

Deep-learning-based seismic data interpolation: A preliminary result

Benfeng Wang¹, Ning Zhang², Wenkai Lu², and Jialin Wang²

ABSTRACT

Seismic data interpolation is a longstanding issue. Most current methods are only suitable for randomly missing cases. To deal with regularly missing cases, an antialiasing strategy should be included. However, seismic survey design using a random distribution of shots and receivers is always operationally challenging and impractical. We have used deep-learning-based approaches for seismic data antialiasing interpolation, which could extract deeper features of the training data in a nonlinear way by self-learning. It can also avoid linear events, sparsity, and low-rank assumptions of the traditional interpolation methods. Based on convolutional neural networks, eight-layers residual learning networks (ResNets) with a better back-propagation property for deep layers is designed for interpolation. Detailed training analysis is also performed. A set of simulated

data is used to train the designed ResNets. The performance is assessed with several synthetic and field data. Numerical examples indicate that the trained ResNets can help to reconstruct regularly missing traces with high accuracy. The interpolated results in the time-space domain and the frequency-wavenumber (f - k) domain demonstrate the validity of the trained ResNets. Even though the accuracy decreases with the increase of the feature difference between the test and training data, the proposed method can still provide reasonable interpolation results. Finally, the trained ResNets is used to reconstruct dense data with halved trace intervals for synthetic and field data. The reconstructed dense data are more continuous along the spatial direction, and the spatial aliasing effects disappear in the f - k domain. The reconstructed dense data have the potential to improve the accuracy of subsequent seismic data processing and inversion.

INTRODUCTION

Because of existing obstacles and economic restrictions, there are always traces missing in acquired seismic data. The missing traces are always irregularly distributed along spatial coordinates or regularly distributed on a coarse grid. Missing traces can affect the accuracy of amplitude-variation-with-offset analysis, surface-related multiple elimination (SRME), inversion, and migration; therefore, seismic data interpolation is a prerequisite for these mentioned algorithms (Wang et al., 2015a, 2016).

There are different kinds of interpolation methods, such as sparse mathematical transform-based methods (Duijndam and Schoneville, 1999; Trad et al., 2002; Liu and Sacchi, 2004; Abma and Kabir, 2006; Trad, 2009; Wang et al., 2014, 2015b; Wang, 2016), rank-reduction-based methods (Trickett et al., 2010; Gao et al.,

2013, 2015; Ma, 2013; Chen et al., 2016), frequency space (f - x) prediction filter-based methods (Spitz, 1991; Porsani, 1999; Naghizadeh and Sacchi, 2007), and wave-equation-based methods (Ronen, 1987; Fomel, 2003). For the wave-equation-based interpolation method, it is not only computationally intensive, but it also requires the velocity model as a prerequisite, even though it can handle randomly missing cases and regularly missing cases. The f - x prediction filter-based methods can deal with regularly sampled seismic data; however, it uses the predictability of linear events in the f - x domain. This means that a local window strategy has to be used to guarantee the linear property for nonlinear events. The window selection can affect the interpolation performance, and optimal window selection is impractical currently. Rank-reduction-based interpolation methods assume that seismic data in a specific arrangement are of low rank at a given frequency component

Manuscript received by the Editor 29 July 2017; revised manuscript received 24 May 2018; published ahead of production 25 September 2018; published online 23 November 2018.

¹Tongji University, State Key Laboratory of Marine Geology, School of Ocean and Earth Science, Institute for Advanced Study, Shanghai 200092, China and Tsinghua University, Easysignal Group, State Key Laboratory of Intelligent Technology and Systems, Tsinghua National Laboratory for Information Science and Technology, Department of Automation, Beijing 100084, China. E-mail: wbf1232007@126.com.

²Tsinghua University, Easysignal Group, State Key Laboratory of Intelligent Technology and Systems, Tsinghua National Laboratory for Information Science and Technology, Department of Automation, Beijing 100084, China. E-mail: zhng1230@163.com; lwkmf@mail.tsinghua.edu.cn; sherry_wangjl12@126.com. © 2019 Society of Exploration Geophysicists. All rights reserved.

(Trickett et al., 2010; Gao et al., 2013, 2017) and that the missing traces can increase its rank. Seismic data interpolation can be achieved through rank-reduction algorithms, and the rank selection is always determined approximately by the number of linear events in the analyzed local window. Rank-reduction interpolation algorithms are always suitable for linear or quasilinear events theoretically; therefore, they should be implemented for normal-moveout (NMO)-corrected data or carried out using local strategy for pre-stack curved events. Even though they can be implemented directly on curved prestack seismic data in practice, it is difficult to determine the optimal rank selection, resulting in unsatisfactory interpolation results (Ma, 2013). Tensor completion is also introduced for seismic data reconstruction (Kreimer et al., 2013). The above-mentioned methods are suitable for irregularly sampled seismic data obtained using a random under-sampling strategy (Ma, 2013; Gao et al., 2015). For regularly sampled seismic data in which the spatial aliasing effects appear, an antialiasing interpolation strategy has to be included to improve the interpolation performance. In this case, interpolation methods based on sparse mathematical transform can play an important role. In the frequency-wavenumber (f - k) domain, a mask function is first designed using dip scanning to constrain the f - k energy distribution, and then antialiasing interpolation is achieved using the projection onto convex sets method (Abma and Kabir, 2006; Gao et al., 2012), based on the Fourier transform. Fourier-transform-based methods assume that seismic events are linear or quasilinear. Even though they can handle curved events in practice, it is impractical to get optimal parameters, which can affect the final interpolation performance. In the curvelet transform domain, the mask function in the aliasing-free scales is first established using a thresholding strategy, and then it is upscaled to estimate the mask function in the aliased scales for antialiasing seismic data interpolation (Naghizadeh and Sacchi, 2010). The computational cost is high for the curvelet transform. Antialiasing seismic data interpolation in the seislet domain is also researched (Gan et al., 2015; Liu et al., 2016), where the event slope is first estimated using lower frequency components, then updated using higher frequency components during interpolation iterations. Even though transform-based methods can handle aliased seismic data interpolation issues to some extent, the performance is always sensitive to the selected parameters. For the Fourier and curvelet transforms, the estimation error of the mask function can affect the interpolation performance directly. For the seislet transform, the estimation error of the local event slope also affects the final performance of the reconstructed seismic data. Therefore, more strategies about antialiasing seismic data interpolation should be investigated for subsequent seismic data processing and refined reservoir description and characterization.

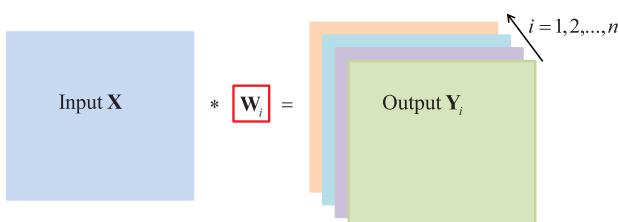


Figure 1. A simple cartoon for describing CNN, \mathbf{X} denotes the input, \mathbf{W}_i represents one of the kernels, and \mathbf{Y}_i is the output corresponding to the kernel \mathbf{W}_i .

With the rise of compressive sensing, rank-reduction-based methods and sparse mathematical transform or tight-frame-based methods have drawn much attention. However, most of them are just suitable for randomly missing cases, and seismic survey design based on a random distribution of shots and receivers is always challenging and impractical. This is the reason why weave sampling and double-weave sampling strategies with regular sampling patterns are proposed (Naghizadeh, 2015). For regularly sampled seismic data with spatial aliasing, antialiasing strategies should be included into these mentioned methods. Because they characterize seismic data in a linear way, the characterization accuracy can decrease for complicated seismic data. With the development of machine learning, Jia and Ma (2017) explore the potential of support vector regression (SVR) in seismic data interpolation for regularly missing cases, which extracts low-level features contained in seismic data. Deep learning (DL), which belongs to the machine-learning category, has attracted much attention in recent years because it can extract high-level features of training data in a nonlinear way through deep self-learning compared with SVR. DL has been successfully used in superresolution image reconstruction (Dong et al., 2014; Kim et al., 2016), gas-chimney detection (Xu et al., 2017), time-lapse reservoir property change estimation (Cao and Roy, 2017), and geologic features and identification (Huang et al., 2017). Except for the above-mentioned applications, wider applications are under way.

