponds to a multi-channel feature map. The number of channels is shown on bottom of the cube. The x-z size is provided at the lower left edge of the cube (example shown for  $25 \times 19$  in lower resolution). The arrows denote the different operations, and the size of the corresponding parameter set is defined in each box. The abbreviations shown in the explanatory frame, i.e., conv, max-pooling, BN, Relu, deconv and skip connection + concatenation, are defined in Table 3.

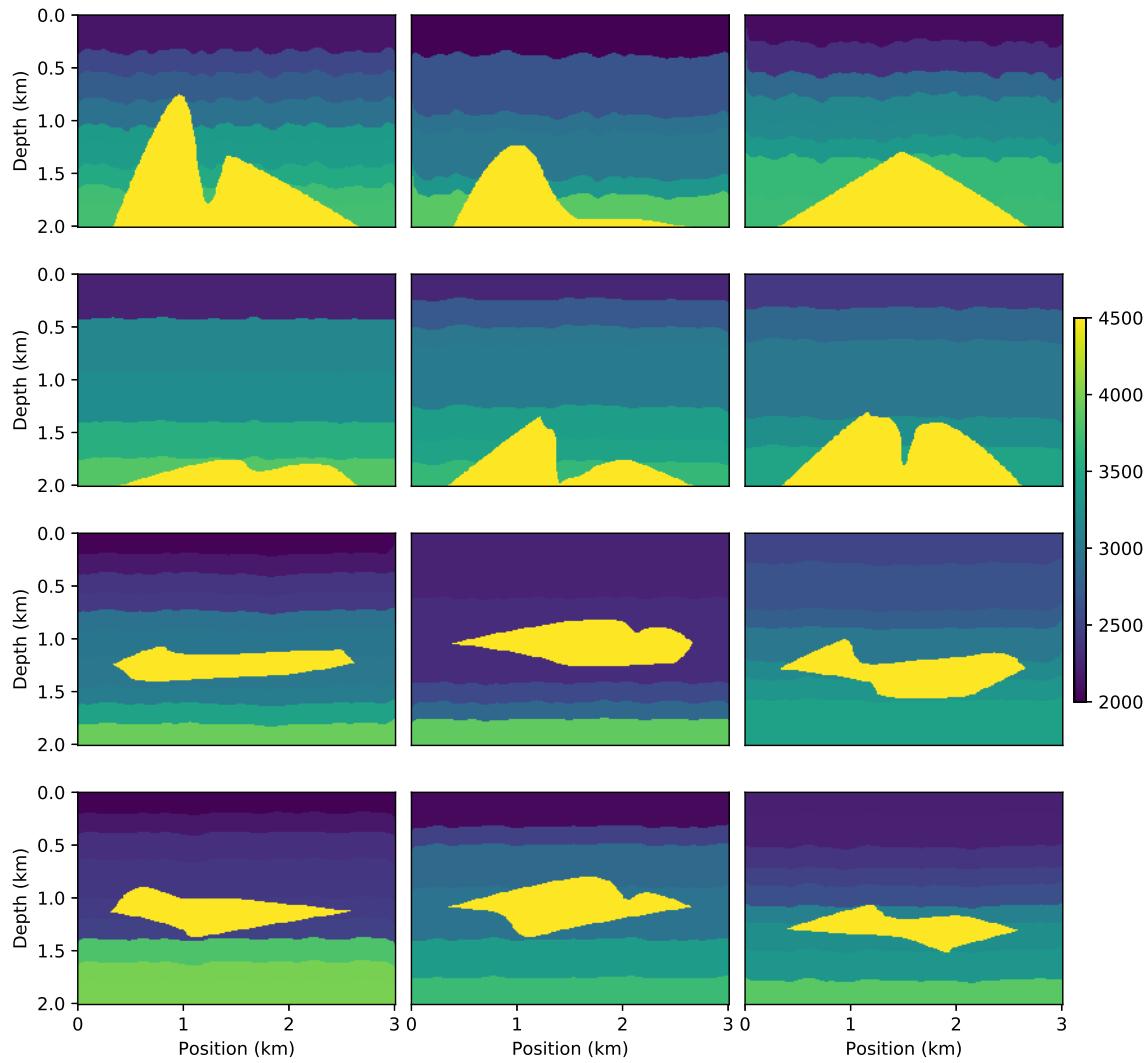


Figure 5: Twelve representative samples from 1600 simulated training velocity models.

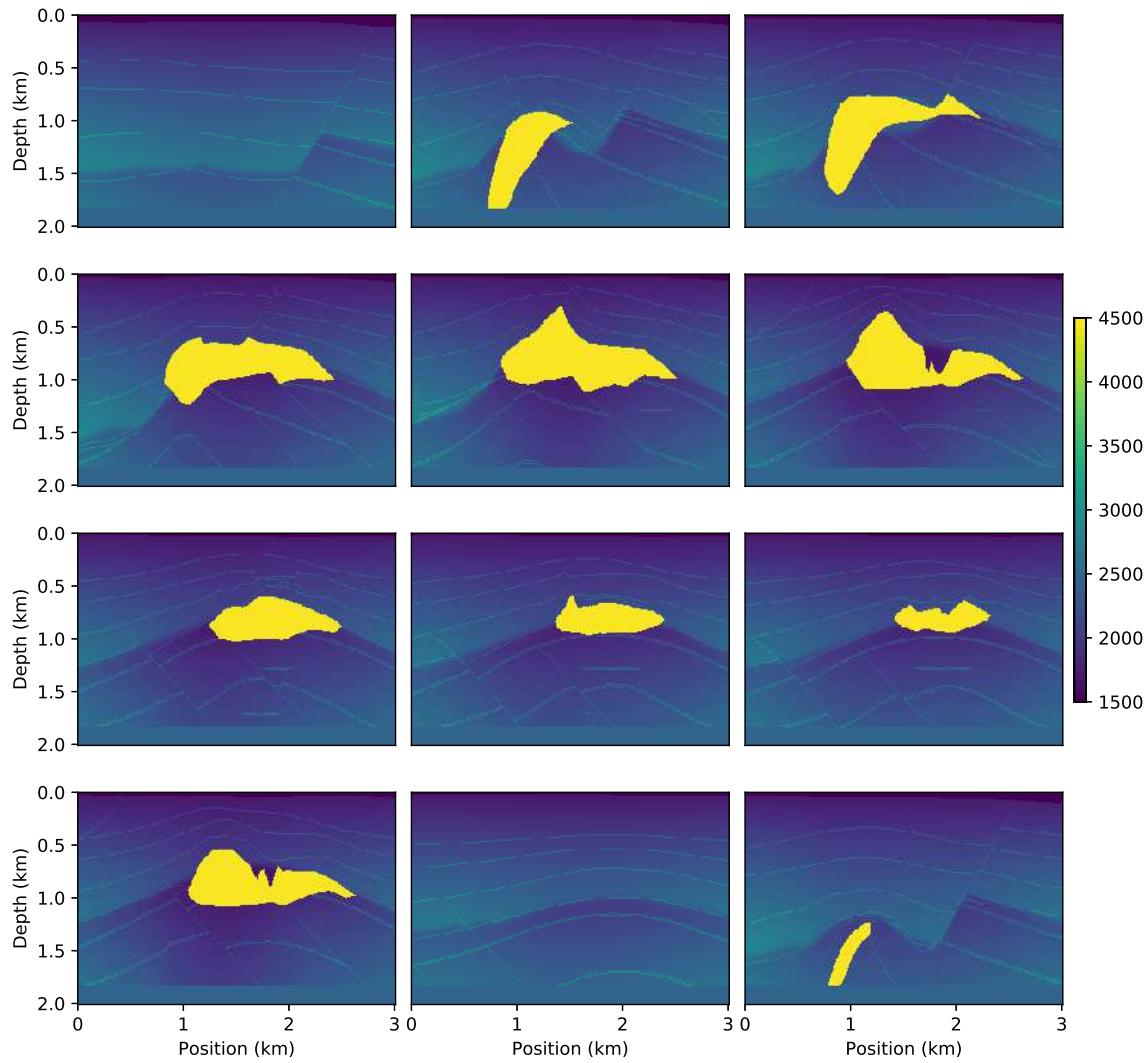


Figure 6: Twelve representative samples from 130 SEG-salt training velocity models.

