Deep Learning for Enhancing Multisource Reverse Time Migration

Yaxing Li, Xiaofeng Jia, Xinming Wu, *Associate Member, IEEE*, and Zhicheng Geng

***Abstract*— Reverse time migration (RTM) is a technique used to obtain high-resolution images of underground reflectors; however, this method is computationally intensive when dealing with large amounts of seismic data. Multisource RTM can significantly reduce the computational cost by processing multiple shots simultaneously. However, multisource-based methods frequently result in crosstalk artifacts in the migrated images, causing serious interference in the imaging signals. Plane-wave migration, as a mainstream multisource method, can yield migrated images with plane waves in different angles by implementing phase encoding of the source and receiver wavefields; however, this method frequently requires a trade-off between computational efficiency and imaging quality. We propose a method based on deep learning for removing crosstalk artifacts and enhancing the image quality of plane-wave migration images. We designed a convolutional neural network that accepts an input of seven plane-wave images at different angles and outputs a clear and enhanced image. We built over 500 1024 × 256 velocity models, and employed each of them using plane-wave migration to produce raw images at 0°, ±10°, ±20°, and ±30° as input of the network. Labels are high-resolution images computed from the corresponding reflectivity models by convolving with a Ricker wavelet. Random sub-images with a size of 512 × 128 were used for training the network. Numerical examples demonstrated the effectiveness of the trained network in crosstalk removal and imaging enhancement. The proposed method is superior to both the conventional RTM and plane-wave RTM (PWRTM) in imaging resolution. Moreover, the proposed method requires only seven migrations, significantly improving the computational efficiency. In the numerical examples, the processing time required by our method was approximately 1.6% and 10% of that required by RTM and PWRTM, respectively.**

***Index Terms*— Computational efficiency, deep learning, high-resolution imaging, seismic waves.**

Manuscript received 29 December 2021; revised 29 June 2022; accepted 4 August 2022. Date of publication 14 September 2022; date of current version 29 September 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 41774121 and Grant 42074125, in part by the Joint Open Fund of Mengcheng National Geophysical Observatory under Grant MENGO-202005, and in part by the China National Petroleum Corporation (CNPC) Innovation Foundation under Grant 2020D-5007-0303. *(Corresponding author: Xiaofeng Jia.)*

Yaxing Li is with the Institute of Advanced Technology, University of Science and Technology of China, Hefei 230026, China, and also with the Laboratory of Seismology and Physics of Earth's Interior, School of Earth and Space Sciences, University of Science and Technology of China, Hefei 230026, China (e-mail: yxli2017@mail.ustc.edu.cn).

Xiaofeng Jia and Xinming Wu are with the Laboratory of Seismology and Physics of Earth's Interior, School of Earth and Space Sciences, University of Science and Technology of China, Hefei 230026, China, and also with the National Geophysical Observatory at Mengcheng, University of Science and Technology of China, Hefei 230026, China (e-mail: xjia@ustc.edu.cn).

Zhicheng Geng is with the Bureau of Economic Geology, University of Texas at Austin, Austin, TX 78713 USA.

Digital Object Identifier 10.1109/TGRS.2022.3206283

I. Introduction

SEISMIC exploration is a geophysical method to infer the geometries and properties of underground rocks by analyzing the propagation pattern of seismic waves generated by artificial seismic sources at the surface. This method is regarded as one of the most effective methods to explore oil and gas resources. Complex underground structural imaging has become particularly valuable in recent years with increasing resource extraction; therefore, acquiring high-quality seismic data and developing accurate imaging algorithms has gained importance. Although wide- and full-azimuth acquisition techniques can yield adequate subsurface reflection signals, these massive data also entail a huge computational overhead [1], [2]. Seismic migration is an imaging technique that returns the reflections or diffractions from data to their true subsurface locations, constructing a high-resolution image of underground structures [3]. Several migration methods have been developed and applied in earlier studies [4], [5], for example, Kirchhoff migration [6], [7], [8] and Gaussian-beam migration [9], [10] based on ray theory, generalized screen propagators [11], [12] based on one-way wave equation, and reverse time migration (RTM) [13] based on two-way wave equation.

RTM, one of the most accurate migration methods, utilizes the full seismic wavefields and is capable of handling arbitrarily complex media with amplitude fidelity [5], [14]. Fig. 1 illustrates the workflow for executing RTM. We start by extracting the seismic records as common-shot gathers and preparing a migration velocity model. Subsequently, we execute the shot loop. For each shot, we need to forward-extrapolate the source wavefield and backward-extrapolate the receiver wavefield by loading the source and corresponding data, respectively. The zero-time lag cross correlation imaging condition applied on the two wavefields generates one single-shot image. After the shot loop is finished, all the single-shot images are stacked to obtain the final imaging result. However, in this process, the propagation of the wavefields is extremely time-consuming. Assuming N sources, a total of N times of migration is required, corresponding to $2N$ times of wavefields propagation, which includes N times each of both forward and backward wavefields propagation. The computational burden increases significantly with an increasing number of shots.

To reduce the computational cost of RTM, several multisource RTM methods have been proposed and developed [15], [16], [17], [18], [19], [20]. Unlike conventional RTM that deals with each shot separately, multisource migration methods

1558-0644 © 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: University of Science & Technology of China. Downloaded on January 26,2023 at 01:10:50 UTC from IEEE Xplore. Restrictions apply.

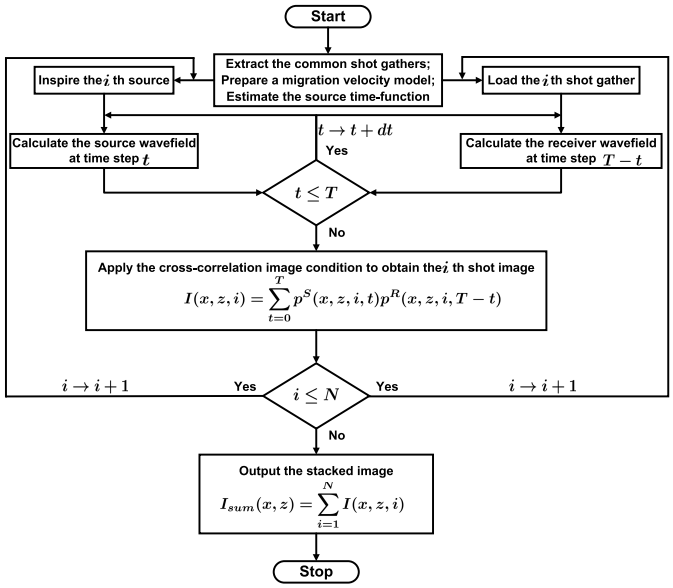


Fig. 1. RTM execution workflow. Here i denotes the shot number, t denotes the time step, and T is the total time steps. $p^S(x,z,i,t)$ and $p^R(x,z,i,t)$ represent the source and receiver wavefields of the ith source at time step t, respectively.

focus on forming a virtual source and the corresponding virtual shot gather by weighted summing multiple shots and their corresponding shot gathers simultaneously. The virtual shot and its corresponding shot gather are migrated to generate an image. Multisource RTM reduces the number of migrations significantly, thus improving computational efficiency. However, a serious drawback of these multisource methods is that mismatched sources and shot gathers produce crosstalk artifacts under nonlinear imaging conditions [18], severely hurting the imaging quality. The study of multisource RTM in suppressing crosstalk noise has two main tracks. One is to design different coding functions, and the other is to develop several denoising or smoothing methods as a post-stack denoising step or as a constraining operator in least-square reverse time migration (LSRTM).

For the first research track, such encoding methods are primarily classified into phase and amplitude encoding. Numerous scientists have investigated phase-coding methods [19], [21], [22], [23], wherein they concentrated on shifting the crosstalk positions in the migrated images to suppress them. To date, several types of phase-coding methods have been developed, such as plane-wave encoding, random phase encoding, modulation encoding, and harmonic encoding. Among them, the plane-wave RTM (PWRTM) is the most widely deployed multisource migration method. In PWRTM, different linear time delays imposed on the source and the receiver wavefields yield planar wavefields with different angles. The cross correlation imaging condition is placed on all the angular planar wavefields to generate the corresponding single-angle images, which are called common-angle image gathers (CAIGs). PWRTM images are created by stacking the CAIGs. The time delay and the number of angles affect the computational efficiency. A few research works based on amplitude encoding [16], [17]. In these methods, the imaging results of several different encoding functions are stacked to accomplish signal enhancement and noise suppression [16]. Jia *et al.* [24] presented a method for simultaneously encoding the amplitude and phase. This technique involves fitting the wavefield of a virtual super-shot in the time domain by encoding the amplitude and phase of the wavefields excited by multiple sources at the surface.

For the second research track in suppressing crosstalk noise, a series of methods have been proposed and developed, especially in terms of LSRTM. Xue *et al.* [25] introduced shaping regularization as a structure-oriented smoothing operator to suppress crosstalk noise in simultaneous-source data and incomplete data. For further improving the imaging quality of multisource LSRTM, Chen *et al.* [26] developed a local low-rank constrained LSRTM by applying a local singular spectrum analysis operator to denoising the crosstalk and preserving the steeply dip-angle events. Chen *et al.* [27] (2022) employed a low-rank decomposition operator for LSRTM of simultaneous-source data and incomplete data and achieved crosstalk noise suppression and edge preserving. Unlike the regularization-based approach described above, Zhang *et al.* [28] (2019) applied an excitation amplitude imaging condition for LSRTM to achieve favorable suppression of crosstalk noise. A dictionary learning approach was proposed to mute the crosstalk noise in LSRTM [29].

Although multisource RTM- and LSRTM-based methods have made great progress, they involve a trade-off between computational efficiency and suppression of noise. There is an urgent demand for a method that can guarantee a significant increase in computational efficiency while eliminating crosstalk noise.