In this paper, we introduce the DL strategy into seismic data interpolation to provide accurately reconstructed dense data. The traditional interpolation methods require different parameter selections for different data sets to generate the optimal interpolation performance. This requires a great deal of human-computer interaction with a low degree of automation, especially for large volumes of data. As a comparison, the main computational cost of the DL strategy comes from the optimal network training, which happens once up front. After training, the interpolation computational cost is negligible, thus making the overall computational cost efficient in an automatic way, especially for large volumes of data. For the DL-based interpolation, the convolutional-neural-network-(CNN)-based residual learning networks (ResNets) are used. The “Theory” section provides a detailed procedure and parameter selection to design a suitable framework using ResNets blocks. A set of simulated data is used to train the designed ResNets, and the performance is assessed by different data sets. Furthermore, the trained ResNets is used to provide dense data with halved trace intervals for synthetic and field data. Finally, conclusions and further research suggestions are provided.

Theory

A brief illustration of CNN

CNN was first proposed by LeCun et al. (1998). It is a special neural network, which is well known for the existence of plenty of convolutional operations. A simple cartoon for the first convolutional layer of CNN is shown in Figure 1.

In Figure 1, the left blue matrix represents the input X in the first convolutional layer of CNN, the kernel \mathbf{W}_i is marked by the red square, and $*$ it represents the convolution operator. The output \mathbf{Y}_i of this layer can be obtained by the convolution between \mathbf{W}_i and \mathbf{X} . Here, \mathbf{Y}_i is also called the feature map, which can be used for data characterization. If several kernels exist, a series of feature

maps can be obtained, as shown in the right part of Figure 1. Here, we assume that n kernels exist. Even though CNN has been proposed for a long time, the applications were limited until the ImageNet Large Scale Visual Recognition Challenge in 2012, during which AlexNet (Krizhevsky et al., 2012), based on CNN, achieved the first prize with 10% higher accuracy than that of the second prize. Since then, a quantity of researchers and companies started to study CNN for deep research, and many classic network architectures have been proposed, such as Visual Geometry Group (Simonyan and Zisserman, 2015), GoogLeNet (Szegedy et al., 2015), and ResNets (He et al., 2016). In this paper, we introduce CNN-based ResNets blocks into seismic data processing, i.e., seismic data interpolation. The reason that we choose the ResNets framework for seismic data interpolation is because the ResNets allow deeper neural networks with a better back-propagation property, and it is also easier to optimize with a higher reconstruction accuracy compared with other network frameworks (He et al., 2016). Detailed illustrations and statements are provided in the following sections.

Definition of the loss function using CNN

In the missing positions, the regularly sampled data \mathbf{d}_{obs} are inconsistent with the true complete data \mathbf{d}_c . To improve the consistency, \mathbf{d}_{obs} is initially interpolated for the following higher training accuracy using the bicubic interpolation algorithm to generate \mathbf{d}_0 , i.e., the input of the designed network. The loss function using CNN can be characterized by

$$\varphi(\boldsymbol{\theta}) = \frac{1}{2} \|f(\mathbf{d}_0; \boldsymbol{\theta}) - \mathbf{d}_c\|_2^2 + \lambda \|\mathbf{W}_{\text{set}}\|_2^2, \quad (1)$$

where \mathbf{d}_c denotes the true complete data, which is also regarded as a label; f denotes the designed network; and $\boldsymbol{\theta}$ represents the network parameters, which include the network depth, size, and number of kernels in each layer, the elements in each kernel, and the bias corresponding to each kernel. The term λ denotes the regularization parameter, and \mathbf{W}_{set} is the collection of all kernels distributed in each layer. When the trained network $\boldsymbol{\theta}_{\text{out}}$ is established, the output of the trained network can be achieved using

$$\mathbf{d}_{\text{int}} = f(\mathbf{d}_0; \boldsymbol{\theta}_{\text{out}}), \quad (2)$$

where \mathbf{d}_{int} is the interpolated data, which can be suitable for subsequent migration or signal-processing steps.

ResNets design for interpolation

CNN has been introduced into many fields. Dong et al. (2014) first introduce CNN to the superresolution image reconstruction issue, followed by more research, such as the very deep convolutional network for superresolution (Kim et al., 2016), deeply recursive convolutional network (Kim et al., 2016), and superresolution generative adversarial network (Ledig et al., 2016). Here, we regard seismic data interpolation along the spatial direction as a superresolution image reconstruction issue, and we design a ResNets architecture using ResNets blocks (He et al., 2016).

Figure 2a shows the CNN-based ResNets for seismic interpolation, which includes the first layer, the last layer, and three ResNets blocks. For a detailed interpretation of each ResNets block, Figure 2b is provided for the second ResNets block, in which there

are two convolutional layers. Therefore, the designed ResNets framework contains eight convolutional layers in total. In Figure 2, the red blocks represent the corresponding convolutional layer, with the size, name, and number of kernels illustrated. For example, there are 64 kernels with size 3×3 in the first seven layers. To make the designed ResNets robust, the nonlinear function, i.e., rectified linear units (ReLUs, $\max(0, x)$) (Krizhevsky et al., 2012), is used for nonlinear projection toward the obtained convolutional outputs in the first seven convolutional layers. The convolutional output is obtained by implementing the convolution operator toward the input of this layer plus the corresponding bias. The orange blocks denote the temporary results during ResNets training. The green block represents the input \mathbf{d}_0 of the designed ResNets and the corresponding label \mathbf{d}_c . The term \mathbf{d}_0 denotes the initially interpolated data using the bicubic interpolation algorithm. The purple block denotes the interpolation output $f(\mathbf{d}_0; \boldsymbol{\theta})$ of the designed ResNets, which is then compared with the label \mathbf{d}_c to construct the loss function, as marked by the blue block. The ResNets block in Figure 2b, as a subnetwork structure, is designed based on the residual learning strategy proposed by He et al. (2016). This residual learning structure can be beneficial to the back propagation of gradient information for deeper layers, and it can be useful for the network training with a high reconstruction accuracy.

Detailed training analysis

For convenient analysis, we denote the input of a specific convolutional layer as \mathbf{X} with size of $nt \times nx \times ch$, where nt and nx represent the sampling points of the local patch along the time axis and the space axis, respectively. For the input of a specific layer, ch denotes the number of channels. In this paper, local seismic data with size $41 \times 41 \times 1$ are regarded as the input of the first convolutional layer because, in this case, the number of channels equals unity for seismic data. We assume that there are n kernels ($\mathbf{W}_i, i = 1, 2, \dots, n = 64$) in each convolutional layer for the first

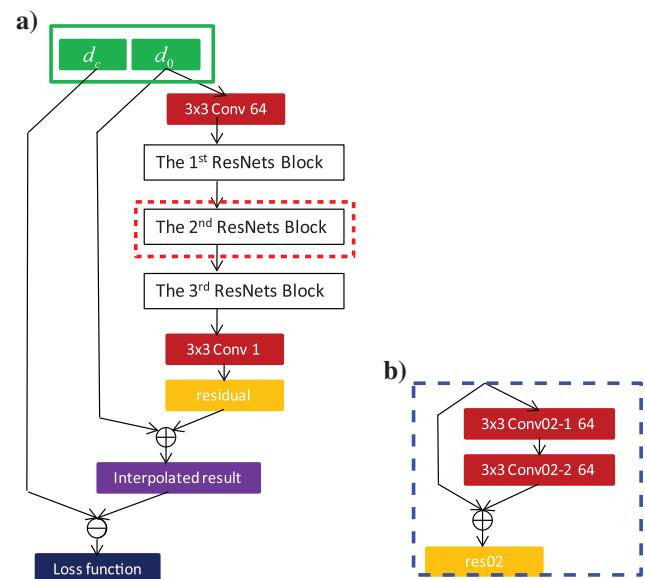


Figure 2. Designed ResNets with specified convolutional layers. (a) ResNets with eight convolutional layers and (b) detailed illustration for the ResNets block.

seven layers. The kernel size is $k \times k \times ch$ with $k = 3$. The output of this layer is denoted as \mathbf{Y} with size of $nt \times nx \times n$, and \mathbf{Y}_i , $i = 1, 2, \dots, n$ represents its components, which can be summarized as follows using the ReLUs nonlinear operator, i.e., $\max(0, x)$:

$$\mathbf{Y}_i = \max(0, \mathbf{W}_i * \mathbf{X} + \mathbf{b}_i), \quad i = 1, 2, \dots, n, \quad (3)$$

where \mathbf{W}_i represents the weighting matrix and \mathbf{b}_i is the corresponding bias variable of the kernel. For the input \mathbf{X} , a convolution operator is first implemented using \mathbf{W}_i , then, the bias variable \mathbf{b}_i corresponding to the kernel is summed, followed by the nonlinear ReLUs projection to obtain the final output \mathbf{Y}_i . If there are several kernels, a series of feature maps can be obtained to characterize the true seismic data nonlinearly.