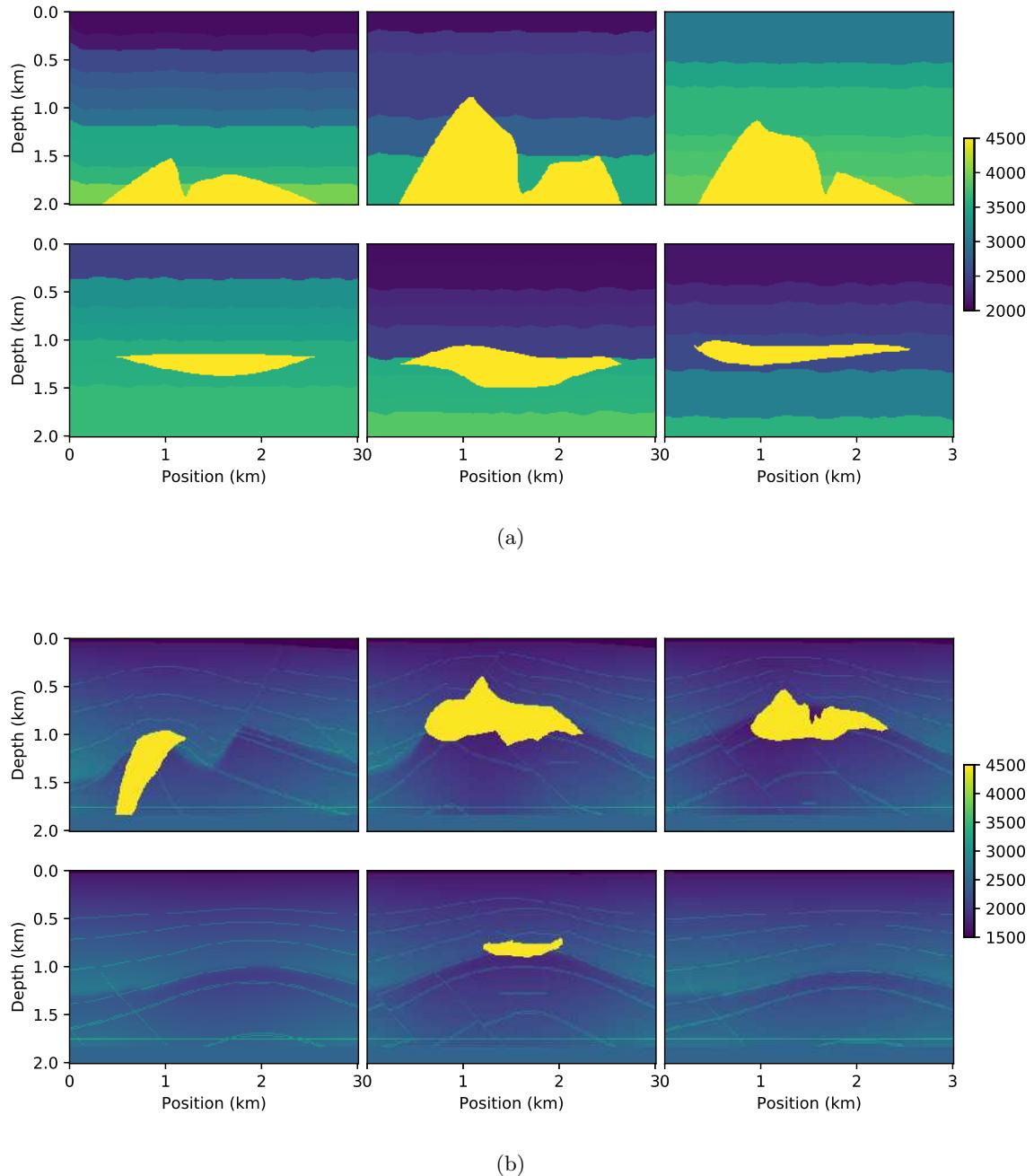


Figure 7: Typical samples from testing dataset for velocity inversion: (a) six velocity models of the simulated dataset; (b) six velocity models of the Society of Exploration Geophysics (SEG) dataset.

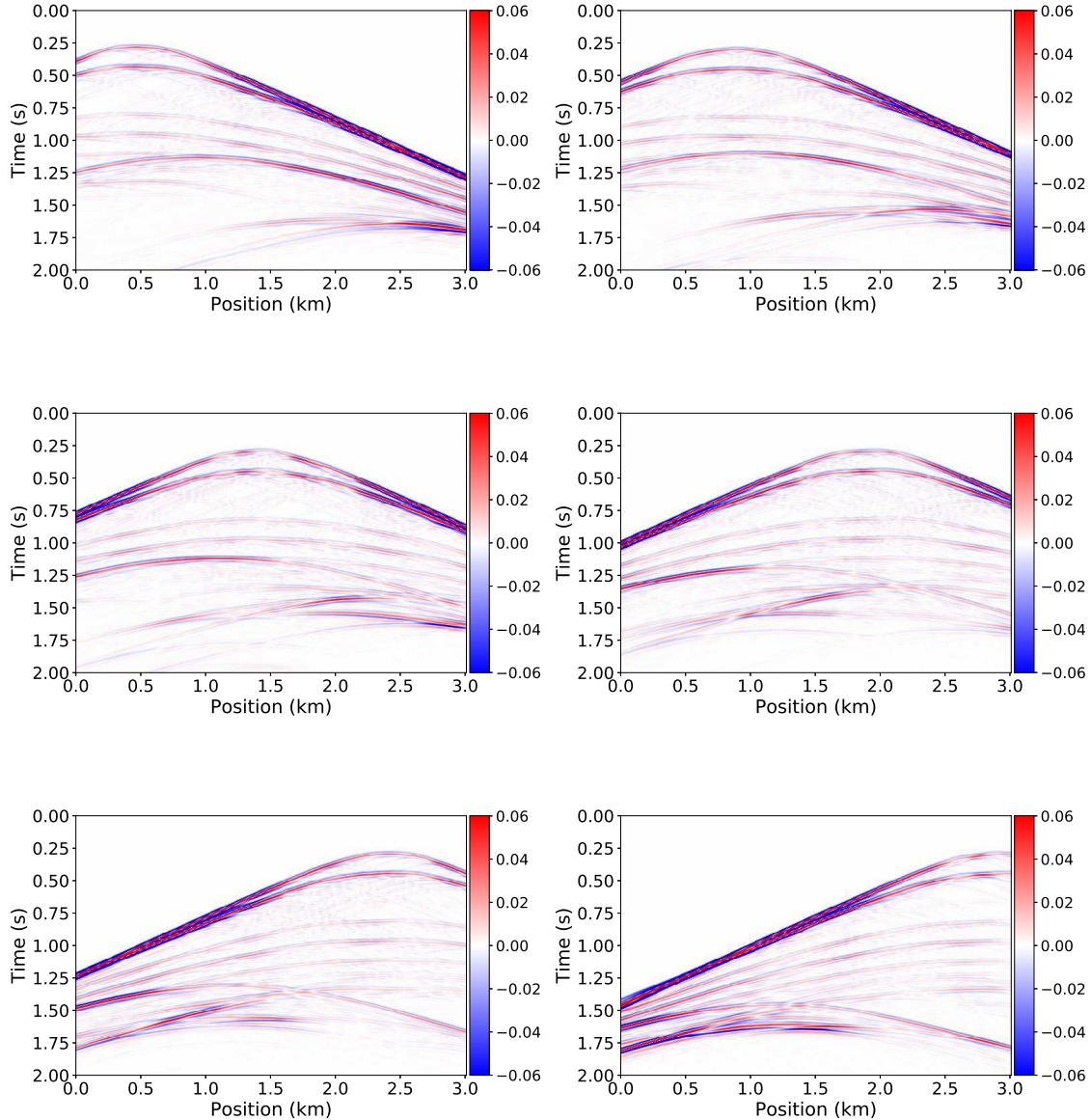


Figure 8: Six shots of the seismic data generated by the finite-difference scheme. The corresponding velocity model is the first model shown in Figure 5.