Deep learning has advanced rapidly in the field of computer vision in recent years owing to the widespread use of convolutional neural networks (CNNs). CNN-based methods have progressed substantially in image classification [30], segmentation [31], object detection [32], and denoising [33] tasks. Compared with traditional methods, deep learning does not require manually defined criteria and automatically picks up the deep-level information of images. Deep learning has been successfully employed for data processing and interpretation in seismic exploration, for example, seismic data noise attenuation [34], [35], [36], [37], first-arrival picking [38], [39], [40], velocity model building [41], [42], [43], missing seismic data reconstruction [44], impedance inversion [45], [46], and seismic structural interpretation [47], [48], [49]. Although, a few advances in deep learning have been reported for simultaneous-source data deblending [50], [51], [52], few studies used deep learning for image denoising in multisource seismic migration.

We propose a CNN to eliminate the crosstalk in the migrated images and enhance the PWRTM image quality. The input of the network are CAIGs, that is, a combination of plane-wave images with seven incident angles (0°, ±10°, ±20°, and ±30°). The input has seven channels. The desired output is a clear and high-resolution image convolved from the corresponding reflectivity model, and we called it "a convolved image". The designed network architecture is a modified

Authorized licensed use limited to: University of Science & Technology of China. Downloaded on January 26,2023 at 01:10:50 UTC from IEEE Xplore. Restrictions apply.

U-net network [53] with three parts: an encoder, decoder, and refinement module. The encoder learns the features of input noisy image gathers at different levels. The decoder restores the high-level feature map to the same size as that of the input images. The refinement module predicts a clear image. To generate rich training data, we began by building about 500 2-D velocity models with different types of structures. With these velocity models, we placed the source and geophones at the surface to simulate the corresponding synthetic seismic records using the finite difference method. A slight smoothing was conducted on these velocity models, and the smoothed velocity models were employed for PWRTM to obtain CAIGs. Labels (high-resolution convolved images) were built by true velocity models. We defined a Laplacian pyramid-$\mathcal{L}1$ loss function for training the network. Using the training data and network architecture, we trained the CNN model with the defined loss function. Through numerical examples, we compared the stacked images of input CAIGs (i.e., the PWRTM images), predicted results, and ground truths. The comparisons revealed that the trained model realizes crosstalk removal and improves the imaging resolution. In addition, we tested the model with complex structures and compared the results for a stacked image of input CAIGs (i.e., simplified PWRTM image with only seven angles), traditional RTM image with 277 shots, and PWRTM image with 71 angles in terms of the image quality and processing time.

II. Theory

A. Plane-Wave RTM

The 2-D first-order velocity-stress equation in the acoustic medium is written as

$$\begin{cases} \dfrac{\partial v_x}{\partial t} = -\dfrac{1}{\rho}\dfrac{\partial p}{\partial x} \\ \dfrac{\partial v_z}{\partial t} = -\dfrac{1}{\rho}\dfrac{\partial p}{\partial z} \\ \dfrac{\partial p}{\partial t} = -\rho c^2\left(\dfrac{\partial v_x}{\partial x} + \dfrac{\partial v_z}{\partial z}\right). \end{cases} \tag{1}$$

Here v_x and v_z denote the x- and z-components of particle velocity, respectively; p is the pressure wavefield. ρ and c indicate the density and acoustic velocity of the media, respectively. In the conventional RTM, the source wavefield $p^S(x, z, i, t)$ of shot i is cross-correlated with its corresponding receiver wavefield $p^R(x, z, i, t)$, and we can obtain a single-shot image as follows:

$$I(x, z, i) = \sum_{t=0}^{T} p^S(x, z, i, t)\, p^R(x, z, i, T - t) \tag{2}$$

where T represents the recording time. Summing all the single-shot images, we obtain a stacked image $I_{\text{sum}}(x, z)$, represented by

$$I_{\text{sum}}(x, z) = \sum_{i=1}^{N} I(x, z, i) \tag{3}$$

where N is the number of shots. For this method, we need N forward extrapolations of the source wavefield and N backward extrapolations of the receiver wavefield, leading to a total of $2N$ simulation cycles. As the number of shots increases, the computational consumption of the conventional RTM is enormous.

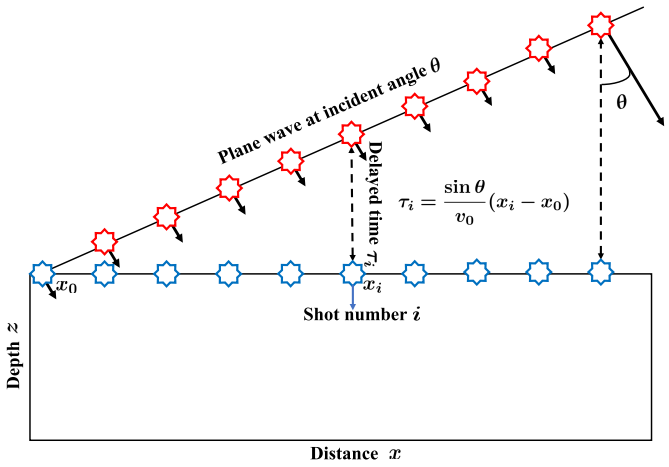


Fig. 2. Schematic of plane wave encoding. The blue octagons indicate the original shots or shot records, and the red octagons indicate the encoding shots or shot records.

For the PWRTM, linear time delay encoding functions are executed on the shot records to construct plane waves, and the encoding functions are expressed as [15]

$$d_{\text{sum}}(x_g, t, \theta) = \sum_{i=1}^{N} d_i(x_g, t) * \delta(t - \tau_i(\theta)) \tag{4}$$

where $d_i(x_g, t)$ and τ_i denote ith shot record and its corresponding time-shift function, respectively. The time-shifted shot records are stacked to obtain a plane wave record $d_{\text{sum}}(x_g, t, \theta)$ with the surface incidence angle of θ. The surface incidence angle θ is defined as the angle between the propagation direction of the incident plane wave and the depth direction (see Fig. 2). The shots and shot records are encoded in the same way, and the encoding source-time function for the ith shot is expressed as

$$W_i(t, \theta) = f(t) * \delta(t - \tau_i(\theta)) \tag{5}$$

where $W_i(t, \theta)$ denotes the encoding source-time function of the ith shot, and $f(t)$ denotes the original source-time function. Fig. 2 shows the schematic of plane wave encoding. The blue and the red octagons indicate the original and the encoding shots, respectively. Assuming that the incidence angle of the plane wave is θ, the time shift of the ith shot is expressed as

$$\tau_i(\theta) = \frac{\sin\theta}{v_0}(x_i - x_0) \tag{6}$$

where v_0 represents the seismic velocity at surface and x_i represents the position of ith shot. As the incident angle θ is large, the time shift τ_i increases.

In fact, the encoding functions applied to the shots and shot records are equivalent to the encoding functions acting on the source and receiver wavefields. The source and receiver wavefields of each shot are first encoded by time delay, and the encoding source and receiver wavefields of all shots

Authorized licensed use limited to: University of Science & Technology of China. Downloaded on January 26,2023 at 01:10:50 UTC from IEEE Xplore. Restrictions apply.

are, respectively, stacked to obtain the corresponding stacked wavefields. The stacked source p^S_{sum} and receiver wavefields p^R_{sum} at the incident angle of θ, respectively, assume the form

$$\begin{cases} p^S_{\text{sum}}(x,z,t,\theta) = \sum_{i=1}^{N} p^S(x,z,i,t-\tau_i(\theta)) \\ p^R_{\text{sum}}(x,z,t,\theta) = \sum_{i=1}^{N} p^R(x,z,i,t-\tau_i(\theta)). \end{cases} \tag{7}$$

Various time delay patterns form different incident-angle wavefields. Using the source wavefield as an example, Fig. 3 shows a schematic for constructing a plane wave. Fig. 3(a) illustrates a plane wave with 0° incidence angle, in which case the plane wave does not need to perform any time delay. The red octagon here indicates the encoded sources. When a linear time delay is applied to the sources at the surface, a plane wave with an incidence angle of θ is formed as shown in Fig. 3(b).

The cross correlation imaging condition is applied to the stacked wavefields and one plane-wave image I^{plane} at an angle of θ can be obtained, expressed by

$$I^{\text{plane}}(x,z,\theta) = \sum_{t=0}^{T} p^S_{\text{sum}}(x,z,t,\theta) p^R_{\text{sum}}(x,z,T-t,\theta). \tag{8}$$

Owing to the complexity of underground structures, obtaining a desirable migrated image from only one angle is difficult, and accomplishing multiple angles migration is necessary. Summing the multiangle plane-wave images yield a final stacked image, represented by

$$I^{\text{plane}}_{\text{sum}}(x,z) = \sum_{\theta} I^{\text{plane}}(x,z,\theta). \tag{9}$$

Here we set the number of incident angles θ to M. By comparing the conventional (3) and PWRTM (9) methods, we find that the migration number of conventional RTM is the same as the number of shots N, while the migration number of plane-wave migration is equal to the number of plane-wave angles M. For PWRTM, the choice of plane-wave angles is critical, as it significantly affects imaging accuracy and computational efficiency.