Different from the first seven convolutional layers, the eighth layer of the designed ResNets just includes one kernel without the nonlinear ReLUs operator. Then, the output of this layer can be characterized using equation 4:

$$\mathbf{Y} = \mathbf{W} * \mathbf{X} + \mathbf{b}, \quad (4)$$

where \mathbf{W} and \mathbf{b} represent the weighting matrix and the corresponding bias variable in the eighth layer, respectively. After these operations in the eight convolutional layers, we can obtain the final interpolation output by summing with the initially interpolated data \mathbf{d}_0 . Next, the interpolation output is compared to the label \mathbf{d}_c to assess the current performance of the trained ResNets. To guarantee the designed ResNets with good performance, the initial ResNets parameters should be chosen properly.

In this paper, all of the numerical examples are carried out on Caffe platform (Jia et al., 2014). The size of the input training data is chosen as $41 \times 41 \times 1$. Thus, the true data should be divided into small patches with the size of 41×41 and the shifting step is set as 21 to overlap adjacent patches. The training method is based on the stochastic gradient descent with momentum (Sutskever et al., 2013), where the batch size is set as 128, the momentum parameter is set as 0.9, and the regularization weighting parameter is set as 0.0001. The learning rate is initially set as 0.01, and then it decreases to 0.001 after 200,000 iterations. After 400,000 iterations, the learning rate decreases even more to 0.0001, until the maximum iteration 520,000. To prevent gradient explosion, the parameter of the clip gradient is set as 0.5. These network parameters are set according to He et al. (2016), and more detailed illustrations and analysis can be found in that paper.

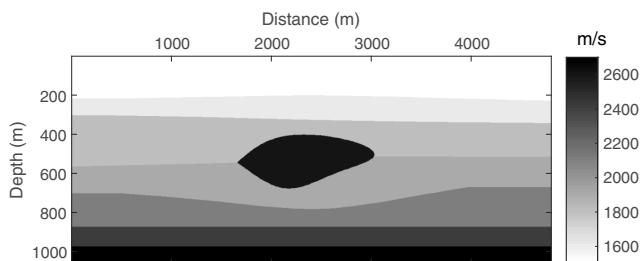


Figure 3. Velocity model of a simplified salt structure.

Numerical examples

Because the DL-based interpolation algorithm requires lots of training data to establish the suitable networks, which is proper for characterizing universal seismic data, we first simulate a set of synthetic data. These synthetic data are obtained using a finite-difference method based on a designed velocity model. The simulated data are divided into two parts: the training set and the test set. The training set is used to train the designed ResNets, and the test set is used to demonstrate the validity of the trained ResNets. To further assess the performance of the trained ResNets, we select several other synthetic and field data sets to demonstrate the flexibility and validity of the trained ResNets. Finally, the trained ResNets is used to provide dense data with halved trace intervals for subsequent accurate seismic data migration or more seismic data processing steps.

Training set generation.—To generate a set of synthetic data for ResNets training, a velocity model is designed, as shown in Figure 3. The receivers are equally distributed from 800 to 3848 m along the surface with 12 m as the receiver spacing, and the source is shifted forward from the location of the first receiver to the last one.

Using a finite-difference method, we obtain the simulated data. The total data set includes 255 shots, and 255 receivers per shot. There are 256 samples along the time axis with a 4 ms time interval. The simulated data form a 3D seismic cube. To see the details of the simulated data, three slices from different perspectives are shown in

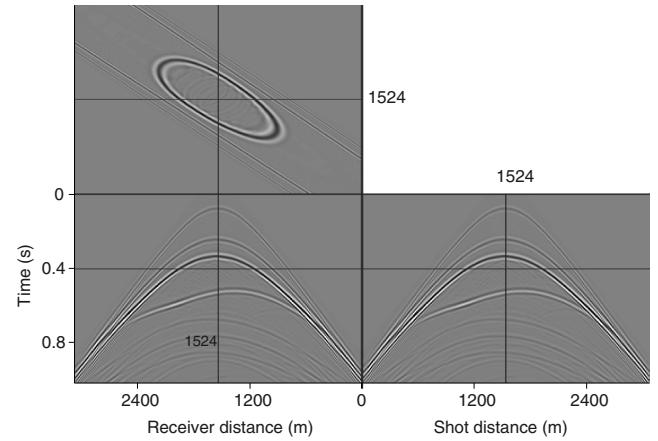


Figure 4. Simulated data view from different perspectives.

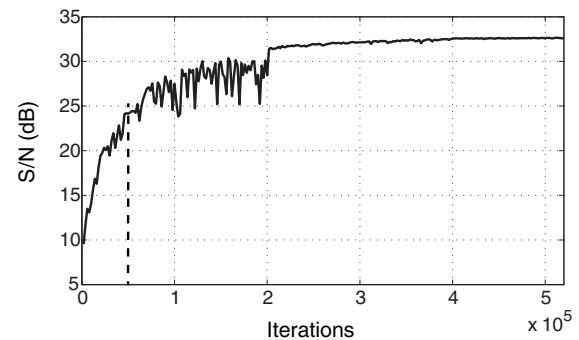


Figure 5. The recovered S/N using the ResNets during training.

Figure 4, including a common-shot gather in the bottom-left corner, a common receiver gather in the bottom-right corner, and a time slice in the top-left corner.

ResNets training.—Among 255 simulated shot gathers, we randomly select 230 shot gathers as the training set, with the remaining shot gathers regarded as the test set. During the training process, the initially interpolated data \mathbf{d}_0 are obtained using the bicubic interpolation method for the regularly sampled data \mathbf{d}_{obs} . Then, \mathbf{d}_0 is regarded as the input and the true complete data \mathbf{d}_c is regarded as a corresponding label, as shown in Figure 2. The designed ResNets is trained on the Caffe platform (Jia et al., 2014) using the parameters listed in the “Detailed training analysis” section. To monitor the training process, we store the temporary parameters of the ResNets and assess its performance using the recovered signal-to-noise ratio (S/N), which is characterized by

$$\text{S/N(dB)} = 20 \log_{10} \frac{\|\mathbf{d}_c\|_2}{\|\mathbf{d}_c - \mathbf{d}_{\text{int}}\|_2}, \quad (5)$$

where \mathbf{d}_c is the labeled data and \mathbf{d}_{int} denotes the interpolated data. The recovered average S/N using the test set is shown in Figure 5, which tells that the performance of the designed ResNets increases with iterations, especially in the early stage. The recovered S/N can achieve more than 20 dB at 50,000 iterations (marked by the dashed vertical line). It also shows that the performance increases just a little bit after 400,000 iterations; thus, the maximum iteration is set as 520,000 by trial and error. After training, the established ResNets can be used for testing and for regularly sampled seismic data interpolation.

ResNets testing.—After we obtain the final ResNets θ_{out} , the test set is used to quantitatively assess the performance via the recovered S/Ns. The results are listed in Table 1. For a better comparison, the results from the f - x prediction filter-based interpolation method (Spitz, 1991; Naghizadeh and Sacchi, 2009) are also listed in Table 1. Because the f - x prediction interpolation method is suitable for linear events theoretically, we divide the whole data into small patches with 50 sampling points along the temporal axis and 20 sampling points along the spatial axis. The highest-frequency component to be processed is 80 Hz, and the length of the prediction filter is set as five, which is determined by trial and error. Though the parameter may be not the optimal one, it can illustrate the shortcomings of the f - x interpolation method to some extent. The recovered average S/N is 17.0 and 32.87 dB for the f - x prediction interpolation algorithm, and the DL-based interpolation method, respectively. By comparison, the DL-based interpolation achieves better interpolation performance

with the minimum and maximum recovered S/Ns as 30.1 and 34.3 dB, respectively (see the bold text in Table 1). The recovered average S/N 32.87 dB is also consistent with the training performance shown in Figure 5.

To show the interpolation performance qualitatively, the results from the 12th shot gather are provided. Figure 6a and 6b shows the complete data and the regularly sampled seismic data with 50%

Table 1. Recovered S/N (dB) for the remainder shots in the test set.