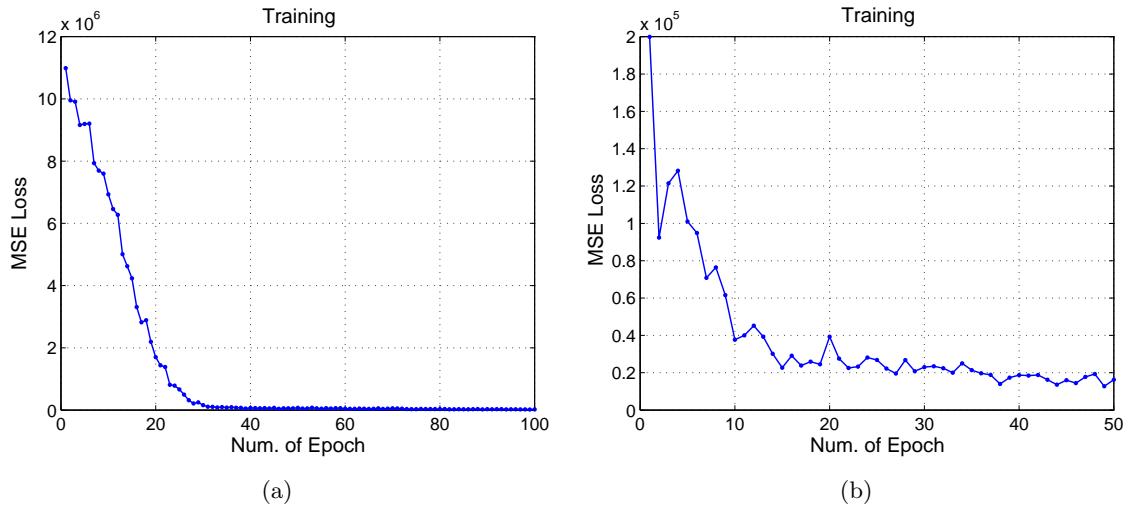


Figure 9: Loss decreases during the training process: (a) mean-squared error for the simulated velocity inversion; (b) mean-squared error for the SEG velocity inversion.

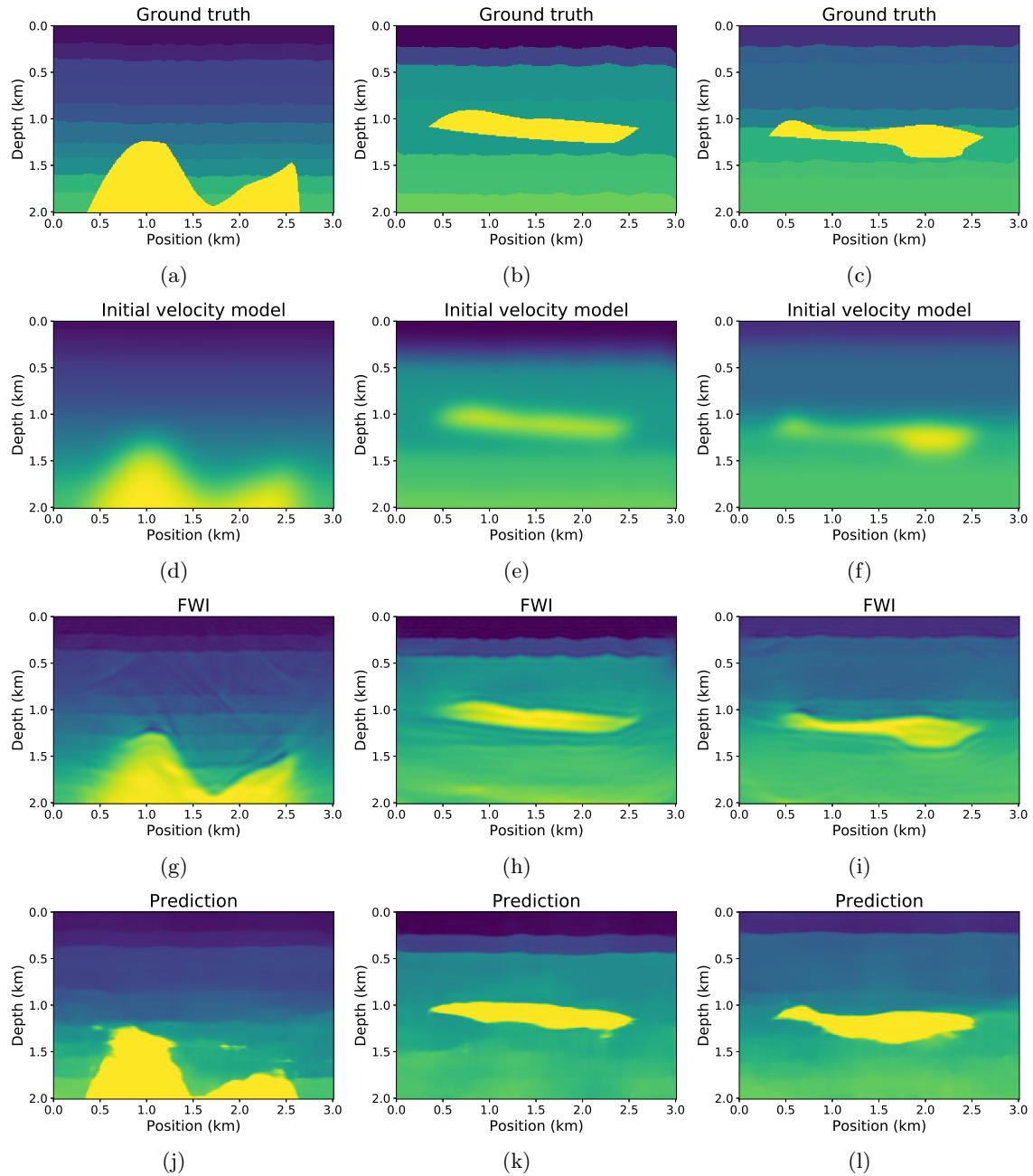


Figure 10: Comparisons of the velocity inversion (simulated models): (a)–(c) ground truth; (d)–(f) initial velocity model of FWI; (g)–(i) results of FWI; (j)–(l): prediction of our method.

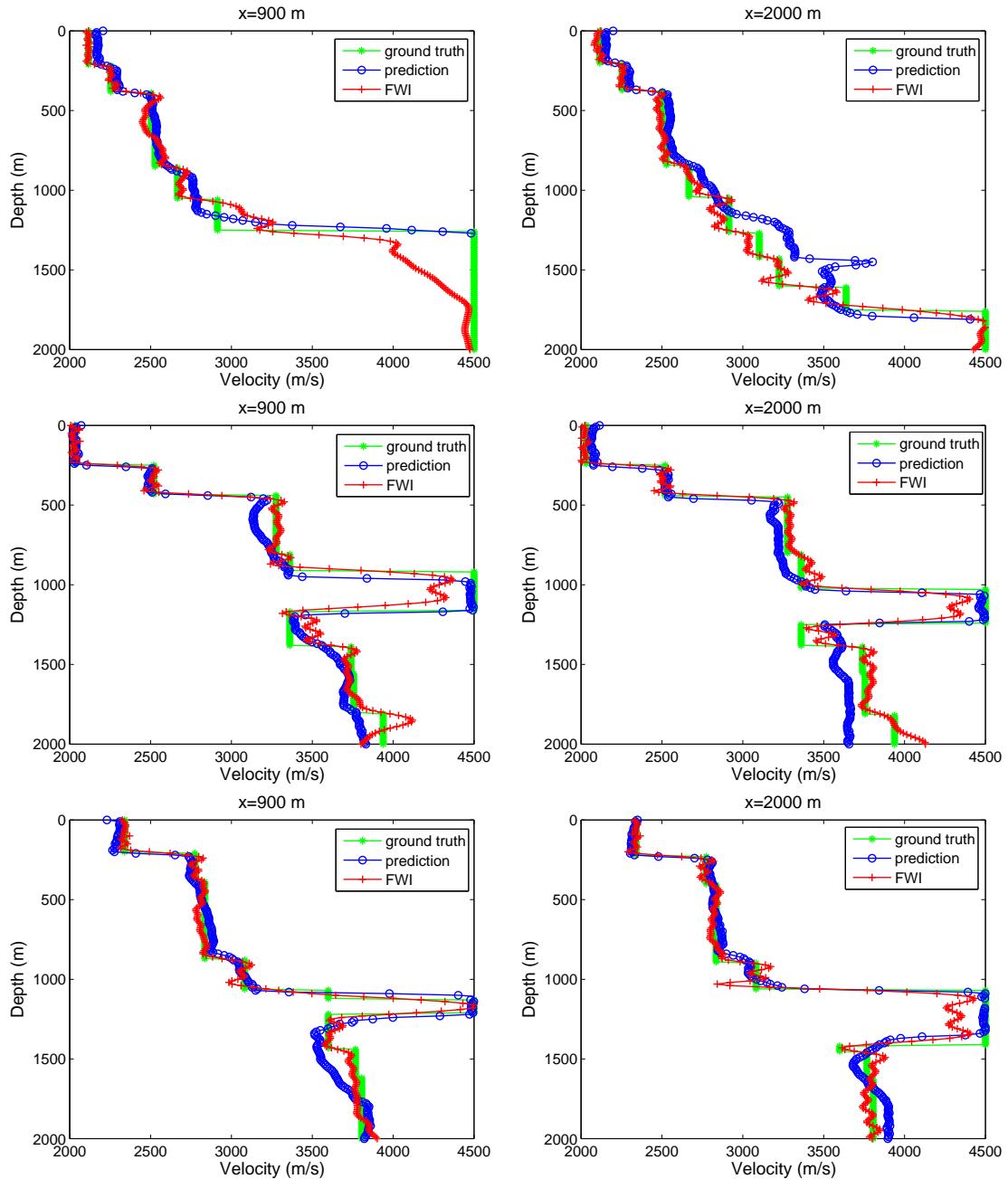


Figure 11: Vertical velocity profiles of our method and FWI. For the three test samples given in Figure 10, the prediction, FWI, and ground-truth velocities in the velocity versus depth profiles at two horizontal positions ( $x = 900$  m,  $x = 2000$  m) are presented in each row.