B. CNN Architectures

As mentioned earlier, plane-wave migration requires a suitable number of angles, i.e., the same number of migrations, to achieve a balance between computational consumption and imaging quality. It is a challenge to obtain high-accuracy imaging results with a small number of migrations. In addition, (9) is a linear superposition of plane-wave migration images of multiple angles, and such a linear superposition provides limited scope for suppressing crosstalk and improving imaging quality. To resolve this limitation, we designed an end-to-end CNN architecture (Fig. 4) to denoise and enhance the plane-wave migrated images. The proposed architecture accepts CAIGs, that is, a combination of plane-wave images with seven incident angles (0°, ±10°, ±20°, ±30°) as input. The 3-D CAIG is input by channels, so the input feature maps are 2-D and the number of input channels is 7. The desired output of the network is a noise-free high-resolution image

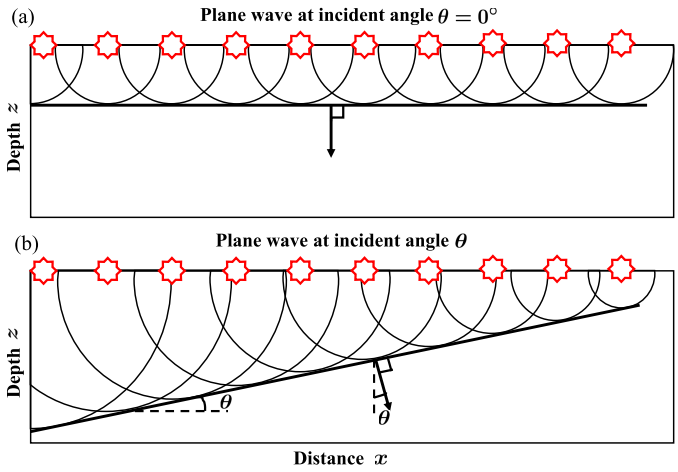


Fig. 3. Schematic of plane wave synthesis. (a) Plane wave at incident angle $\theta = 0°$. (b) Plane wave at incident angle θ.

convolved from the corresponding reflectivity model corresponding to the Ricker wavelet with a dominant frequency of approximately 70 Hz. The network has three potential functions: optimal stacking of the images with multiple incidence angles, crosstalk noise removal, and imaging enhancement. The proposed architecture is designed by replacing the convolutional blocks in a U-net network [53], which comprises three parts: an encoder, decoder, and refinement module [54], [55]. Regarding the encoder and decoder structure, the U-net can help extract multiscale structural features of seismic images.

Several layers of ResNet-18 [56] are used to constitute the encoder part. ResNet-18 comprises 18 layers of residual net and can address the degradation problem of neural networks with depth. A series of residual blocks are stacked to form the Resnet-18. Fig. 5 illustrates a residual block, which comprises two 3×3 convolution filters. The shortcut of the residual block enables the network to learn the seismic structural features easily [56]. The details of the layers or blocks used in the CNN architecture are shown in Table I. When we input raw plane-wave CAIGs, after two 5×5 convolution layers followed by batch normalization and rectified linear unit (ReLU) activation function, the feature map size becomes 1/2 of the input image. After max pooling and four blocks (Block1–Block4) of Resnet-18, the feature map size successively becomes 1/4, 1/8, 1/16, and 1/32 of the input. Here, each block among Block1–Block4 contains two residual blocks (Fig. 5).

The decoder part is designed to recover the desired size of the feature map from the encoder part. We used the upsampling block with reference to [55], [57]. As shown in Fig. 6, the upsampling block, which is somewhat similar to the residual block (Fig. 5), has a shortcut. This upsampling structure allows for efficient access to global information. For the upsampling block, the input feature map is first bilinearly interpolated to enlarge the size. The expanded-size feature map is then subjected to double convolutional layers and one convolutional layer, respectively. Finally, the two parts are superimposed with a shortcut. Using five upsampling blocks (UpBlock0-UpBlock4), the feature map size becomes 1/16, 1/8, 1/4, 1/2, and 1 the size of the input images, successively. We concatenated different scale layers between the encoder and decoder parts (dark blue arrows in Fig. 4).

Authorized licensed use limited to: University of Science & Technology of China. Downloaded on January 26,2023 at 01:10:50 UTC from IEEE Xplore. Restrictions apply.

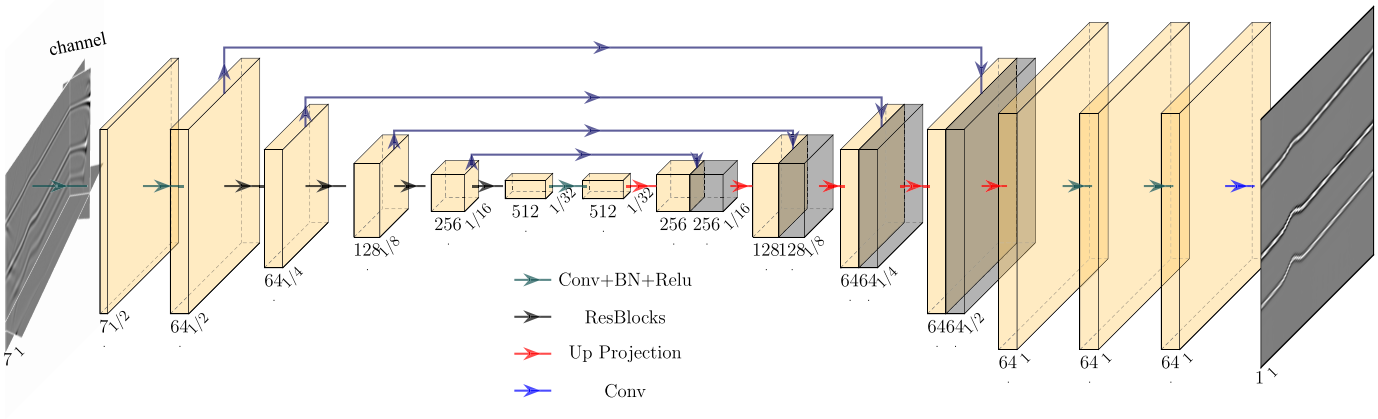


Fig. 4. Proposed CNN architecture. The input is noisy CAIGs of PWRTM, which is a combination of plane-wave images with seven incident angles (0°, ±10°, ±20°, and ±30°). The 3-D CAIG is fed into a CNN in a channel-wise fashion; the feature map is 2-D and the number of input channels is 7. The output is a clear and enhanced image. The CNN architecture comprises encoder, decoder, and refinement parts. The yellow rectangles are the layers inside the CNN. The size of these rectangles denotes the size of each layer's feature maps and the numbers below them denote the number of channels for each layer. The arrows indicate the operations between different layers.

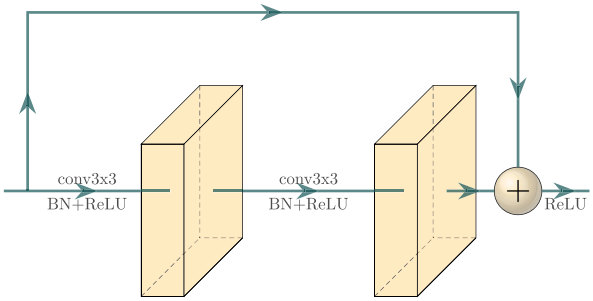


Fig. 5. Structure of residual block.

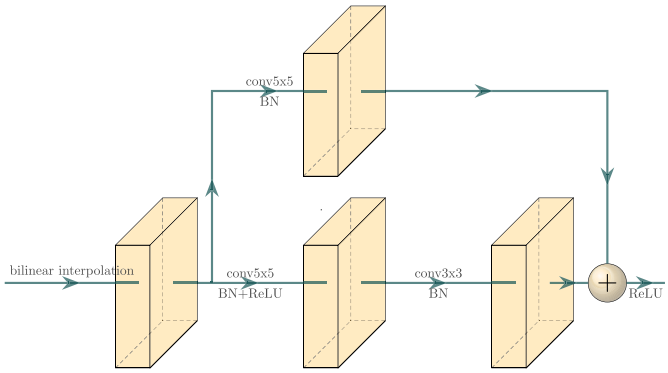


Fig. 6. Structure of upsampling block.

Regarding the refinement part, we stacked three convolutional layers (Conv4-Conv6). Conv4 and Conv5 operate with a 5 × 5 convolution kernel, batch normalization, and ReLU. The last convolutional layer (Conv6) is a straightforward 5 × 5 convolution kernel to obtain the final output of the network.

C. Loss Functions

We defined a loss function based on the Laplacian pyramid-$\mathcal{L}1$ loss to train and validate the proposed network. The loss function follows the form presented in [58]:

$$\mathcal{L}_{\text{lap}} = \sum_{j}^{M} 2^{2j} \left| \left(L^j\left(y_k^p\right) - L^j\left(y_k\right) \right| \right. \tag{10}$$

where $L^j(y_k^p)$ and $L^j(y_k)$ represent the jth level Laplacian pyramid features of the predicted and label images, respectively. We first impose average pooling with different sizes on the predicted and the label images to obtain different levels Laplacian pyramid features. Then $\mathcal{L}1$ losses of predicted and label image features at different scales are calculated separately, and these $\mathcal{L}1$ losses are weighted and summed to obtain the final loss. In our training process, M equaled 5,

TABLE I

DETAILS OF THE LAYERS OR BLOCKS USED IN EACH SECTION OF THE CNN ARCHITECTURE, INCLUDING KERNEL SIZE, SCALE, AND THE NUMBER OF CHANNELS OF THE INPUT AND OUTPUT FEATURE MAPS

Section	Layer/Block	Kernel Size	Scale	Input Channel	Output Channel
Encoder	Conv1	5	1/2	7	7
	Conv2	5	1/2	7	64
	Pool1	3	1/4	64	64
	Block1		1/4	64	64
	Block2		1/8	64	128
	Block3		1/16	128	256
	Block4		1/32	256	512
Decoder	Conv3	1	1/32	512	512
	UpBlock0		1/16	512	256
	UpBlock1		1/8	512	128
	UpBlock2		1/4	256	64
	UpBlock3		1/2	128	64
	UpBlock4		1	128	64
Refinement	Conv4	5	1	64	64
	Conv5	5	1	64	64
	Conv6	5	1	64	64

Authorized licensed use limited to: University of Science & Technology of China. Downloaded on January 26,2023 at 01:10:50 UTC from IEEE Xplore. Restrictions apply.