Index	1	2	3	4	5	6	7	8	9	10	11	12	13
ResNets	30.1	31.6	32.0	32.0	32.2	33.3	33.1	33.9	33.8	34.2	34.0	34.1	33.8
f - x	15.6	16.1	17.4	17.4	17.6	16.5	16.9	17.3	17.3	17.8	16.1	17.6	16.0
Index	14	15	16	17	18	19	20	21	22	23	24	25	Mean
ResNets	33.7	33.8	33.6	34.3	33.8	32.8	32.8	32.7	32.1	31.9	31.3	31.0	32.87
f - x	15.7	16.6	16.6	17.8	16.9	19.6	18.6	18.0	18.3	16.0	15.9	16.2	17.0

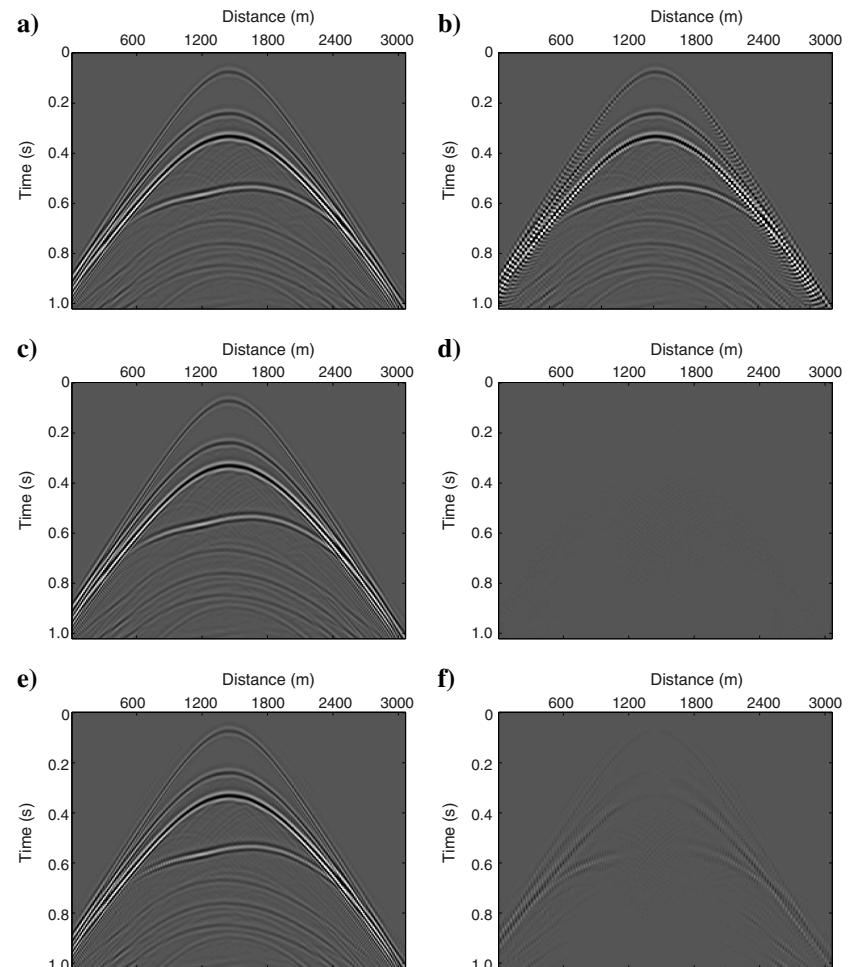


Figure 6. The 12th shot for testing. (a) Complete data, (b) regularly sampled data with 50% missing, i.e., from 255 traces into 128 traces, (c) interpolated data using the trained ResNets, (d) residual of (c), (e) interpolated data using the f - x prediction interpolation algorithm, and (f) residual of (e).

traces missing. Because of the regularly missing traces, the distance between adjacent receivers is doubled, which generates a visual serration effect. Figure 6c and 6d represents the interpolated seismic data using the trained ResNets and its residual. The result shows that the interpolated data are consistent with the true complete data (Figure 6a) with high accuracy and the residual is small. Figure 6e and 6f represents the results from the f - x prediction interpolation algorithm. The result shows that the continuity of the recovered data is enhanced; however, the residual appears to be of comparable amplitude with the signal. The recovered S/Ns are 34.1 and 17.6 dB for the DL-based and the f - x prediction-based interpolation methods, respectively. The performance of the f - x prediction interpolation algorithm is a little weak for this simulated data with curved events because it is suitable for linear events theoretically. Even though using a local window strategy can extend it to curved events cases, the optimal parameter selection is always unreachable.

For a better comparison, Figure 7 shows the f - k spectra of the 12th shot, in which the vertical axis denotes the frequency, and the horizontal axis represents the normalized wavenumber. Figure 7a and 7b represents the f - k spectra of the complete data and the regularly sampled seismic data, respectively. From Figure 7b, we see that there is some spatial aliasing because regularly missing traces make the distance between adjacent traces doubled and the Nyquist wavenumber halved. Figure 7c and 7d represents the f - k spectra of the interpolated seismic data using the trained ResNets and the f - x prediction interpolation method, respectively. Figure 7c is consistent with Figure 7a, which further illustrates the validity of the trained ResNets in regular missing

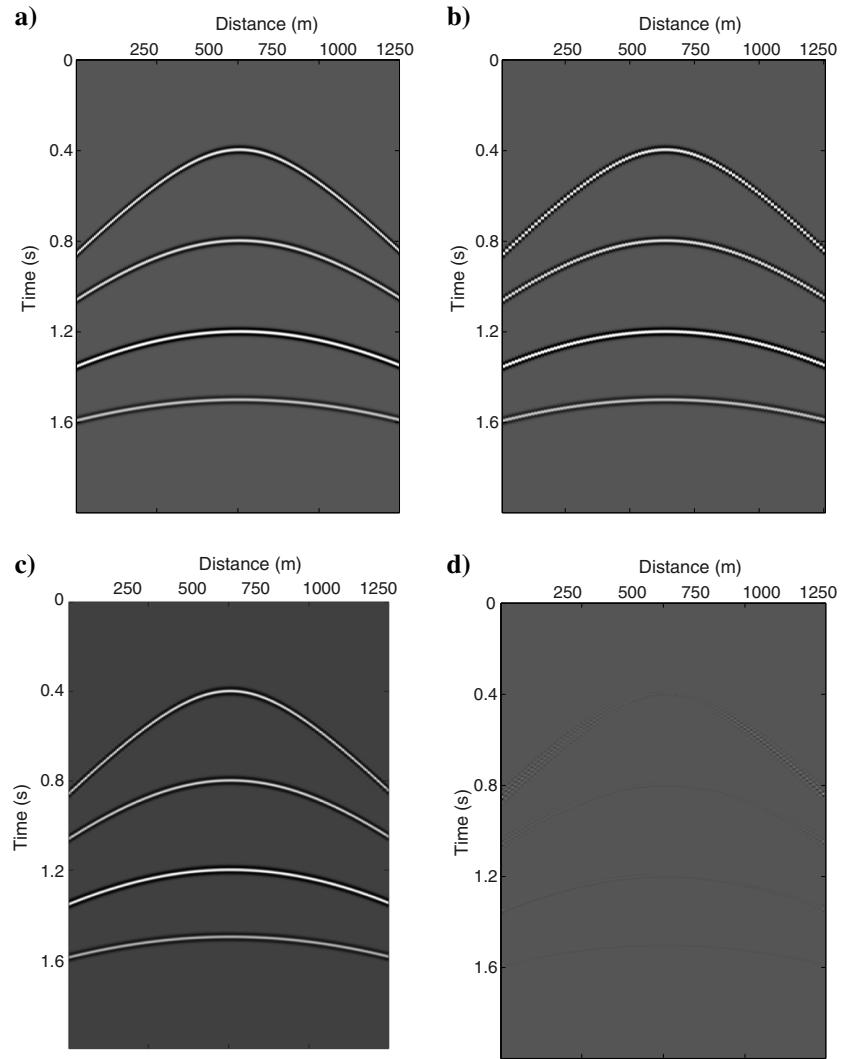


Figure 8. The five-layer model for testing. (a) Complete data, (b) regularly sampled data with 50% missing with 25 m in the trace interval, (c) interpolated data using the trained ResNets, and (d) the residual.