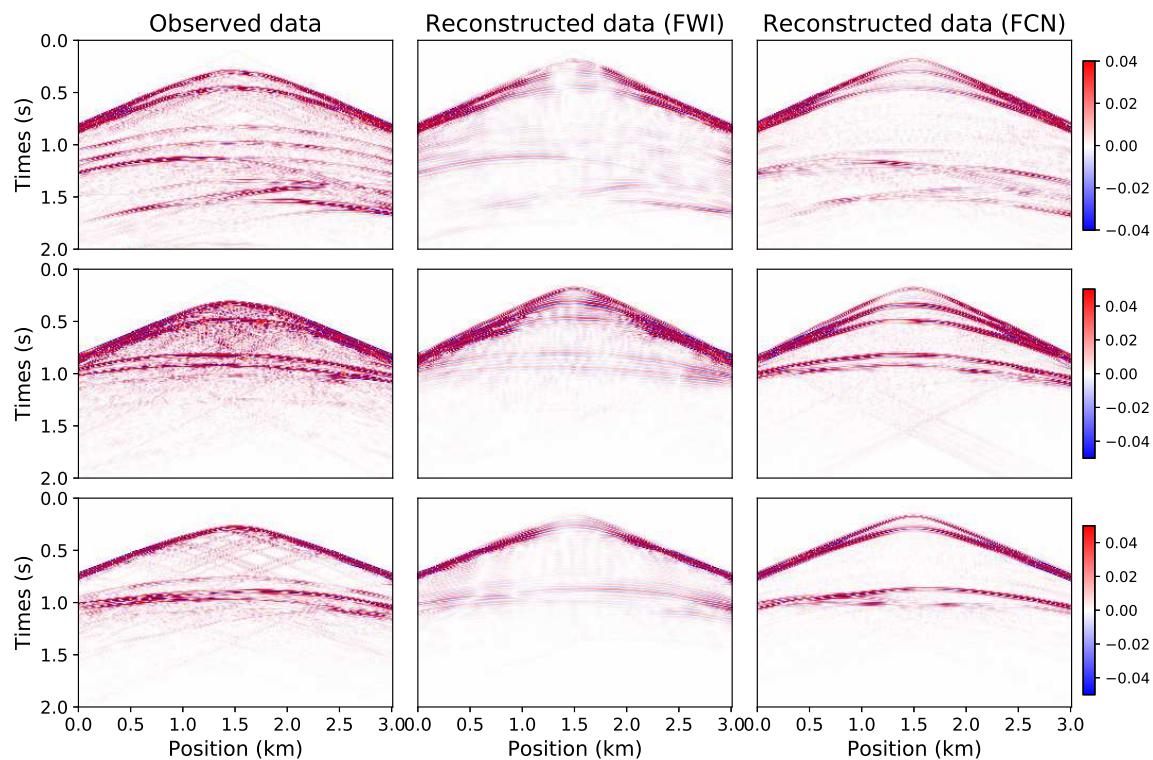


Figure 12: Shot records of the 15th receiver. Given from left to right in each row are the observed data according to the ground-truth velocity model (shown in Figure 10(a)–Figure 10(c)), reconstructed data by obtained by forward modeling of the FWI inverted velocity model (shown in Figure 10(g)–Figure 10(i)), and reconstructed data obtained by forward modeling the prediction of the FCN-based inversion method (shown in Figure 10(j)–Figure 10(l)).

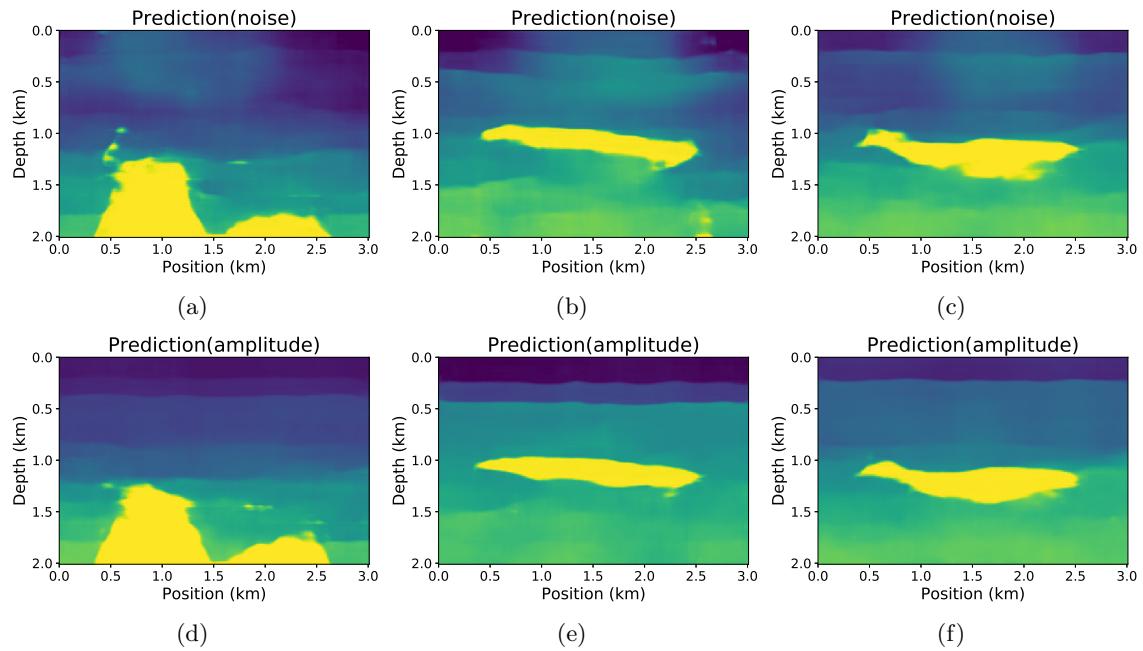


Figure 13: Sensitivity of the proposed method to noise and amplitude (simulated models): (a)–(c) prediction results with the noisy seismic data; (d)–(f) prediction results with magnified seismic amplitude. Our method showed acceptable results when the input data were perturbed.