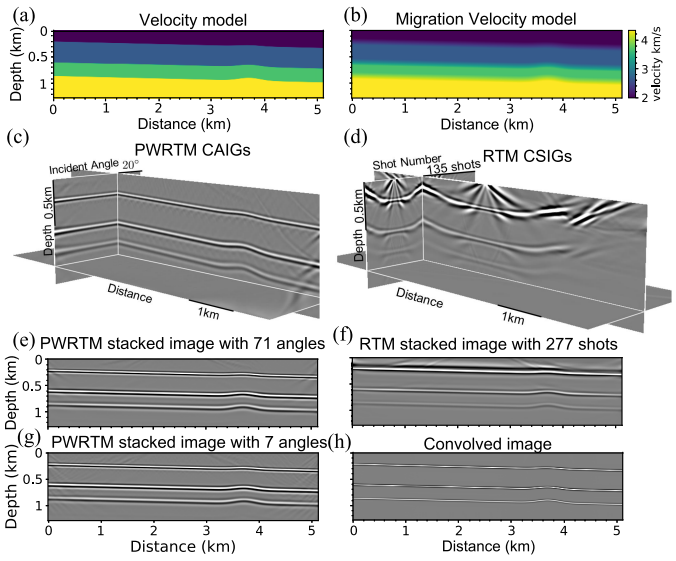


Fig. 7. Four-layer velocity model and its imaging results. (a) True velocity model. (b) Migration velocity model. (c) CAIGs obtained with PWRTM with 71 incident angles (from $-35° \sim 35°$ with an interval of 1°). (d) CSIGs obtained using typical RTM with 277 shots. (e) PWRTM stacked image of 71 incident angles, produced by stacking the image gathers as shown in Fig. 7(c) along the angular direction. (f) Conventional RTM stacked image with 277 shots. (g) PWRTM stacked image of seven incident angles ($-30° \sim 30°$ with an interval of 10°). (h) High-resolution image convolved from the corresponding reflectivity model using Ricker wavelet with a dominant frequency of approximately 70 Hz.

which means that $\mathcal{L}1$ loss function was conducted on images at five different scales. The Laplacian pyramid loss is of potential importance in multiscale image denoising and enhancement.

III. Experiments

A. Comparison Experiment of RTM and PWRTM

To visually compare the regular and PWRTM, we designed a simple velocity model with a size of 1024×256 and a grid interval of 5 m [Fig. 7(a)]. The sources and geophones were arranged on the surface and the seismic records were synthesized by seismic modeling using the finite difference method. We employed the first-order derivative of a Gaussian wavelet with a dominant frequency of 25 Hz as the source-time function. The number of shot gathers was set as 277 and shot interval as 20 m. The recording time was set as 6 s and time interval as 0.4 ms. The true velocity shown in Fig. 7(a) was smoothed to obtain a migration velocity [Fig. 7(b)]. With the migrated velocity model and synthetic seismic records, we performed regular RTM and PWRTM to obtain the imaging results. Fig. 7(c) illustrates the CAIGs obtained using the PWRTM (5), which has 71 angles from $-35°$ to $35°$ with an interval of 1°. When performing the PWRTM, the maximum-delayed time was set as 2.8 s; therefore, the total simulation time was 8.8 s, which is approximately 1.5 times the original record length. The time delay implies an increment in computational effort. We stacked all 71 plane-wave migrated images to obtain the image shown in Fig. 7(e). Although this image reflects the underground structure satisfactorily and the crosstalk artifacts are well suppressed, it requires 71 migrations and is computationally inefficient. To reduce the computational burden, we selected the images corresponding to $-30° \sim 30°$ with an interval of 10° for the CAIGs [Fig. 7(c)], and then summed them to obtain the PWRTM stacked image with seven incident angles, as shown in Fig. 7(g). It is evident that although the image in Fig. 7(g) can generally reveal the underground structures, it is heavily affected by crosstalk artifacts.

Fig. 7(d) shows the common-shot image gathers (CSIGs) using conventional RTM (2). For the generation of image gathers shown in Fig. 7(d), 277 shots correspond to 277 cycles of migration, which requires a large amount of computation time. The total images of 277 shots are stacked to obtain the image shown in Fig. 7(f). The imaging result is satisfactory, except for the presence of low-frequency artifacts in the shallow part. The required computation time for obtaining the results shown in Fig. 7(e)–(g) is 4596, 29 027, and 468 s, respectively. Comparing the three images in Fig. 7(e)–(g), that in Fig. 7(g) requires the least computational consumption; however, it suffers the greatest interference from artifacts. Fig. 7(h) presents a high-resolution image derived by convolving the reflectivity of this model with a Ricker wavelet with an approximately 70 Hz dominant frequency. The reflectivity can be acquired from the true velocity model [Fig. 7(a)]. The high-resolution convolved image is our expected ideal image of the subsurface structure. Based on the above findings, this study aimed to obtain high-quality imaging results with the minimum possible computation time, and deep learning provides a useful tool to help achieve this goal.

B. Training Datasets

A large amount of input data and corresponding labels are essential during the training and validation of a CNN model. To implement image denoising and enhancement of plane-wave migration by training a CNN, we need diverse plane-wave migration image gathers and corresponding targets, that is, high-resolution images convolved from the corresponding reflectivity models. By extracting multiple sub-images from a migrated image and the corresponding target image, we can obtain a large amount of training data. We designed a workflow to build velocity models and their corresponding migrated image gathers, as shown in Fig. 8. In this workflow, we first created a vast number of 2-D true velocity models of size 1024×256. Using the velocity models, we placed sources and geophones at the surface and synthesized seismic records through finite-difference simulation. Velocity models were also employed to derive the reflectivities, which were then convolved with a Ricker wavelet with a dominant frequency of approximately 70 Hz to produce clear and high-resolution convolved images. The true velocities were smoothed and were used for the plane-wave migration to obtain the imaging gathers. At each iteration during training, the image gathers were fed into the network, producing predicted images. We calculated the residuals between the predicted images and the corresponding clear images (i.e., the high-resolution convolved images), and we back-propagated the residuals to the network. In training, the network adjusts the parameters until they meet our requirements.

Authorized licensed use limited to: University of Science & Technology of China. Downloaded on January 26,2023 at 01:10:50 UTC from IEEE Xplore. Restrictions apply.

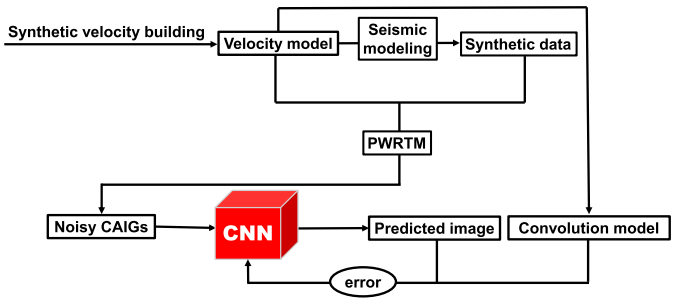


Fig. 8. Workflow for building training datasets.

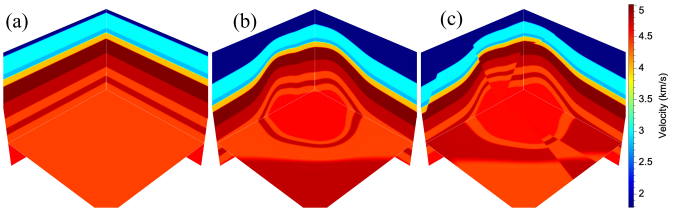


Fig. 9. Construction of 3-D velocity models. (a) Flat model. (b) Folded model. (c) Fault model.

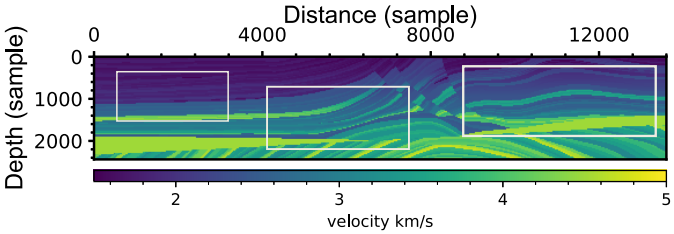


Fig. 10. Marmousi velocity model. To create the velocity models, we randomly extracted different regions (white rectangle boxes) of the model and applied affine transformations on the extractions to obtain diverse structural velocities.

1) Synthetic Velocity Models and Seismic Records: As shown in the workflow in Fig. 8, the formation of image gathers and labels requires velocity models and seismic records. The structural diversity of velocity models has a strong impact on the quality of output images. We used two schemes to automatically build different types of velocities. For the first scheme, we followed the work of Wu *et al.* [59] to create the velocity models. We started with 3-D flat velocity models [Fig. 9(a)], and then added the folding [Fig. 9(b)], and faulting [Fig. 9(c)]. Thus, we constructed 100 3-D velocity models embodying different structures. We extracted one vertical slice from these 3-D velocity models along the inline direction and obtained 100 2-D-velocity models.

In the second scheme, we performed random extractions from the Marmousi velocity model [60] (Fig. 10) and conducted affine transformations on these extractions using the augmentor tool [61]. The affine transformations include elastic distortions, perspective transforms, zooms, and their combinations. 405 velocity models were created. Notably, although the sizes of the randomly chosen regions were not fixed, they are all resampled to the same size of 1024(distance) $\times$ 256(depth).

Fig. 11 shows six typical velocity models, and we observed that these velocity models contain either simple or complex structures, such as horizontal layers, folds, and faults. The richness of models can help the proposed CNN perform well

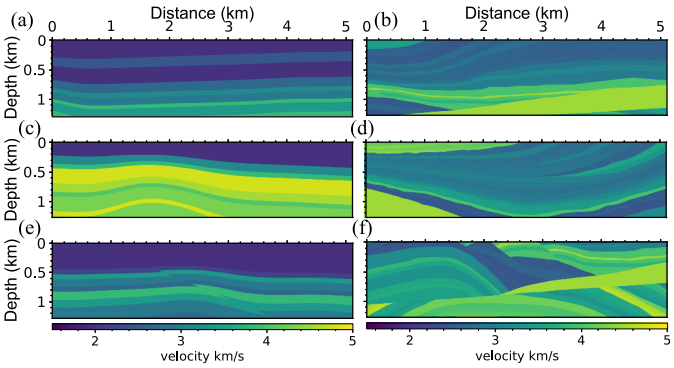


Fig. 11. Velocity models with different structures.