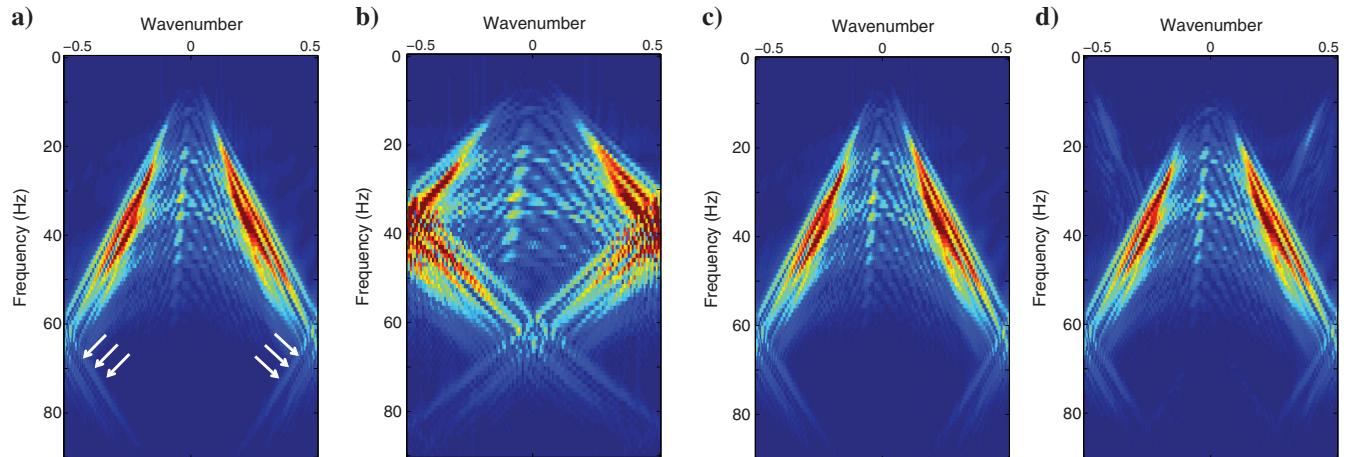


Figure 7. The f - k spectra comparisons of the 12th shot. (a) Complete data, (b) regularly sampled data with 50% missing, i.e., from 255 traces into 128 traces, (c) interpolated data using the trained ResNets, and (d) interpolated data using the f - x prediction interpolation method.

cases. In Figure 7d, some artifacts exist because the regularly missing is not properly recovered using the f - x prediction interpolation algorithm. Interpolation results using the test set demonstrate the validity of the established ResNets.

To further prove the flexibility of the trained ResNets, other synthetic data and field data with a few different features from the training set are used. Qualitative and quantitative analysis are provided. Figure 8a shows synthetic data with four events, which includes 199 traces with a 12.5 m trace interval. There are 1001 time samples per trace with 2 ms as the time interval. Figure 8b shows the regularly sampled seismic data with 25 m as the trace interval, in which visual serration effects appear. The interpolated result using the trained ResNets is shown in Figure 8c, which is consistent with Figure 8a, and it proves the effectiveness of the trained ResNets. Figure 8d shows the residual, from which we can see that some biases exist as the offset increases in the shallow part because the features of the used test data are different from those of the training data. Fortunately, the bias is minor with the recovered S/N equal to 24.54 dB. After interpolation, the trace interval is halved and the visual serration effects along the spatial direction are weakened. To further assess its performance, the f - k spectra are provided in Figure 9a–9c, which represent the spectrum of the true complete data, regularly sampled data with doubled trace interval, and interpolated data using the trained ResNets, respectively. Figure 9b shows some spatial aliasing effects for the regularly sampled data (marked by the arrows). After interpolation using the trained ResNets, the spatial aliasing disappears and the obtained f - k spectrum is consistent with the true one, which validates the proposed method.

A field data set is provided to further assess the flexibility of the trained ResNets. Figure 10a shows one field shot gather obtained by towed streamers, which includes 179 traces and 1500 sampling points per trace. The trace interval and the time-sampling interval are 12.5 m and 4 ms, respectively. Figure 10b represents the regularly sampled data with 25 m as the trace interval, which causes visual serration effects along the spatial axis. Obviously, the reason to carry out infilling would be that some other processing step requires halved trace intervals, and then the trained ResNets is used to interpolate the regularly sampled data. The interpolated result is shown in Figure 10c, which is consistent with the true complete data (Figure 10a). The recovered S/N is 19.24 dB, which proves the validity of the trained ResNets. Figure 10d shows the interpolated residual, and there is some signal leakage because of the existing different features

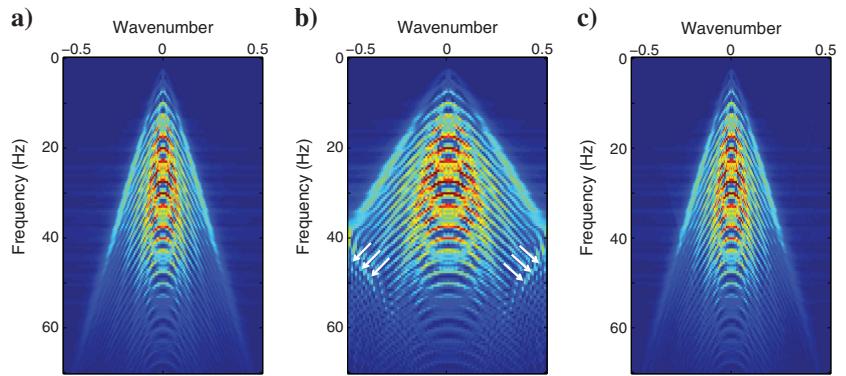


Figure 9. The f - k spectra comparisons of the layered model. (a) Complete data, (b) regularly sampled data with 25 m the trace interval, and (c) interpolated data using the trained ResNets.

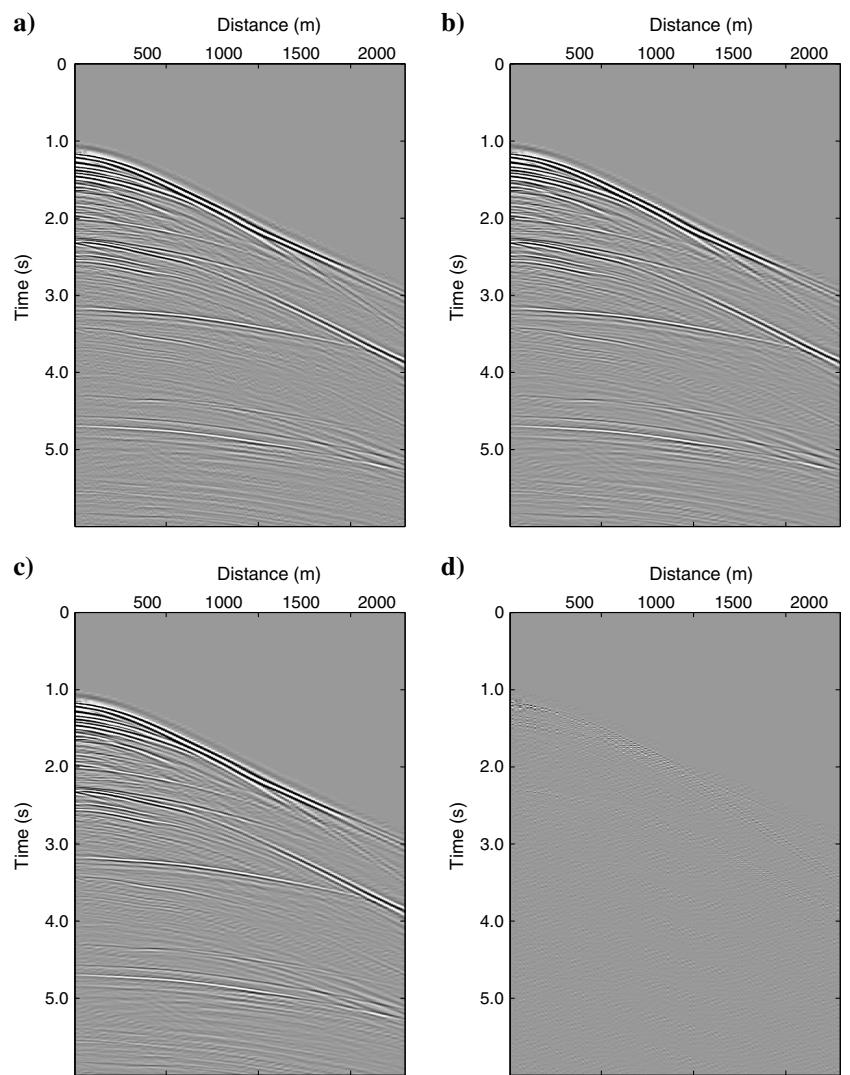


Figure 10. Field data with towed streamers for testing. (a) Complete data, (b) regularly sampled data with 50% missing, (c) interpolated data using the trained ResNets, and (d) the residual.

between this data and the trained data set. The f - k spectra are provided in Figure 11a–11c. The trace interval is doubled for the regularly sampled data, resulting in a halved Nyquist wavenumber, which generates some spatial aliasing effects (marked by the arrows). After interpolation, the spatial aliasing disappears and the obtained f - k spectrum (Figure 11c) is consistent with the true one. All of these tests prove the flexibility and validity of the trained ResNets.