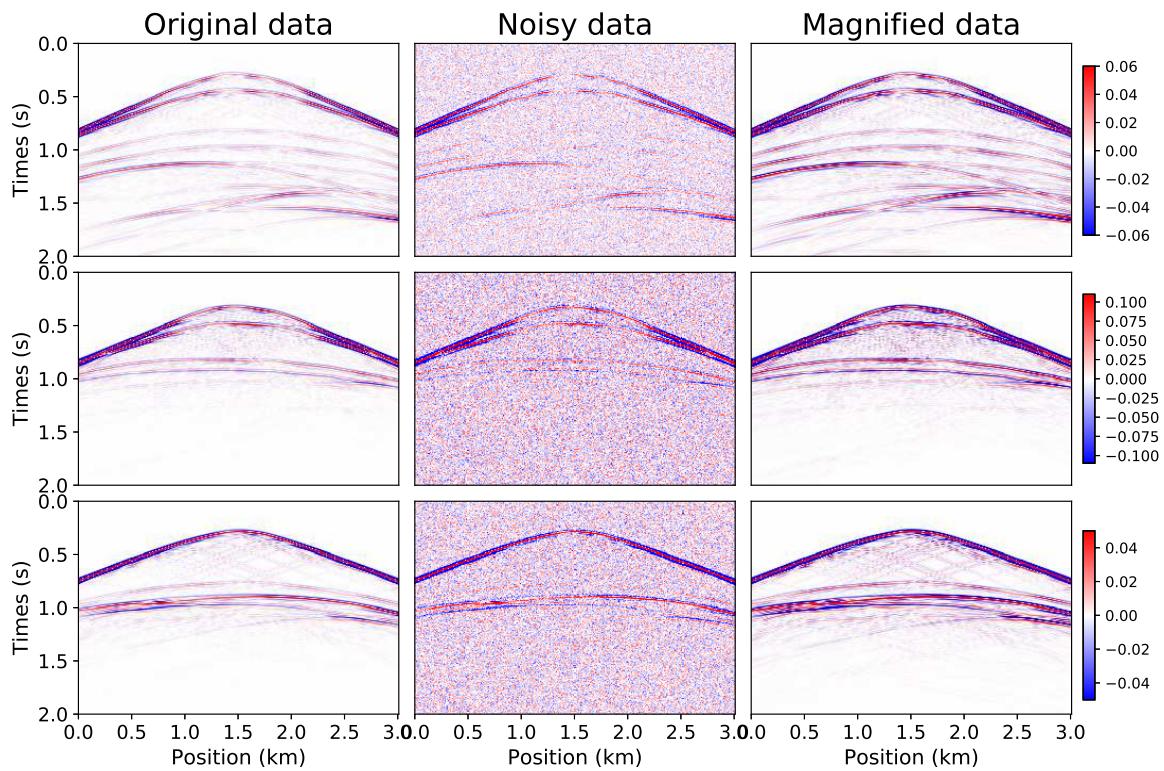


Figure 14: Comparison of records of simulated seismic data. Given in each row from left to right are original data, noisy data (with added Gaussian noise), and magnified data (to twice as large). The corresponding velocity models of each row are the three models shown in Figure 10(a)-Figure 10(c), respectively.

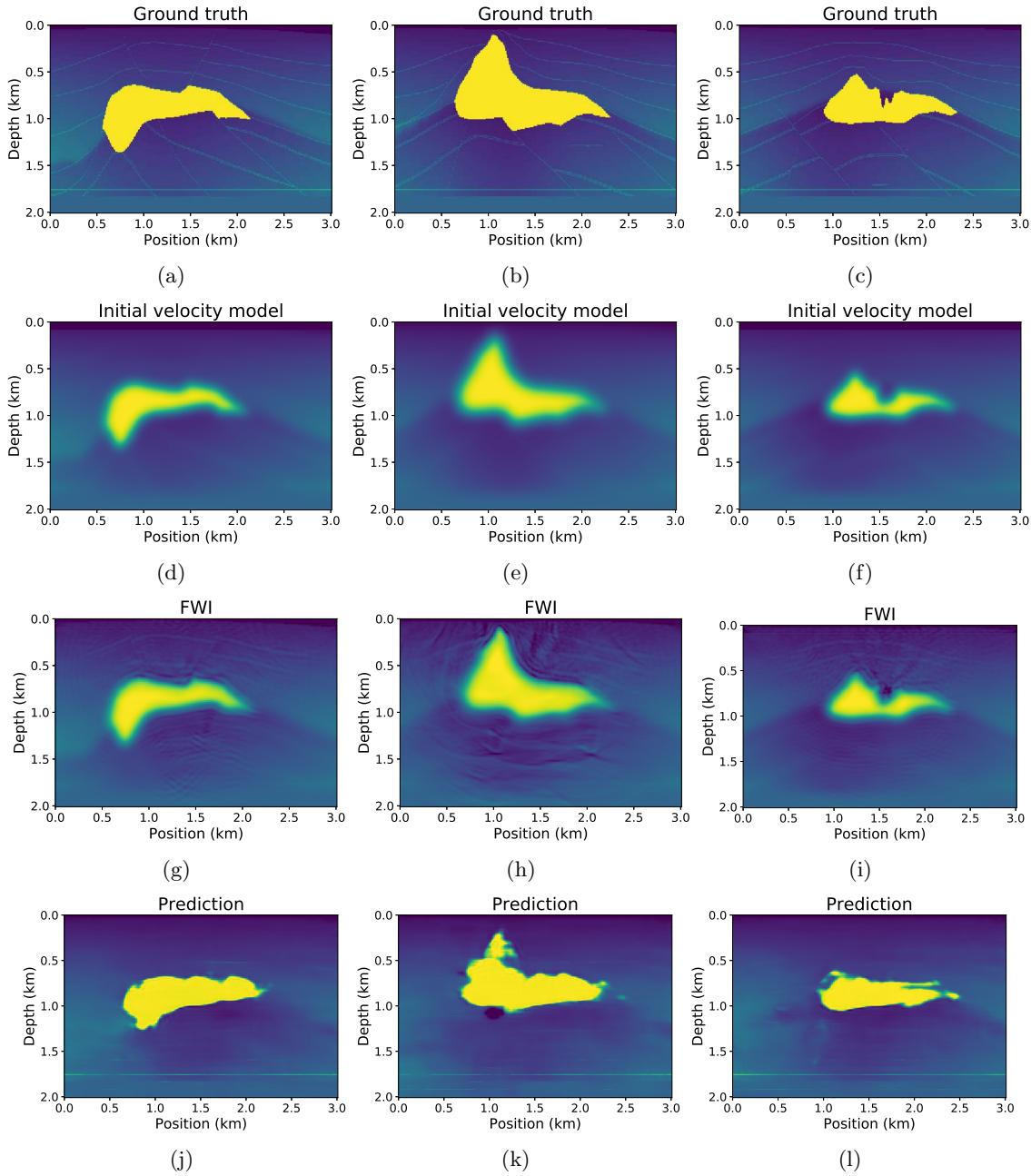


Figure 15: Comparisons of the velocity inversion (SEG salt models): (a)–(c) ground truth; (d)–(f) initial velocity model of FWI; (g)–(i) results of FWI; (j)–(l) prediction of our method.