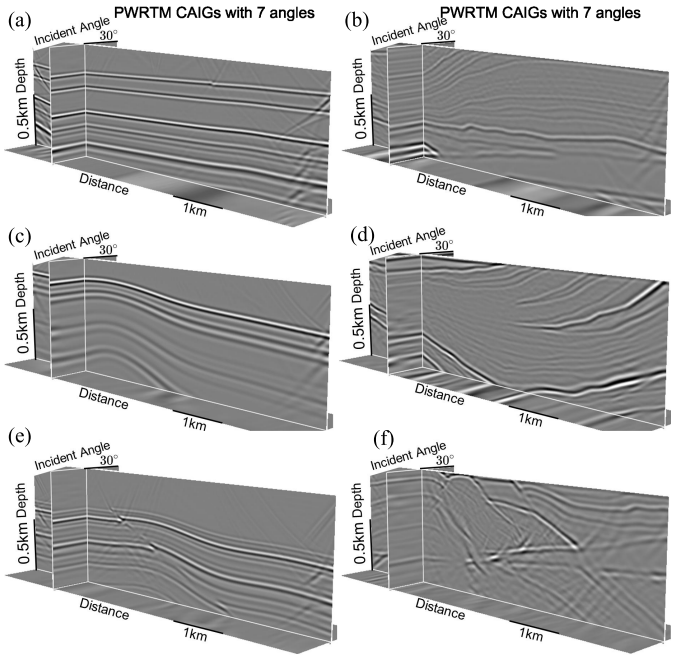


Fig. 12. Plane-wave image gathers corresponding to the velocity models shown in Fig. 11, respectively. These imaging gathers have seven incidence angles (0°, $\pm$10°, $\pm$20°, $\pm$30°) and are heavily disturbed by crosstalks.

at learning the image features. Using the velocity models, we simulated the corresponding seismic records. The simulations were performed with four dominant frequencies of the source wavelets: 15, 20, 25, and 30 Hz. The sources and geophones were distributed on the entire surface. The number of shots was 277.

2) Synthetic Migrated Images and High-Resolution Images Convolved From the Corresponding Reflectivity Models: Using the velocity models and corresponding synthetic seismic records, we employ PWRTM to obtain migrated image gathers. The smoothed velocities were used for the PWRTM (5), forming CAIGs (Fig. 12). We can observe that these images are interrupted by crosstalk artifacts, which even obscure the true signals.

The true velocities were employed to obtain high-resolution images convolved from the corresponding reflectivity models (Fig. 13). Compared to the noisy image gathers (Fig. 12), the "true" images in Fig. 13 are clear and present the underground structure accurately. In terms of computational efficiency, the noisy image gathers shown in Fig. 12 require only seven cycles

Authorized licensed use limited to: University of Science & Technology of China. Downloaded on January 26,2023 at 01:10:50 UTC from IEEE Xplore. Restrictions apply.

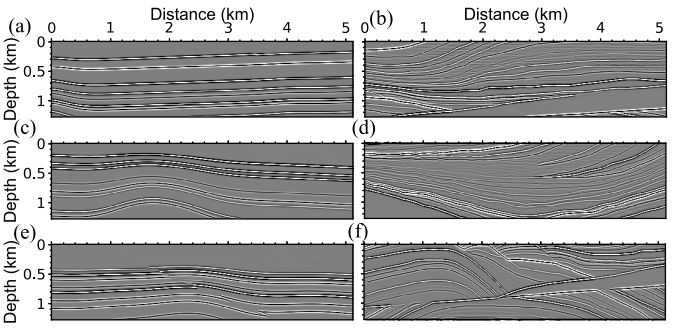


Fig. 13. High-resolution images convolved from the reflectivity models corresponding to the velocities depicted in Fig. 11, respectively. These images have high quality and are free from crosstalk artifacts, as desired for subsurface images.

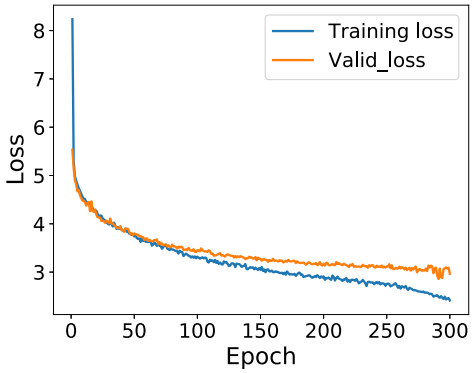


Fig. 14. Loss of designed network for training and validation datasets.

of migration because of the implementation of simplified PWRTM with seven angles to generate the raw images.

C. Training and Validation

Utilizing the above-discussed method, we synthesized over 500 datasets. We utilized approximately 80% of the datasets for training and the remaining for validation. To expand the training data, four randomly extracted data of size 512 × 128 were trained for each epoch of the dataset during training and validation of the network. We normalized the input image gathers along channel directions to ensure their amplitude consistency.

We used the Pytorch version 1.0 [62] for the denoising and enhancement of seismic images. The network was trained using NVIDIA Tesla V100 graphics processing unit (GPU) with 32 GB memory. The batch size was set as 32, with an initial learning rate of 10^{-4}.

We performed training with the Adam optimizer [63] on these hyperparameters and normalized seismic image gathers to update the model parameters. We terminated the training process after 300 epochs of training and validation. Fig. 14 shows the loss of the trained model, and it is visible that both the training and valid losses converge to small values.

Fig. 15 the feature maps of predicted images and labels at five different levels, and these images reflect the features of different resolutions. Performing loss functions on the images at five levels can help the network to extract effective features well.

IV. Results

To test the trained model, we designed velocity models with different structures, as shown in Fig. 16. These velocity models contained horizontal layers, folds, faults, and their combinations. The sampling size of the velocity models was 1024 × 256. With these velocity models, we employed the previously mentioned approach of building training datasets to obtain the corresponding plane-wave image gathers and high-resolution images. For plane-wave migration, different dominant frequencies of seismic wavelets were used to produce images with varying resolutions. Allowing for responses to different frequencies, the models shown in Fig. 16 were setup with different dominant frequencies of wavelets, including 20, 25, 25, 25, 30, and 30 Hz.

To demonstrate the effectiveness of our method, we compared three types of images: stacked PWRTM image (i.e., stack of the input images), predicted image, and ground truth. The imaging results corresponding to the velocity models shown in Fig. 16 are illustrated in Fig. 17. The plane-wave stacked images are displayed in the first column in Fig. 17, and severe interference from crosstalk noise can be observed, with low resolution and poor imaging quality. Similar to the ground truths (third column in Fig. 17), the predicted images (second column in Fig. 17) are free from interference from crosstalk artifacts and the signals are enhanced. In addition, the predicted images have high resolution and imaging quality. Owing to the frequency-band limitation of the wavelets and the effect of crosstalk noises, thin layers are difficult to recognize in the plane-wave stacked images, whereas the same are clearly visible in the high-resolution predicted images (marked by red arrows in Fig. 17). Hence, the designed network achieves optimal superposition of imaging gathers, suppression of crosstalk artifacts, and imaging enhancement.

Next, used the trained network to test a model with greater complexity, as illustrated in Fig. 18. Fig. 18(a) shows the velocity model with a size of 1024 × 256, with dense layers and steeply dipping folds and multiple faults. This complex model tends to present difficulties in seismic imaging. The parameters are similar to the model shown in Fig. 7 except that the delay time is 3.6 s. To demonstrate the merits of the trained model, we compared the predicted image [Fig. 18(f)] with the high-resolution image [Fig. 18(b)] convolved from the corresponding reflectivity model, PWRTM stacked image [Fig. 18(c)], traditional RTM stacked image [Fig. 18(d)], and stacked image of input image gathers [Fig. 18(e)]. The PWRTM stacked image shown in Fig. 18c was produced by summing plane-wave images at 71 angles (from $-35° \sim 35°$ with an interval of 2°), which requires 71 migrations. The RTM stacked image exhibited in Fig. 18(d) was generated by summing single-shot images of 277 shots, which requires 277 migrations. Both these types of stacked images [Fig. 18(c) and (d)] allow for clear imaging of subsurface structures without crosstalk artifacts. However, they are limited in resolution and insensitive to fine layers owing to the influence of wavelet frequencies [red arrows in Fig. 18(c) and (d)]. The plane-wave CAIGs of seven angles (from $-30° \sim 30°$ with an interval of 10°) were fed into the network to obtain the predicted image shown in [Fig. 18(f)]. The stacked image of the seven input

Authorized licensed use limited to: University of Science & Technology of China. Downloaded on January 26,2023 at 01:10:50 UTC from IEEE Xplore. Restrictions apply.

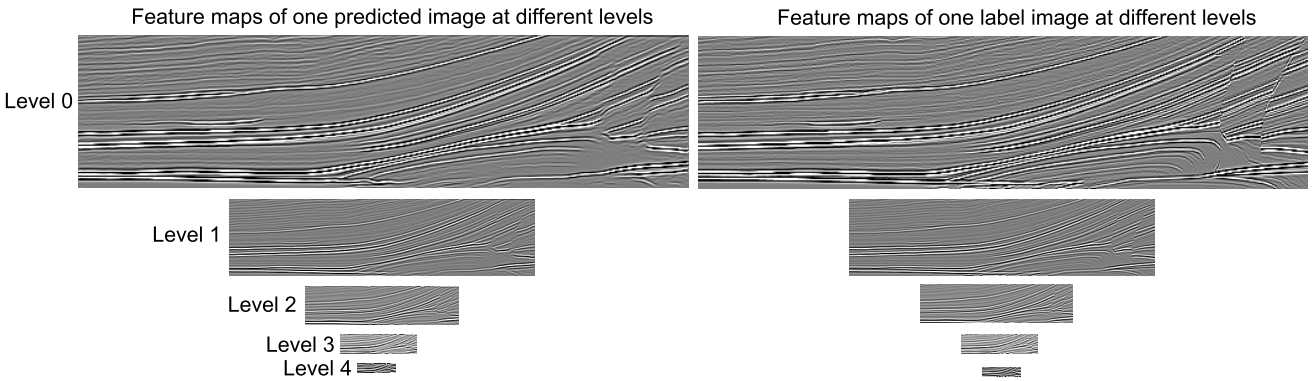


Fig. 15. Feature maps of predicted images and labels at different levels.