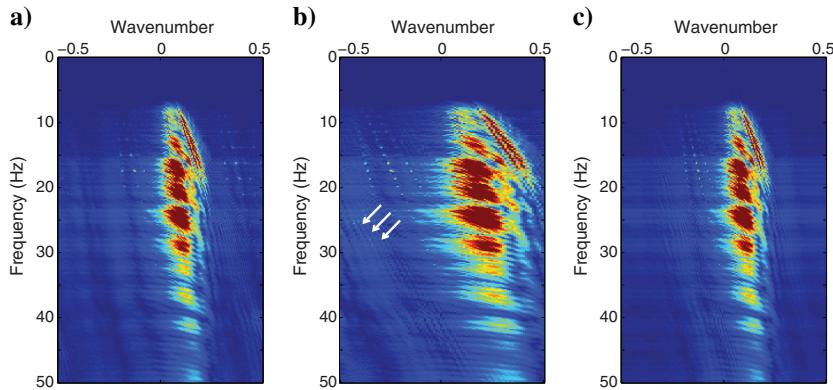


Figure 11. The f - k spectra comparisons of field data. (a) Complete data, (b) regularly sampled data, and (c) interpolated data using the trained ResNets.

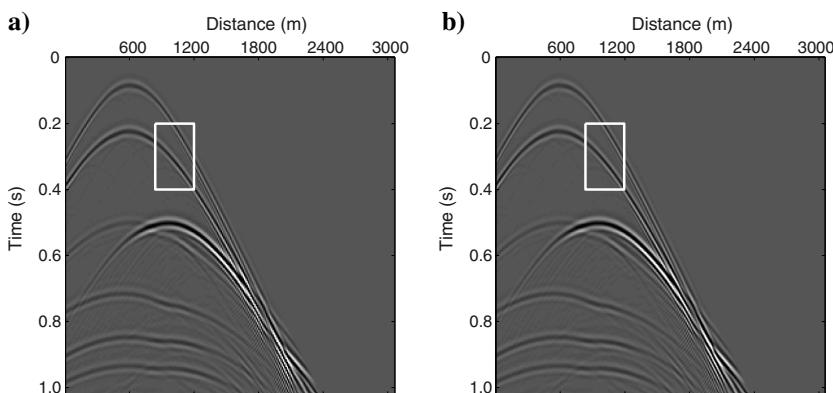


Figure 12. Tests on synthetic data to provide dense data with halved trace intervals, i.e., 6 m. (a) The fifth shot in the test set and (b) interpolated data with halved trace intervals using the trained ResNets.

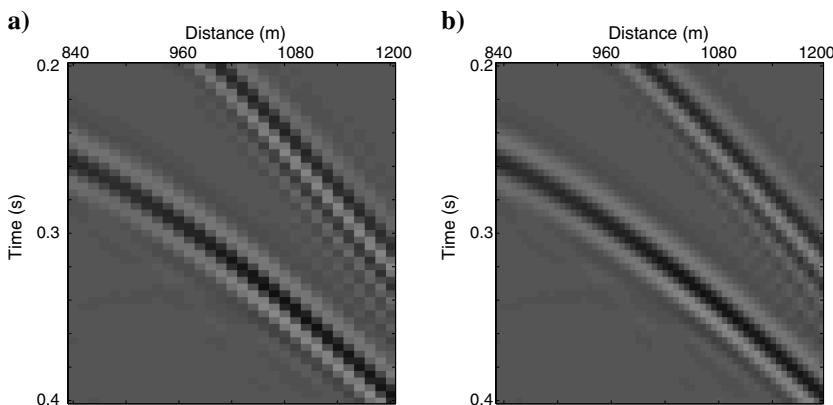


Figure 13. Magnification of marked rectangle in Figure 12. (a) From Figure 12a and (b) From Figure 12b.

It should be noted that the recovered S/N decreases as the feature difference between the test data and the training data increases. To improve the interpolation performance, different kinds of data should be included into the training set, which remains for future work. The above numerical tests demonstrate the validity and flexibility of the trained ResNets. The trained ResNets is used to provide dense data with halved trace intervals for subsequent seismic data processing, such as wave-equation-based inversion and migration.

Dense data reconstruction.—Figure 12a shows the fifth shot in the test set, which has 255 traces with a 12 m trace interval. Sometimes the acquisition trace interval is not sufficient for fine reservoir characterization and specific processing algorithms; thus, interpolation algorithms should be adopted. Using the trained ResNets, we can obtain interpolated data with 509 traces and halved trace intervals, i.e., 6 m between adjacent traces, as shown in Figure 12b. After interpolation, the spatial visual-serration effects are weakened, which can be seen clearly from the enlarged rectangle area in Figure 13. This demonstrates the validity of the trained ResNets in seismic interpolation, and the reconstructed dense data have the potential to improve the accuracy of migration, multiple attenuation, etc.

The f - k spectra are also provided for seismic data before and after interpolation in Figure 14. Because the initial trace interval is coarse, there are some spatial aliasing effects, marked by the arrows in Figure 14a. After interpolation using the trained ResNets, the trace interval is halved and the spatial aliasing disappears, which can improve the accuracy of migration and provide regularized data for SRME (Verschuur and Berkhouwt, 1997) and deghosting (Wang et al., 2017).

Field marine data are also provided to demonstrate the validity of the trained ResNets in providing dense seismic data with halved trace intervals. Figure 15a shows a marine shot gather with 196 traces, and the spatial interval is 6.25 m between adjacent traces. Then, a dense data are provided using the trained ResNets with a halved trace interval, i.e., 3.125 m. Figure 15b shows the reconstructed dense data, which are more continuous with less serration along the spatial axis. We assess the reconstruction performance by analyzing the rectangle area marked in Figure 15. Figure 16a and 16b represents the magnified version before and after interpolation, respectively. We can see that the visual spatial serration effects are weakened effectively after interpolation, as shown in Figure 16b. Figure 17 shows the spectra of seismic data before and after interpolation. Spatial aliasing appears in Figure 17a because the original spatial interval is not fine enough. After interpolation, the spatial aliasing effects disappear in the reconstructed dense data, as shown in Figure 17b. All of the tests demon-

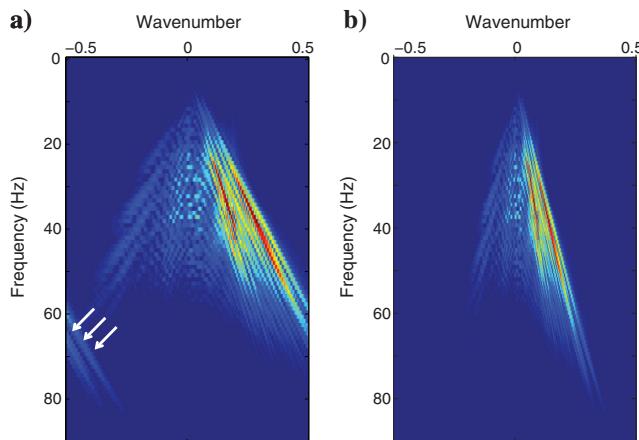


Figure 14. The f - k -spectra comparisons. (a) The f - k spectrum of Figure 12a and (b) the f - k spectrum of Figure 12b.