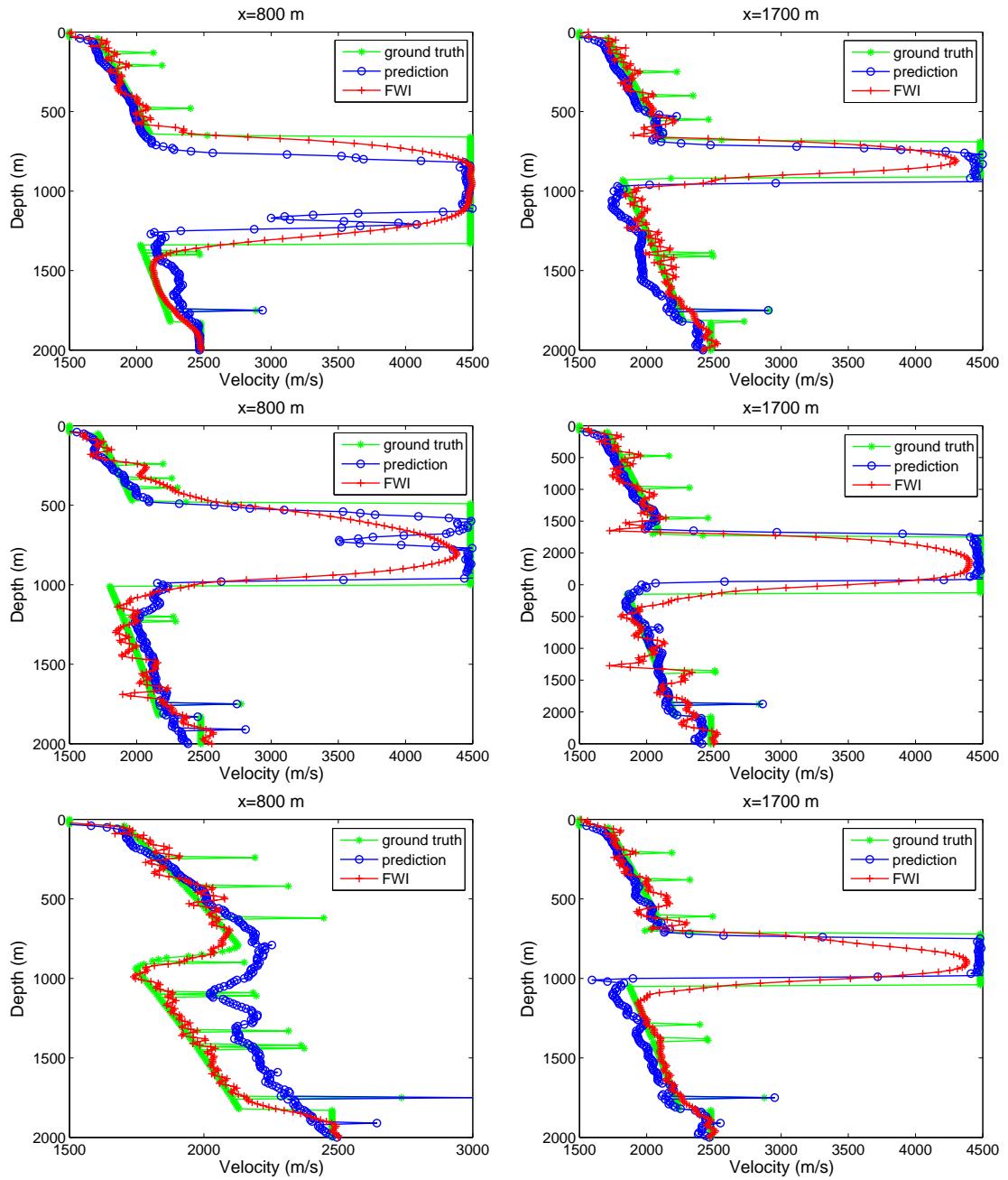


Figure 16: Vertical velocity profiles of our method and FWI. The prediction, FWI, and the ground-truth velocities in the velocity versus depth profiles at two horizontal positions ( $x = 800$  m,  $x = 1700$  m) of the three test samples in Figure 15 are shown in each row.

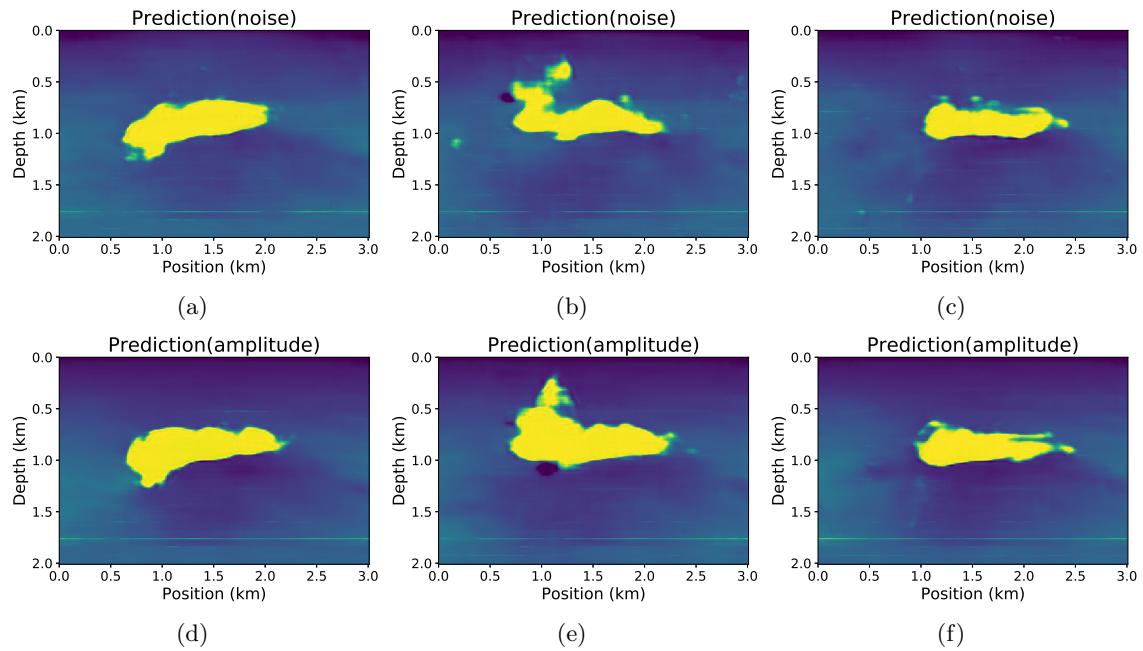


Figure 17: Sensitivity of the proposed method to noise and amplitude (SEG models): (a)–(c) prediction results with the noisy seismic data; (d)–(f) prediction results with magnified seismic amplitude. For the open dataset, our method yielded acceptable predictions when the input data were perturbed.

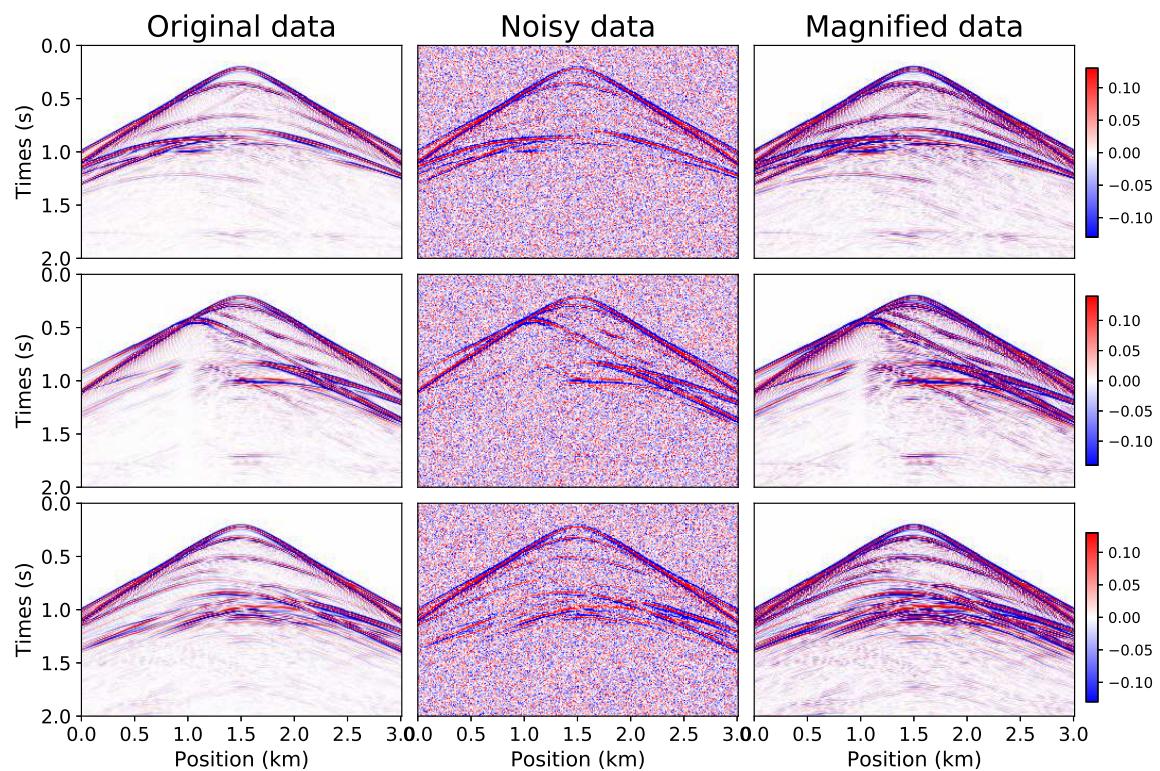


Figure 18: Comparison of records of SEG seismic data. Given in each row from left to right are original data, noisy data (with added Gaussian noise), and magnified data (to twice as larger). The corresponding velocity models of each row are the three models shown in Figure 15(a)–Figure 15(c), respectively.