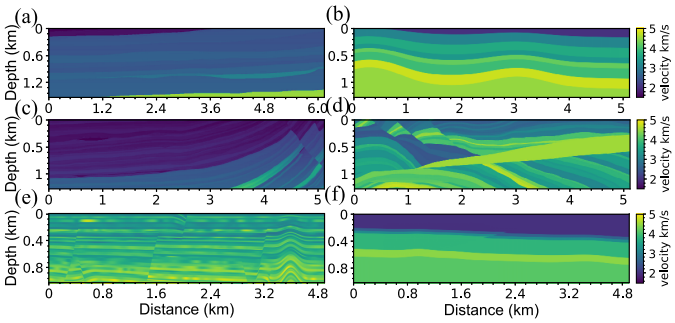


Fig. 16. Examples of velocity models with different structures for testing the trained model.

image gathers is shown in Fig. 18(e). The image in Fig. 18(e) is heavily affected by crosstalk noise and has poor imaging quality. Similar to the ground truth depicted in Fig. 18(b), the predicted image [Fig. 18(f)] eliminates crosstalk artifacts from the input image gathers and improves the imaging resolution for high-quality imaging of subsurface fine structures.

We analyze the wavenumber components by transforming the imaging results in the spatial domain in Fig. 18(c), (d), and (f) to the wavenumber domain via the 2-D Fourier transform. Fig. 19(a)–(c) show the wavenumber-domain images of conventional PWRTM, RTM, and the proposed method, respectively. We observe that compared with the conventional PWRTM and RTM, our method can broaden the frequency band and increase the frequency component, verifying the effectiveness of the proposed method in improving the imaging resolution.

The processing time required to achieve the imaging results illustrated in Fig. 18 is presented in Table II. The RTM stacked image with 277 shots required 28 790 s, and the PWRTM stacked image with 71 angles required 4602 s, which is approximately 16% of the time required for the traditional RTM. The processing time required for the two stacked images is large and unacceptable. In comparison, the proposed method takes only 470 s, which is approximately 1.6% and 10% of the time required for the traditional and PWRTM, respectively. The extremely high-computational efficiency of our method is because of the lower number of migrations. The proposed method, involving only seven migrations, achieves both crosstalk denoising and imaging enhancement through optimal superposition of imaging gathers and maintains high-computational efficiency. The proposed method breaks through

TABLE II

PROCESSING TIME REQUIRED FOR THE DIFFERENT IMAGING RESULTS PRESENTED IN FIG. 18

Methods	RTM with 277 shots	PWRTM with 71 angles	PWRTM with 7 angles	Our method
Processing time (second)	28, 790	4, 602	469	**470**

the contradiction between computational efficiency and imaging accuracy to obtain high-quality imaging results with low-computational cost. The predicted imaging results enable the characterization of fine subsurface structures.

V. DISCUSSION

In this study, we propose a deep learning method for enhancing PWRTM via a CNN. The network predicts high-quality seismic images by inputting plane-wave images at seven incident angles. Seven plane-wave images mean that only seven times of migration need to be executed, which can significantly save computational time. The functions of the network include optimal stacking of input plane-wave images, noise removal, and imaging enhancement. Traditional plane-wave methods perform linear stacking of multiple plane-wave images of different incident angles, limiting their ability in improving imaging quality. In comparison, networks can perform nonlinear stacking of plane-wave images, a task in which the networks are skilled. In fact, seven plane-wave images are fed into a CNN in a channel-wise fashion, and convolutional kernels act on these seven images and stack them. CNNs have been shown to have strong capabilities in image denoising, in our research task, both crosstalk artifacts and random noise can be suppressed by a CNN. The labels are high-resolution images convolved from the corresponding reflectivity models, and this can help the network to improve the resolution of the predicted images and achieve imaging enhancement. The test results demonstrate that the trained model can achieve the three functions mentioned above. The proposed method effectively alleviates the conflict between computational efficiency and accuracy in seismic migration and can achieve high-resolution imaging results with low-computational effort.

A shortcoming of the method is that the salt mound model was not tested. Theoretically, the proposed method can help to improve the imaging quality under the salt dome. For the salt dome models, the plane-wave migrated images at different

Authorized licensed use limited to: University of Science & Technology of China. Downloaded on January 26,2023 at 01:10:50 UTC from IEEE Xplore. Restrictions apply.

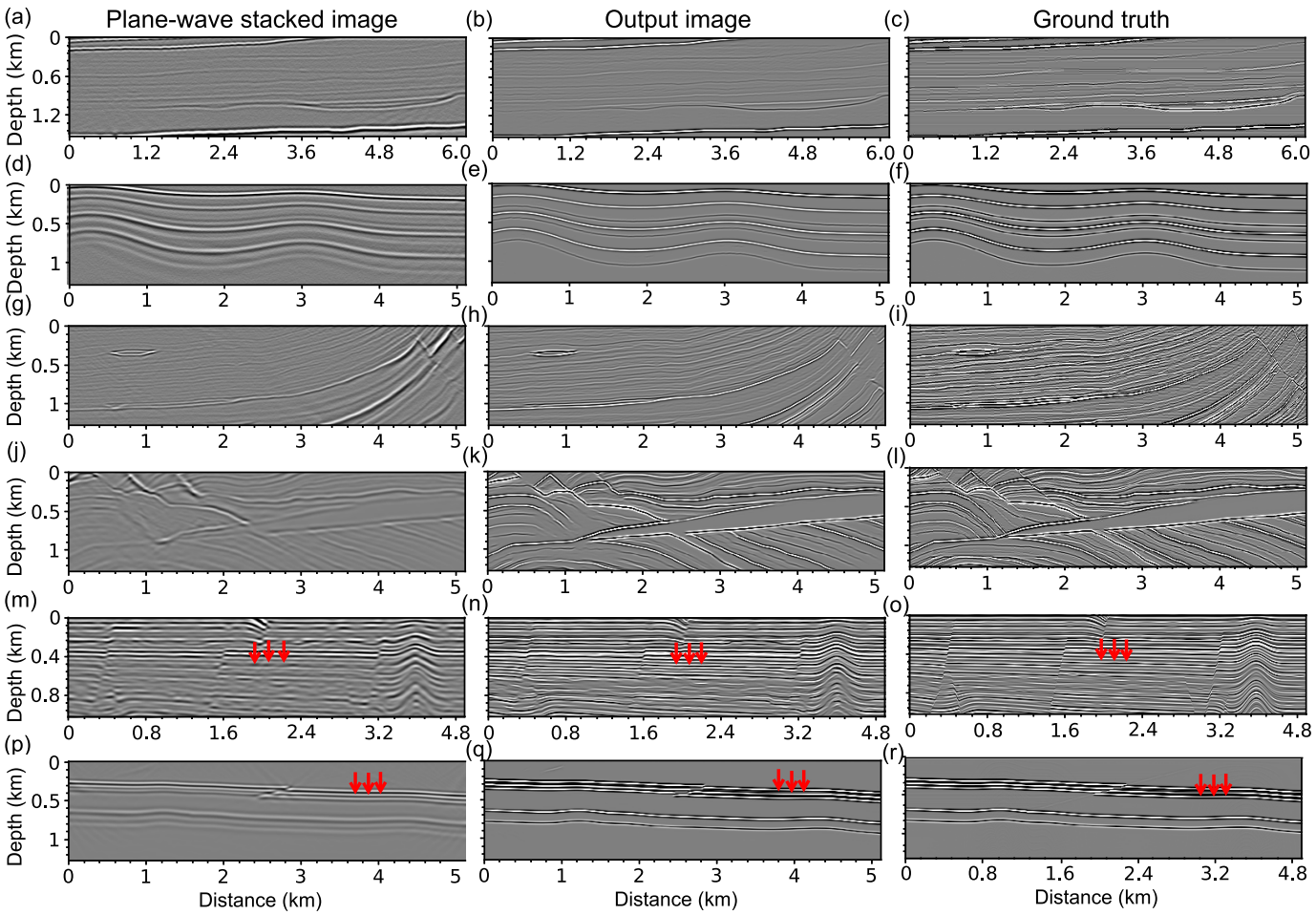


Fig. 17. Imaging results corresponding to the velocities depicted in Fig. 16. Each row of images corresponds to a velocity model in Fig. 16(a)–(f), respectively. The images in the first column are the plane-wave stacked images, which are the summation of input image gathers. The second column shows the output of the network. The images in the third column are ground truths (high-resolution images convolved from the corresponding reflectivity models).

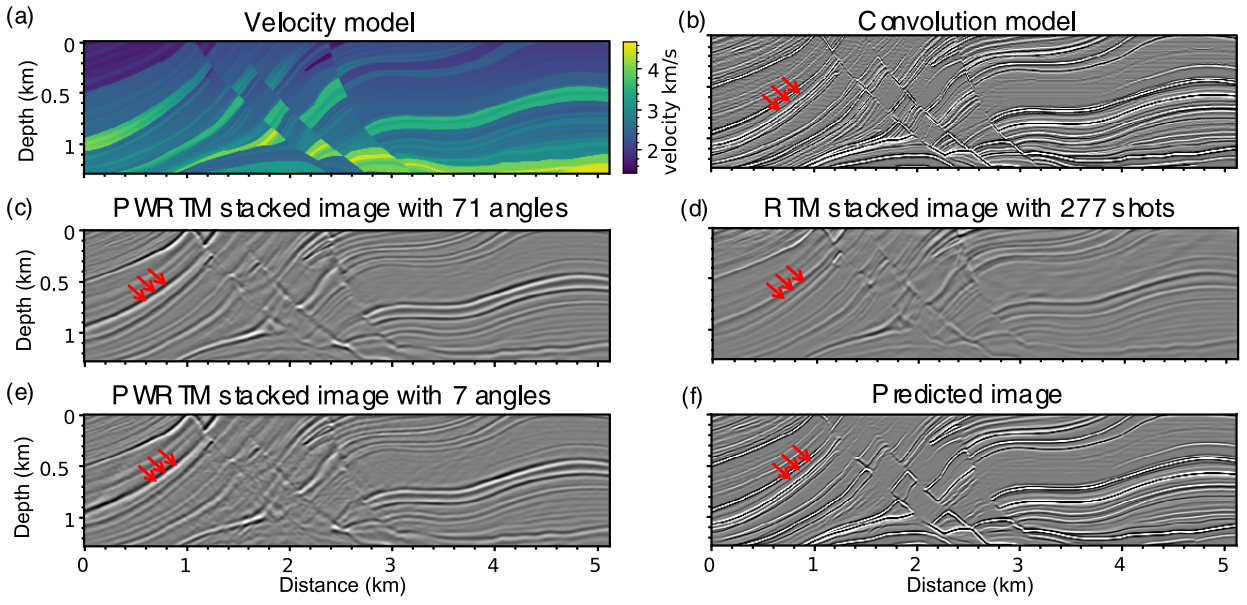


Fig. 18. Velocity model with complex structures and its imaging results. (a) True velocity model. (b) High-resolution image convolved from the corresponding reflectivity model using Ricker wavelet with a dominant frequency of approximately 70 Hz. (c) PWRTM stacked image with 71 incident angles (from $-35^\circ \sim 35^\circ$ with an interval of 1°). (d) RTM stacked image with 277 shots. (e) PWRTM stacked image with seven incident angles ($-30^\circ \sim 30^\circ$ with an interval of 10°). (f) Predicted image.