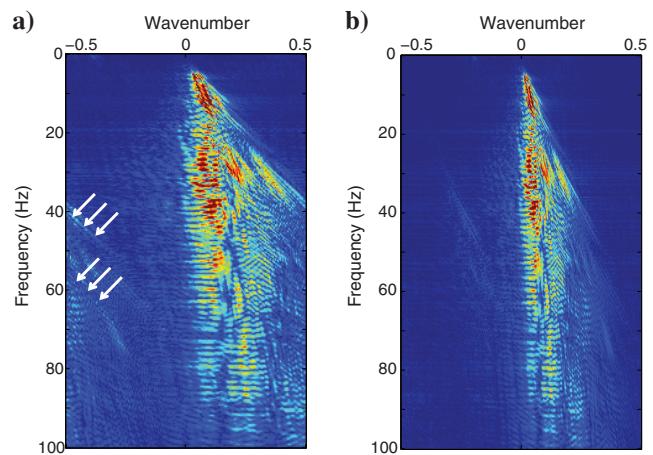


Figure 17. The f - k spectra comparisons. (a) The f - k spectrum of Figure 15a and (b) The f - k spectrum of Figure 15b.

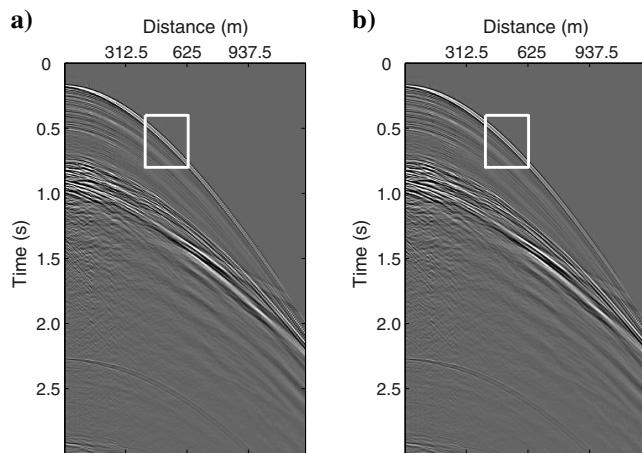


Figure 15. Tests on field marine data to provide dense data with halved trace intervals, i.e., 3.125 m. (a) Observed data with 6.25 m the trace interval and (b) reconstructed dense data using the trained ResNets with 3.125 m the trace interval.

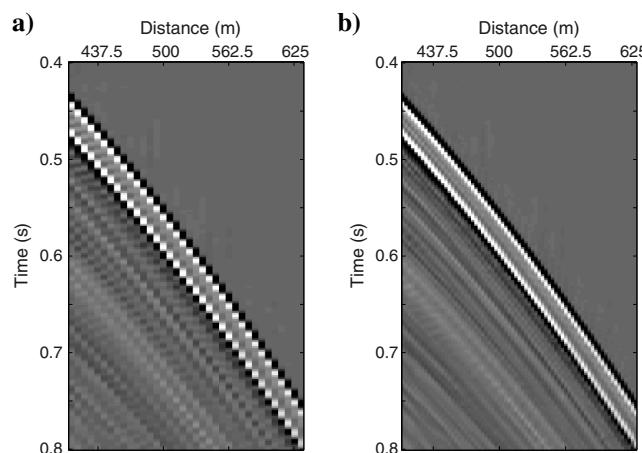


Figure 16. Magnification of marked rectangle in Figure 15. (a) From Figure 16a and (b) from Figure 16b.

strate the effectiveness and validity of the trained ResNets in providing dense data with halved trace intervals. The reconstructed aliasing-free dense data can be beneficial to subsequent seismic data processing steps.

The interpolation performance of synthetic and field data demonstrated the validity and accuracy of the trained ResNets. However, it should be noted that the bias appears when the feature in the test data is different from that in the training data. The bias may increase as the differences increase. Therefore, different kinds of data should be included into the training set to improve the performance of the established ResNets.

CONCLUSION

Seismic data interpolation is achieved for regularly sampled seismic data based on DL strategy using the designed residual learning networks (ResNets). We first designed the ResNets with eight convolutional layers based on the CNN, and we analyzed its implementation details. Then, a set of synthetic data was simulated using a finite-difference method based on a designed velocity model to train the designed ResNets. After training, synthetic data in the test set with similar features to the training set, other synthetic data and field data with different features were used to assess the performance of the trained ResNets. The interpolated results and the recovered S/Ns demonstrated the validity and flexibility of the trained ResNets. Finally, the trained ResNets were used to provide dense data with halved trace intervals for synthetic and field data. The reconstructed dense data have more continuity and less visual serration effects along the spatial axis. Besides, the spatial aliasing effects disappear in the f - k domain after interpolation. The reconstructed dense data can be beneficial in improving the accuracy of subsequent algorithms, such as inversion, migration, and multiple attenuation.

However, the interpolation bias increases when the feature difference between the test data and the training data set increases, and the noise effects are not considered. Therefore, different kinds of data should be included into the training set to improve the performance of the designed ResNets and the robustness of ResNets to noise should be assessed. Instead of ResNets, more related studies on other network architectures based on CNN and generative

adversarial nets and their applications in 5D seismic data interpolation should be carried out in future research.

ACKNOWLEDGMENTS

This research is financially supported by the National Natural Science Foundation of China (grant nos. 41604096, 41874128, 41674116), the Strategic Priority Research Program of the Chinese Academy of Sciences (grant no. XDA14010203), and the Key State Science and Technology Project (grant no. 2016ZX05024001-005). We appreciate the editors and reviewers for the critical and constructive comments, which improved the quality of the paper significantly.

DATA AND MATERIALS AVAILABILITY

Data associated with this research are confidential and cannot be released.