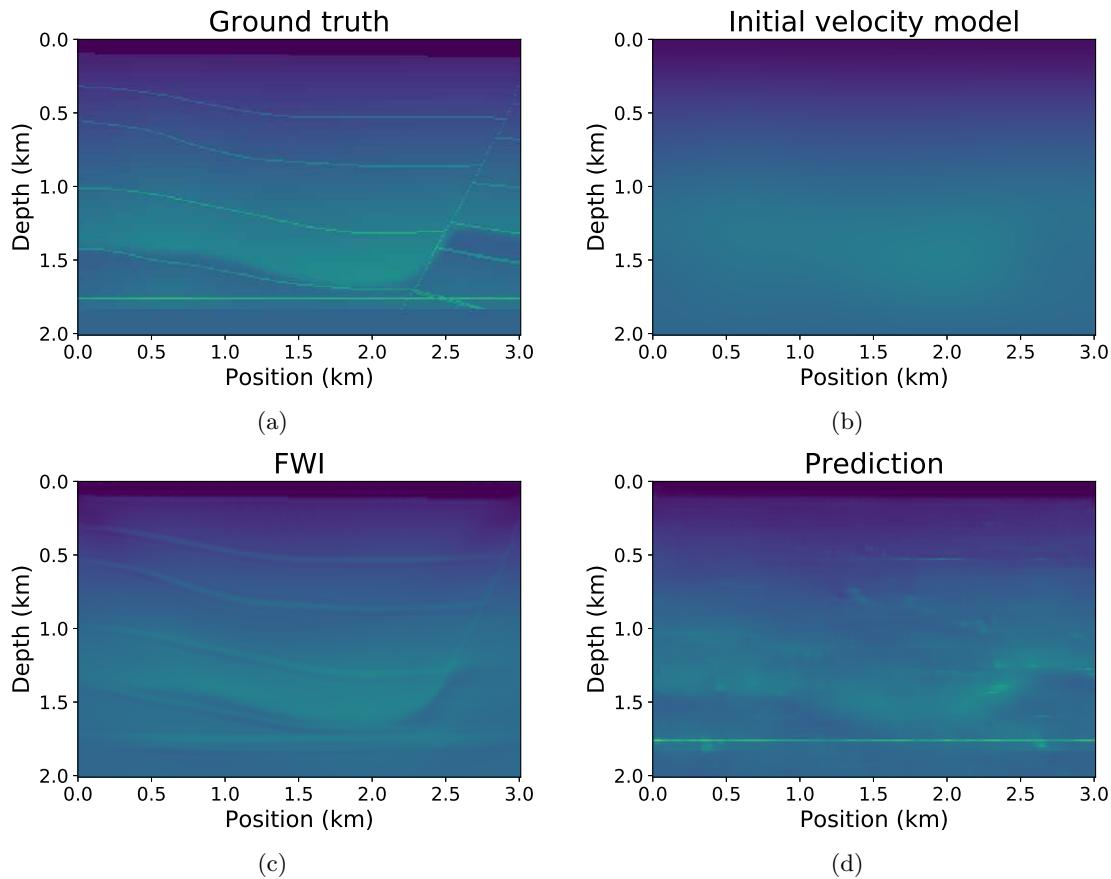


Figure 19: Inversion of velocity model without salt dome: (a) ground-truth velocity model; (b) initial velocity model of FWI; (c) result of FWI; (d) result of the proposed method. Our method showed a slightly lower performance than FWI because only 10 training models without the salt body were utilized. More training data are required to obtain corresponding improvement.

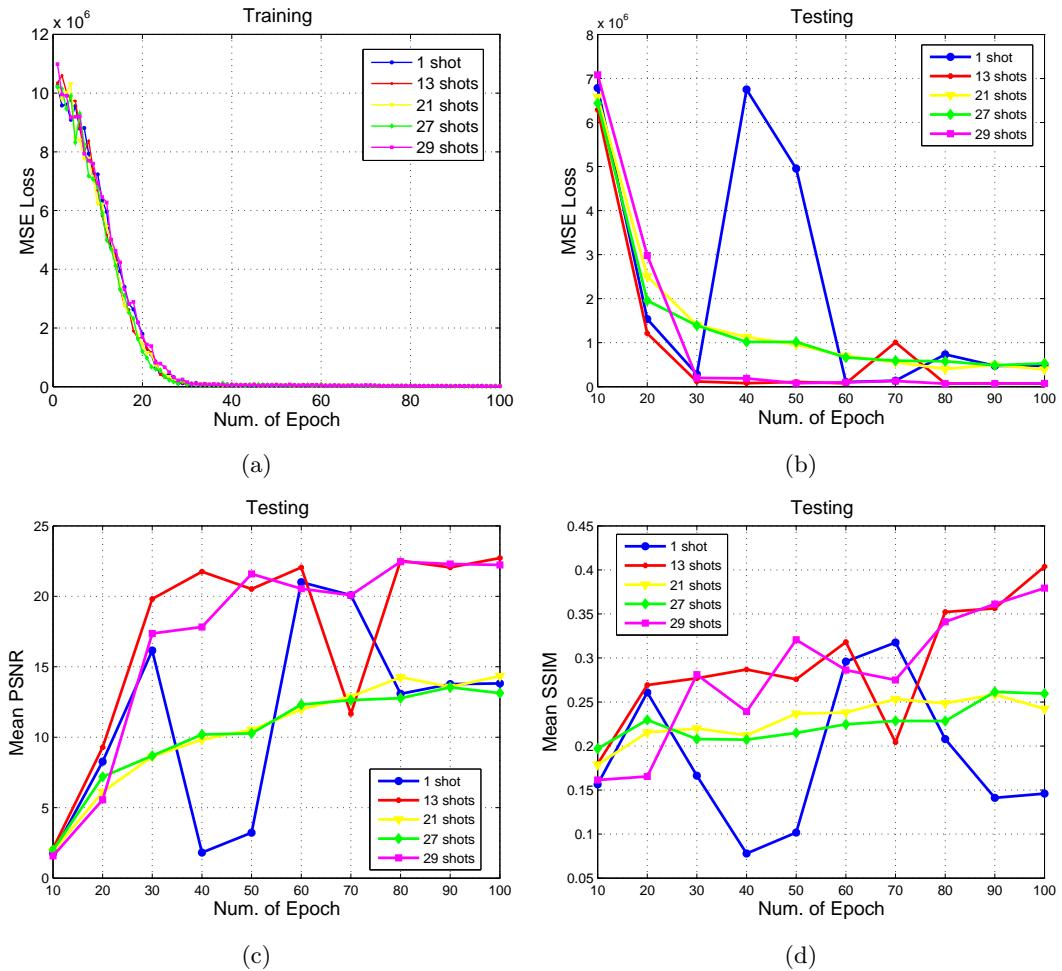


Figure 20: Comparison of performance versus number of epoch between different numbers of training shots: (a) mean-square error during training stage; (b) mean square error during testing stage; (c) mean peak signal-to-noise ratio (PSNR) during the testing stage; (d) mean structural similarity (SSIM) during the testing stage. All testing evaluation was obtained for 100 testing models.

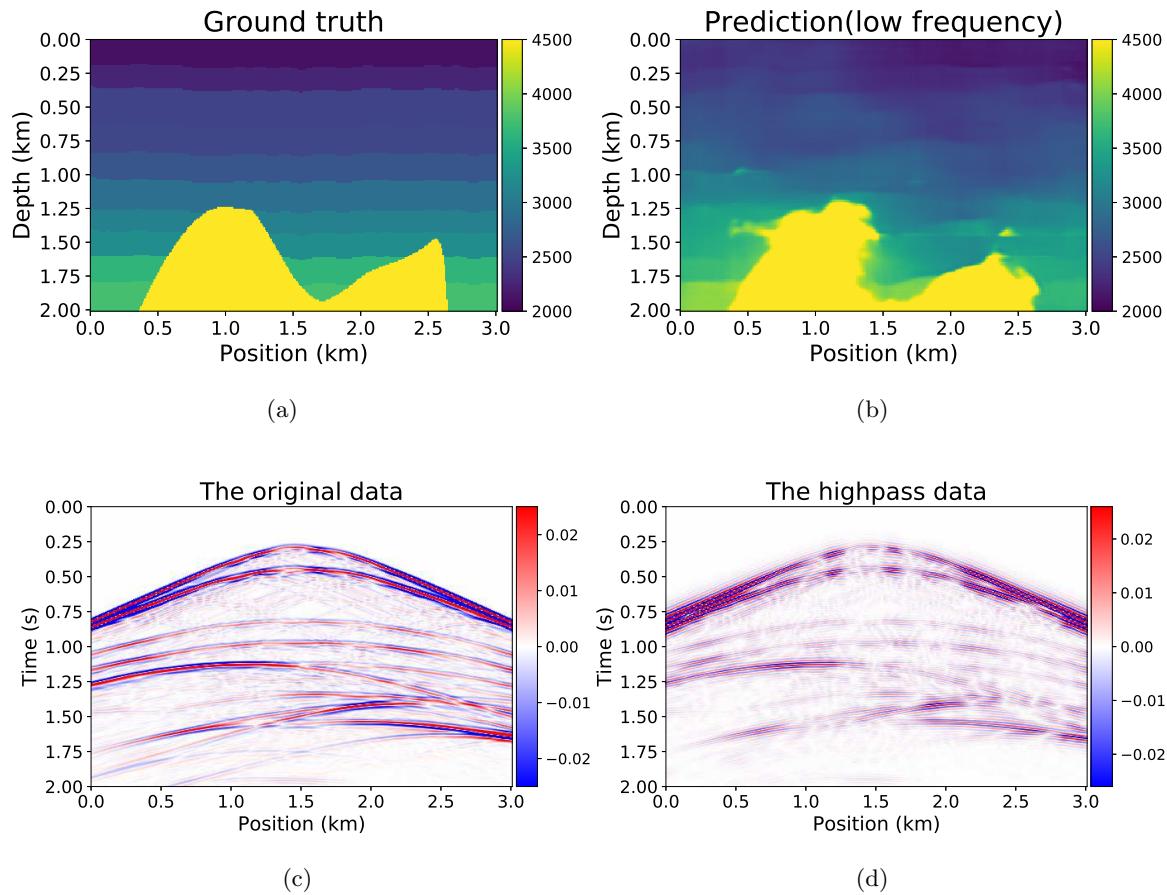


Figure 21: Typical results obtained by using our method when seismic data are lacking in low-frequency components: (a) ground truth; (b) prediction with data lacking low frequencies; (c) original seismic data (15th shot); (d) reconstructed data lacking 1/10 normalized Fourier spectrum.