Authorized licensed use limited to: University of Science & Technology of China. Downloaded on January 26,2023 at 01:10:50 UTC from IEEE Xplore. Restrictions apply.

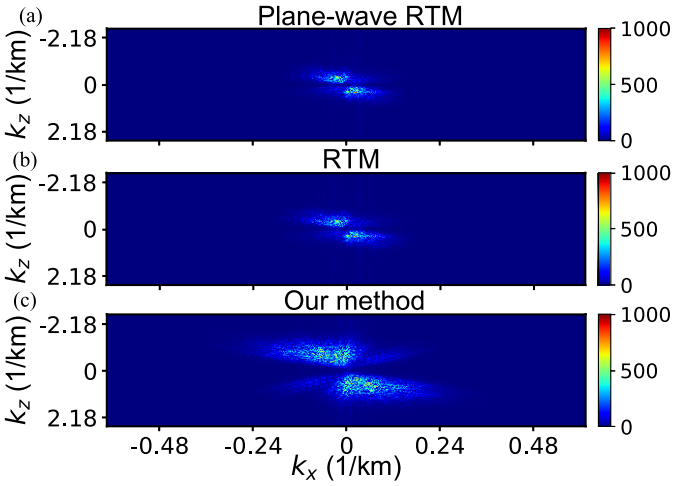


Fig. 19. Wavenumber-domain images corresponding to the different methods. The spatial domain seismic images in Fig. 18(c), (d), and (f) are converted to the wavenumber domain via 2-D Fourier transform. (a) Wavenumber image of conventional PWRTM. (b) Wavenumber image of conventional RTM. (c) Wavenumber image of the proposed method.

incident angles contain the signals of subsalt structures, only its energy is weak. We believe that the designed network has the ability to extract the features of subsalt images. Improving the quality of subsalt images with the studied method requires building a number of salt models as training data. We hope to further extend the application of the research method by adding salt models in later studies.

A problem with seismic migration has a large dependence on the migrated velocity models. Similar to traditional migration methods, the proposed method requires a relatively accurate velocity model. If the velocity is not accurate enough, it is difficult for the proposed method to obtain a favorable imaging result. The study in this article is designed to address the problem of conflicting computational efficiency and accuracy of seismic migration, and do not consider the effect of velocity errors on the predicted images. In contrast, our other two studies are to solve the problem of the strong dependence of seismic migration on the migrated velocities [55], [64]. Geng *et al.* [55] uses a CNN to analyze the features of the distorted imaging gathers induced by velocity errors to predict accurate velocity models. Li *et al.* [64] uses a kernel prediction network to correct for distorted imaging gathers due to velocity errors, achieving accurate imaging under inaccurate velocity models.

VI. Conclusion

In this study, we presented a CNN for enhancing PWRTM. The network performs an optimal stacking of input noisy plane-wave CAIGs and predicts a noise-free and high-quality image. To train the network, we designed a workflow to build numerous training datasets with diverse structures. At each iteration of the training and validation process, random extraction of migration images helped expand the dataset. A Laplacian pyramid loss function was defined to train the network architecture. The results of training and validation experiments demonstrate that our trained CNN is effective in eliminating the artifacts and improving the imaging quality of plane-wave CAIGs. Compared with the traditional RTM and the conventional use of several angles for plane-wave migration, the proposed method requires image gathers with only seven incident angles, improving both computational efficiency and imaging quality. The predicted images have higher resolution and can clearly present the fine subsurface structures. In terms of time consumption, the proposed method requires approximately 1.6% and 10% of the time required by RTM and PWRTM, respectively, for processing a provided test model with a size of 1024×256. Extending the image denoising and enhancing tasks for 2-D plane-wave migration to 3-D migration is straightforward by directly replacing the 2-D convolutional kernel with a 3-D kernel. However, building the training datasets requires a relatively large amount of computation compared with the 2-D case. The method described in this study has the potential for generalization to other geophysical imaging methods.

Acknowledgment

The authors are deeply grateful to the Associate Editor Dr. Yangkang Chen and two anonymous reviewers for their careful work and thoughtful suggestions that have helped improve this article substantially. The authors also acknowledge the critical comments received from Dr. Yunzhi Shi, Wenyang Chen, Bin Chen, and Jintao Li.

References

[1] G. Cambois, S. Ronen, and X. Zhu, "Wide-azimuth acquisition: True 3D at last!" *Lead. Edge*, vol. 21, no. 8, p. 763, 2002.
[2] N. Moldoveanu, J. Kapoor, and M. Egan, "Full-azimuth imaging using circular geometry acquisition," *Lead. Edge*, vol. 27, no. 7, pp. 908–913, 2008.
[3] S. H. Gray, *Seismic, Migration*. Dordrecht, The Netherlands: Springer, 2011, pp. 1236–1244.
[4] J. B. Bednar, "A brief history of seismic migration," *Geophysics*, vol. 70, no. 3, pp. 3MJ–20MJ, 2005.
[5] J. Etgen, S. H. Gray, and Y. Zhang, "An overview of depth imaging in exploration geophysics," *Geophysics*, vol. 74, no. 6, pp. WCA5–WCA17, 2009.
[6] W. S. French, "Computer migration of oblique seismic reflection profiles," *Geophysics*, vol. 40, no. 6, pp. 961–980, Dec. 1975.
[7] W. A. Schneider, "Integral formulation for migration in two and three dimensions," *Geophysics*, vol. 43, no. 1, pp. 49–76, 1978.
[8] B. Biondi, "Kirchhoff imaging beyond aliasing," *Geophysics*, vol. 66, no. 2, pp. 654–666, 2001.
[9] N. R. Hill, "Prestack Gaussian-beam depth migration," *Geophysics*, vol. 66, no. 4, pp. 1240–1250, 2001.
[10] Y. Sun, F. Qin, S. Checkles, and J. P. Leveille, "3-D prestack Kirchhoff beam migration for depth imaging," *Geophysics*, vol. 65, no. 5, pp. 1592–1603, 2000.
[11] S. Jin, R.-S. Wu, and C. Peng, "Seismic depth migration with pseudo-screen propagator," *Comput. Geosci.*, vol. 3, no. 3, pp. 321–335, 1999.
[12] J. M. H. Le Rousseau and M. V. de Hoop, "Modeling and imaging with the scalar generalized-screen algorithms in isotropic media," *Geophysics*, vol. 66, no. 5, pp. 1551–1568, 2001.
[13] E. Baysal, D. D. Kosloff, and J. W. Sherwood, "Reverse time migration," *Geophysics*, vol. 48, no. 11, pp. 1514–1524, 1983.
[14] H.-W. Zhou, H. Hu, Z. Zou, Y. Wo, and O. Youn, "Reverse time migration: A prospect of seismic imaging methodology," *Earth-Sci. Rev.*, vol. 179, pp. 207–227, Apr. 2018.
[15] W. Dai and G. T. Schuster, "Plane-wave least-squares reverse-time migration," *Geophysics*, vol. 78, no. 4, pp. S165–S177, 2013.
[16] J. Godwin and P. Sava, "A comparison of shot-encoding schemes for wave-equation migration," *Geophys. Prospecting*, vol. 61, pp. 391–408, Jan. 2013.

Authorized licensed use limited to: University of Science & Technology of China. Downloaded on January 26,2023 at 01:10:50 UTC from IEEE Xplore. Restrictions apply.

[17] J. Hu, H. Wang, L. Zhao, Y. Shao, M. Wang, and A. Osen, "Efficient reverse time migration with amplitude encoding," *J. Geophys. Eng.*, vol. 12, no. 4, pp. 601–609, 2015.

[18] F. Perrone and P. Sava, "Wave-equation migration with dithered plane waves," *Geophys. Prospecting*, vol. 60, no. 3, pp. 444–465, 2012.

[19] L. A. Romero, D. C. Ghiglia, C. C. Ober, and S. A. Morton, "Phase encoding of shot records in prestack migration," *Geophysics*, vol. 65, no. 2, pp. 426–436, Mar. 2000.

[20] G. T. Schuster, X. Wang, Y. Huang, W. Dai, and C. Boonyasiriwat, "Theory of multisource crosstalk reduction by phase-encoded statics," *Geophys. J. Int.*, vol. 184, no. 3, pp. 1289–1303, 2011.

[21] P. Temme, "A comparison of common-midpoint, single-shot, and plane-wave depth migration," *Geophysics*, vol. 49, no. 11, pp. 1896–1907, 1984.

[22] H. J. Tieman, "Improving plane-wave decomposition and migration," *Geophysics*, vol. 62, no. 1, pp. 195–205, 1997.

[23] X. Jia, B. Chen, and Q. Li, "A bipolar-bisection piecewise encoding scheme for multi-source reverse time migration," *Commun. Comput. Phys.*, vol. 28, no. 2, pp. 723–742, 2020.