REFERENCES

- Abma, R., and N. Kabir, 2006, 3D interpolation of irregular data with a POCS algorithm: *Geophysics*, **71**, no. 6, E91–E97, doi: [10.1190/1.2356088](https://doi.org/10.1190/1.2356088).
- Cao, J., and B. Roy, 2017, Time-lapse reservoir property change estimation from seismic using machine learning: *The Leading Edge*, **36**, 234–238, doi: [10.1190/tle36030234.1](https://doi.org/10.1190/tle36030234.1).
- Chen, Y., D. Zhang, Z. Jin, X. Chen, S. Zu, W. Huang, and S. Gan, 2016, Simultaneous denoising and reconstruction of 5D seismic data via damped rank-reduction method: *Geophysical Journal International*, **206**, 1695–1717, doi: [10.1093/gji/ggw230](https://doi.org/10.1093/gji/ggw230).
- Dong, C., C. C. Loy, K. He, and X. Tang, 2014, Learning a deep convolutional network for image super-resolution: *European Conference on Computer Vision*, 184–199.
- Duijndam, A. J. W., and M. A. Schoneville, 1999, Nonuniform fast Fourier transform: *Geophysics*, **64**, 539–551, doi: [10.1190/1.1444560](https://doi.org/10.1190/1.1444560).
- Fomel, S., 2003, Seismic reflection data interpolation with differential offset and shot continuation: *Geophysics*, **68**, 733–744, doi: [10.1190/1.1567243](https://doi.org/10.1190/1.1567243).
- Gan, S., S. Wang, Y. Chen, Y. Zhang, and Z. Jin, 2015, Dealiased seismic data interpolation using seislet transform with low-frequency constraint: *IEEE Geoscience and Remote Sensing Letters*, **12**, 2150–2154, doi: [10.1109/LGRS.2015.2453119](https://doi.org/10.1109/LGRS.2015.2453119).
- Gao, J., J. Cheng, and M. D. Sacchi, 2017, Five-dimensional seismic reconstruction using parallel square matrix factorization: *IEEE Transactions on Geoscience and Remote Sensing*, **55**, 2124–2135, doi: [10.1109/TGRS.2016.2636864](https://doi.org/10.1109/TGRS.2016.2636864).
- Gao, J., M. Sacchi, and X. Chen, 2013, A fast reduced-rank interpolation method for prestack seismic volumes that depend on four spatial dimensions: *Geophysics*, **78**, no. 1, V21–V30, doi: [10.1190/geo2012-0038.1](https://doi.org/10.1190/geo2012-0038.1).
- Gao, J., A. Stanton, M. Naghizadeh, M. D. Sacchi, and X. Chen, 2012, Convergence improvement and noise attenuation considerations for beyond alias projection onto convex sets reconstruction: *Geophysical Prospecting*, **61**, 138–151, doi: [10.1111/j.1365-2478.2012.01103.x](https://doi.org/10.1111/j.1365-2478.2012.01103.x).
- Gao, J., A. Stanton, and M. D. Sacchi, 2015, Parallel matrix factorization algorithm and its application to 5D seismic reconstruction and denoising: *Geophysics*, **80**, no. 6, V173–V187, doi: [10.1190/geo2014-0594.1](https://doi.org/10.1190/geo2014-0594.1).
- He, K., X. Zhang, S. Ren, and J. Sun, 2016, Deep residual learning for image recognition: *IEEE Conference on Computer Vision and Pattern Recognition*, 770–778.
- Huang, L., X. Dong, and T. E. Cleo, 2017, A scalable deep learning platform for identifying geologic features from seismic attributes: *The Leading Edge*, **36**, 249–256, doi: [10.1190/tle36030249.1](https://doi.org/10.1190/tle36030249.1).
- Jia, Y., and J. Ma, 2017, What can machine learning do for seismic data processing? An interpolation application: *Geophysics*, **82**, no. 3, V163–V177, doi: [10.1190/GEO2016-0300.1](https://doi.org/10.1190/GEO2016-0300.1).
- Jia, Y., E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, 2014, Caffe: Convolutional architecture for fast feature embedding: *22nd ACM International Conference on Multimedia*, 675–678.
- Kim, J., J. Kwon Lee, and K. Mu Lee, 2016, Accurate image super-resolution using very deep convolutional networks: *IEEE Conference on Computer Vision and Pattern Recognition*, 1646–1654.
- Kreimer, N., A. Stanton, and M. Sacchi, 2013, Tensor completion based on nuclear norm minimization for 5D seismic data reconstruction: *Geophysics*, **78**, no. 6, V273–V284, doi: [10.1190/geo2013-0022.1](https://doi.org/10.1190/geo2013-0022.1).
- Krizhevsky, A., I. Sutskever, and G. E. Hinton, 2012, Imagenet classification with deep convolutional neural networks: *Proceedings of the 25th International Conference on Advances in Neural Information Processing Systems*, 1097–1105.
- LeCun, Y., L. Bottou, Y. Bengio, and P. Haffner, 1998, Gradient-based learning applied to document recognition: *Proceedings of the IEEE*, **86**, 2278–2324, doi: [10.1109/5.726791](https://doi.org/10.1109/5.726791).
- Ledig, C., L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, and Z. Wang, 2016, Photo-realistic single image super-resolution using a generative adversarial network, *arXiv preprint arXiv:1609.04802*.
- Liu, B., and M. Sacchi, 2004, Minimum weighted norm interpolation of seismic records: *Geophysics*, **69**, 1560–1568, doi: [10.1190/1.1836829](https://doi.org/10.1190/1.1836829).
- Liu, W., S. Cao, S. Gan, Y. Chen, S. Zu, and Z. Jin, 2016, One-step slope estimation for dealiased seismic data reconstruction via iterative seislet thresholding: *IEEE Geoscience and Remote Sensing Letters*, **13**, 1462–1466, doi: [10.1109/LGRS.2016.2591939](https://doi.org/10.1109/LGRS.2016.2591939).
- Ma, J., 2013, Three-dimensional irregular seismic data reconstruction via low-rank matrix completion: *Geophysics*, **78**, no. 5, V181–V192, doi: [10.1190/GEO2012-0465.1](https://doi.org/10.1190/GEO2012-0465.1).
- Naghizadeh, M., 2015, Double-weave 3D seismic acquisition. Part II: Seismic modeling and subsurface fold analyses: *Geophysics*, **80**, no. 6, WD163–WD173, doi: [10.1190/geo2015-0162.1](https://doi.org/10.1190/geo2015-0162.1).
- Naghizadeh, M., and M. Sacchi, 2009, f-x adaptive seismic-trace interpolation: *Geophysics*, **74**, no. 1, V9–V16, doi: [10.1190/1.3008547](https://doi.org/10.1190/1.3008547).
- Naghizadeh, M., and M. D. Sacchi, 2007, Multistep autoregressive reconstruction of seismic records: *Geophysics*, **72**, no. 6, V111–V118, doi: [10.1190/1.2771685](https://doi.org/10.1190/1.2771685).
- Naghizadeh, M., and M. D. Sacchi, 2010, Beyond alias hierarchical scale curvelet interpolation of regularly and irregularly sampled seismic data: *Geophysics*, **75**, no. 6, WB189–WB202, doi: [10.1190/1.3509468](https://doi.org/10.1190/1.3509468).
- Porsani, M., 1999, Seismic trace interpolation using half-step prediction filters: *Geophysics*, **64**, 1461–1467, doi: [10.1190/1.1444650](https://doi.org/10.1190/1.1444650).
- Ronen, J., 1987, Wave-equation trace interpolation: *Geophysics*, **52**, 973–984, doi: [10.1190/1.1442366](https://doi.org/10.1190/1.1442366).
- Simonyan, K., and A. Zisserman, 2015, Very deep convolutional networks for large-scale image recognition: *International Conference on Learning Representations*, 1–14.
- Spitz, S., 1991, Seismic trace interpolation in the F-X domain: *Geophysics*, **56**, 785–794, doi: [10.1190/1.1443096](https://doi.org/10.1190/1.1443096).
- Sutskever, I., J. Martens, G. Dahl, and G. Hinton, 2013, On the importance of initialization and momentum in deep learning: *International Conference on Machine Learning*, 1139–1147.
- Szegedy, C., W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, 2015, Going deeper with convolutions: *IEEE Conference on Computer Vision and Pattern Recognition*, 1–9.
- Trad, D., 2009, Five-dimensional interpolation: Recovering from acquisition constraints: *Geophysics*, **74**, no. 6, V123–V132, doi: [10.1190/1.3245216](https://doi.org/10.1190/1.3245216).
- Trad, D. O., T. J. Ulrych, and M. D. Sacchi, 2002, Accurate interpolation with high-resolution time-variant Radon transforms: *Geophysics*, **67**, 644–656, doi: [10.1190/1.1468626](https://doi.org/10.1190/1.1468626).
- Trickett, S., L. Burroughs, A. Milton, L. Walton, and R. Dack, 2010, Rank-reduction-based trace interpolation: 80th Annual International Meeting, SEG, Expanded Abstracts, 3829–3833, doi: [10.1190/1.3513645](https://doi.org/10.1190/1.3513645).
- Verschuur, D., and A. Berkhouit, 1997, Estimation of multiple scattering by iterative inversion. Part II: Practical aspects and examples: *Geophysics*, **62**, 1596–1611, doi: [10.1190/1.1444262](https://doi.org/10.1190/1.1444262).
- Wang, B., 2016, An efficient POCS interpolation method in the frequency-space domain: *IEEE Geoscience and Remote Sensing Letters*, **13**, 1384–1387, doi: [10.1109/LGRS.2016.2589260](https://doi.org/10.1109/LGRS.2016.2589260).
- Wang, B., X. Chen, J. Li, and J. Cao, 2016, An improved weighted projection onto convex sets method for seismic data interpolation and denoising: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **9**, 228–235, doi: [10.1109/JSTARS.2015.2496374](https://doi.org/10.1109/JSTARS.2015.2496374).
- Wang, B., R. Wu, Y. Geng, and X. Chen, 2014, Dreamlet-based interpolation using POCS method: *Journal of Applied Geophysics*, **109**, 256–265, doi: [10.1016/j.jappgeo.2014.08.008](https://doi.org/10.1016/j.jappgeo.2014.08.008).
- Wang, B., R.-S. Wu, and X. Chen, 2017, Deghosting based on the transmission matrix method: *Journal of Geophysics and Engineering*, **14**, 1572–1581, doi: [10.1088/1742-2140/aa82da](https://doi.org/10.1088/1742-2140/aa82da).
- Wang, B., J. Li, and X. Chen, 2015a, A novel method for simultaneous seismic data interpolation and noise removal based on the L0 norm constraint: *Journal of Seismic Exploration*, **24**, 187–204.
- Wang, B., R.-S. Wu, X. Chen, and J. Li, 2015b, Simultaneous seismic data interpolation and denoising with a new adaptive method based on dreamlet transform: *Geophysical Journal International*, **201**, 1180–1192, doi: [10.1093/gji/ggv072](https://doi.org/10.1093/gji/ggv072).
- Xu, P., W. Lu, and B. Wang, 2017, Multi-attribute classification based on sparse autoencoder-a gas chimney detection example: 79th Annual International Conference and Exhibition, EAGE, Extended Abstracts, Tu B4 14, doi: [10.3997/2214-4609.201700920](https://doi.org/10.3997/2214-4609.201700920).