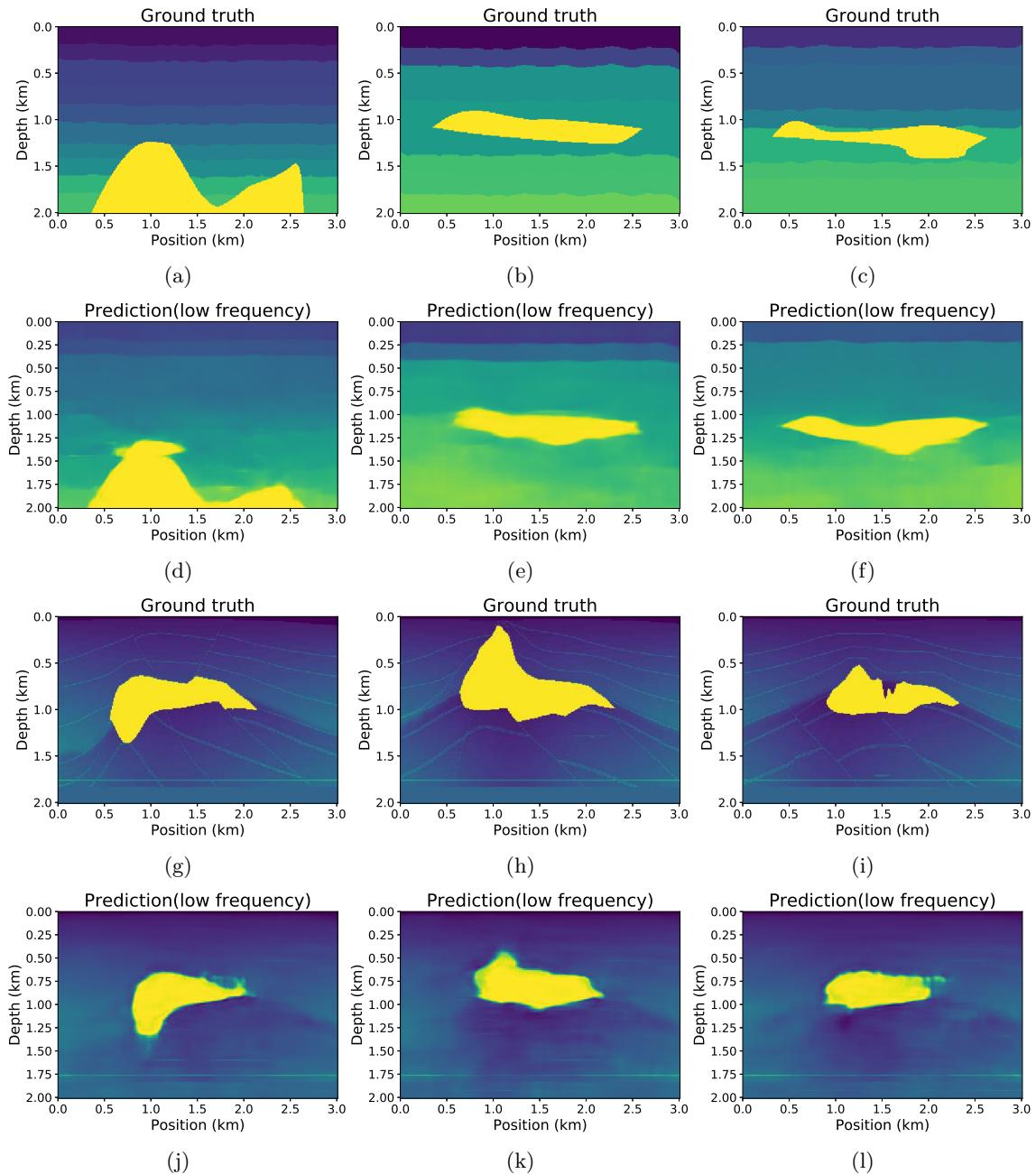


Figure 22: Results of the velocity inversion obtained when the training data lack low frequencies: (a)–(c) ground truth of simulated models; (d)–(f) prediction results; (g)–(i) ground truth of SEG salt models; (j)–(l): prediction results.

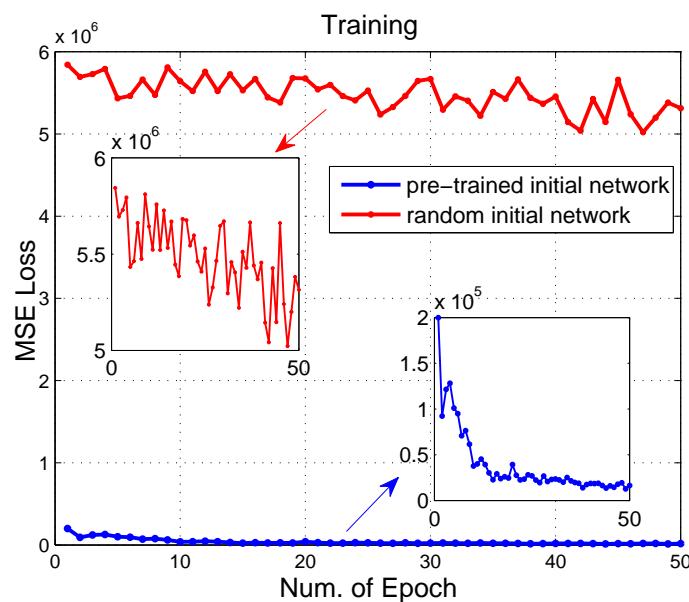


Figure 23: Comparison of training loss versus number of epochs. The red line denotes training with a random initial network. The blue line represents training with a pre-trained initial network (i.e., the trained network for the simulated dataset).

Acronyms	Corresponding definition
FCN	Fully convolutional neural network
DL	Deep learning
FWI	Full-waveform inversion
VMB	Velocity-model building
ML	Machine learning
DNN	Deep neural network
CNN	Convolutional neural network
SEG	Society of exploration geophysics
SGD	Stochastic gradient descent

Table 1: All acronyms used in this paper and their definitions.

Task	Source num	Spatial interval	Sampling time interval	Ricker wave	Maximum travel time
Velocity inversion	29	10 m	0.001 s	25 Hz	2 s

Table 2: Parameters of forward modeling.

Operation (Acronym)	Definition(2D)
Convolution (conv)	$output = K * input + b$
Batch normalization (BN)	$out = \frac{input - \text{mean}[input]}{\sqrt{\text{Var}[input] + \varepsilon}} * \gamma + \beta$
Rectified linear unit (Relu)	$out = \max(0, input)$
Max-pooling (max-pooling)	$out = \max[input]_{w \times h}$
Deconvolution / Transposed convolution (deconv)	$out = \overline{K} * input + b$
Skip connection and concatenation	$out = [input, padding]_{channel}$

Table 3: Definitions of the different operations for our proposed network.

Task	Learning rate	Epoch	Batch size	SGD algorithm	Number of training setd	Number of testing setd
Inversion (simulated model)	1.0e-03	100	10	Adam	1600	100
Inversion (SEG salt model)	1.0e-03	50	10	Adam	130	10

Table 4: Parameters of training process in our proposed network.

Time \ Method	FCN-based method		FWI	
Process				
Training	1078 min	43 min	N/A	N/A
Prediction	2 s	2 s	37 min	25 min

Table 5: Time consumed for the training and testing processes. The three columns of each method from left to right indicate the GPU time for the simulated velocity-model inversion, and SEG salt-model inversion. The training time is the total time required for all training sets; the testing time is for only one model. N/A indicates that FWI had no training time.