[24] X. Jia, W. Chen, and B. Chen, "An efficient super-virtual shot encoding scheme for multisource reverse time migration," *Geophysics*, vol. 85, pp. S405–S416, Nov. 2020.

[25] Z. Xue, Y. Chen, S. Fomel, and J. Sun, "Seismic imaging of incomplete data and simultaneous-source data using least-squares reverse time migration with shaping regularization," *Geophysics*, vol. 81, no. 1, pp. S11–S20, 2016.

[26] Y. Chen, H. Chen, K. Xiang, and X. Chen, "Preserving the discontinuities in least-squares reverse time migration of simultaneous-source data," *Geophysics*, vol. 82, no. 3, pp. S185–S196, 2017.

[27] Y. Chen, M. Bai, Y. Zhou, Q. Zhang, Y. Wang, and H. Chen, "Substituting smoothing with low-rank decomposition—Applications to least-squares reverse time migration of simultaneous source and incomplete seismic data," *Geophysics*, vol. 84, no. 4, pp. S267–S283, 2019.

[28] Q. Zhang, W. Mao, and Y. Chen, "Attenuating crosstalk noise of simultaneous-source least-squares reverse time migration with GPU-based excitation amplitude imaging condition," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 587–597, Aug. 2018.

[29] J. Wu and M. Bai, "Incoherent dictionary learning for reducing crosstalk noise in least-squares reverse time migration," *Comput. Geosci.*, vol. 114, pp. 11–21, May 2018.

[30] S.-J. Lee, T. Chen, L. Yu, and C.-H. Lai, "Image classification based on the boost convolutional neural network," *IEEE Access*, vol. 6, pp. 12755–12768, 2018.

[31] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis. (DV)*, Oct. 2016, pp. 565–571.

[32] S. Ren, K. He, R. Girshick, X. Zhang, and J. Sun, "Object detection networks on convolutional feature maps," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 7, pp. 1476–1481, Jul. 2016.

[33] K. Zhang, W. Zuo, and L. Zhang, "FFDNet: Toward a fast and flexible solution for CNN-based image denoising," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4608–4622, Sep. 2018.

[34] D. Liu, W. Wang, X. Wang, C. Wang, J. Pei, and W. Chen, "Poststack seismic data denoising based on 3-D convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 3, pp. 1598–1629, Nov. 2019.

[35] H. Wang, Y. Li, and X. Dong, "Generative adversarial network for desert seismic data denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 8, pp. 7062–7075, Nov. 2020.

[36] X. Wang and J. Ma, "Adaptive dictionary learning for blind seismic data denoising," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 7, pp. 1273–1277, Jul. 2020.

[37] W. Li and J. Wang, "Residual learning of cycle-GAN for seismic data denoising," *IEEE Access*, vol. 9, pp. 11585–11597, 2021.

[38] C. Guo, T. Zhu, Y. Gao, S. Wu, and J. Sun, "AEnet: Automatic picking of P-wave first arrivals using deep learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 5293–5303, Jul. 2020.

[39] Z. E. Ross, M.-A. Meier, and E. Hauksson, "P wave arrival picking and first-motion polarity determination with deep learning," *J. Geophys. Res.*, vol. 123, no. 6, pp. 5120–5129, Jun. 2018.

[40] S. Han, Y. Liu, Y. Li, and Y. Luo, "First arrival traveltime picking through 3-D U-Net," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2021.

[41] W. Zhang and J. Gao, "High-resolution velocity model building based on common-source migration images and convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2021.

[42] F. Yang and J. Ma, "Deep-learning inversion: A next-generation seismic velocity model building method," *Geophysics*, vol. 84, no. 4, pp. R583–R599, 2019.

[43] M. Araya-Polo, J. Jennings, A. Adler, and T. Dahlke, "Deep-learning tomography," *Lead. Edge*, vol. 37, no. 1, pp. 58–66, 2018.

[44] J. Yu and B. Wu, "Attention and hybrid loss guided deep learning for consecutively missing seismic data reconstruction," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–8, 2021.

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

[46] X. Wu, S. Yan, Z. Bi, S. Zhang, and H. Si, "Deep learning for multidimensional seismic impedance inversion," *Geophysics*, vol. 86, no. 5, pp. R735–R745, 2021.

[47] X. Wu, Y. Shi, S. Fomel, L. Liang, Q. Zhang, and A. Z. Yusifov, "FaultNet3D: Predicting fault probabilities, strikes, and dips with a single convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 9138–9155, Nov. 2019.

[48] Z. Bi, X. Wu, Z. Geng, and H. Li, "Deep relative geologic time: A deep learning method for simultaneously interpreting 3-D seismic horizons and faults," *J. Geophys. Res., Solid Earth*, vol. 126, no. 9, 2021, Art. no. e2021JB021882.

[49] Y. Shi, X. Wu, and S. Fomel, "SaltSeg: Automatic 3D salt segmentation using a deep convolutional neural network," *Interpretation*, vol. 7, no. 3, pp. SE113–SE122, 2019.

[50] G. Matharu, W. Gao, R. Lin, Y. Guo, M. Park, and M. D. Sacchi, "Simultaneous source deblending using a deep residual network," in *Proc. Workshop, Math. Geophys., Traditional Learn.*, Beijing, China, Nov. 2019, pp. 13–16.

[51] S. Zu, J. Cao, S. Qu, and Y. Chen, "Iterative deblending for simultaneous source data using the deep neural network," *Geophysics*, vol. 85, no. 2, pp. V131–V141, 2020.

[52] A. Richardson and C. Feller, "Seismic data denoising and deblending using deep learning," 2019, *arXiv:1907.01497*.

[53] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 234–241.

[54] Z. Geng, X. Wu, Y. Shi, and S. Fomel, "Deep learning for relative geologic time and seismic horizons," *Geophysics*, vol. 85, no. 4, pp. WA87–WA100, 2020.

[55] Z. Geng, Z. Zhao, Y. Shi, X. Wu, S. Fomel, and M. Sen, "Deep learning for velocity model building with common-image gather volumes," *Geophys. J. Int.*, vol. 228, no. 2, pp. 1054–1070, 2021.

[56] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

[57] I. Laina, C. Rupprecht, V. Belagiannis, F. Tombari, and N. Navab, "Deeper depth prediction with fully convolutional residual networks," in *Proc. 4th Int. Conf. 3D Vis. (DV)*, 2016, pp. 239–248.

[58] P. Bojanowski, A. Joulin, D. Lopez-Paz, and A. Szlam, "Optimizing the latent space of generative networks," 2017, *arXiv:1707.05776*.

[59] X. Wu, Z. Geng, Y. Shi, N. Pham, S. Fomel, and G. Caumon, "Building realistic structure models to train convolutional neural networks for seismic structural interpretation," *Geophysics*, vol. 85, no. 4, pp. WA27–WA39, 2020.

[60] G. S. Martin, R. Wiley, and K. J. Marfurt, "Marmousi2: An elastic upgrade for Marmousi," *Lead. Edge*, vol. 25, no. 2, pp. 156–166, Jan. 2006.

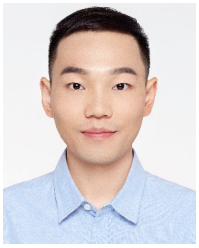
[61] M. D. Bloice, P. M. Roth, and A. Holzinger, "Biomedical image augmentation using Augmentor," *Bioinformatics*, vol. 35, no. 21, pp. 4522–4524, 2019.

[62] A. Paszke *et al.*, "PyTorch: An imperative style, high-performance deep learning library," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, Dec. 2019, pp. 8026–8037.

[63] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[64] Z. Li, X. Wu, L. Liang, and X. Jia, "Kernel prediction network for common image gather stacking," in *SEG Technical Program Expanded Abstracts*. Houston, TX, USA: SEG, 2020, pp. 2988–2992.

Authorized licensed use limited to: University of Science & Technology of China. Downloaded on January 26,2023 at 01:10:50 UTC from IEEE Xplore. Restrictions apply.



Yaxing Li received the Ph.D. degree in geophysics from the University of Science and Technology of China (USTC), Hefei, China, in 2022.

He is the Post-Doctoral Fellow with the Computational Interpretation Group (CIG), USTC. His research interests include geophysical data processing including denoising, seismic migration, and inversion.

Xinming Wu (Associate Member, IEEE) received the Ph.D. degree in geophysics from the Colorado School of Mines, Golden, CO, USA, in 2016.

He was a Post-Doctoral Fellow with the Bureau of Economic Geology, University of Texas at Austin, Austin, TX, USA. He is a Professor with the School of Earth and Space Sciences, University of Science and Technology of China (USTC), Hefei, China. His research interests include image processing, machine learning, 3-D seismic interpretation, subsurface modeling, and geophysical inversion.

Dr. Wu received the Society of Exploration Geophysicists (SEG's) awards for J. Clarence Karcher Award, 2020 and the Honorary Lecturer, South and East Asia, 2020. He was also a recipient of the Best Paper Award in Geophysics in 2016, the Best Student Poster Paper Award at the 2017 SEG Annual Convention, and the Honorable Mention Award for Best Paper at the 2018 SEG Annual Convention.

Xiaofeng Jia received the Ph.D. degree in geophysics from Peking University, Beijing, China, in 2005.

He was a Post-Doctoral Fellow with the University of California, Santa Cruz, CA, USA. He is an Associate Professor with the School of Earth and Space Sciences, University of Science and Technology of China, Hefei, China. His research interests include seismic numerical modeling, seismic processing and migration, and machine learning.

Zhicheng Geng received the B.S. degree in geophysics from Tongji University, Shanghai, China, in 2017, and the Ph.D. degree in geophysics from the University of Texas at Austin, Austin, TX, USA, in 2022, where he was supervised by Dr. Sergey Fomel.

He is an Applied Scientist with the Amazon Machine Learning Solutions Laboratory, Austin. He currently focuses on deep-learning applications to solve geophysical problems.

Authorized licensed use limited to: University of Science & Technology of China. Downloaded on January 26,2023 at 01:10:50 UTC from IEEE Xplore. Restrictions apply.