

Accelerating seismic scattered noise attenuation in offset-vector tile domain: Application of deep learning

Dawei Liu¹, Xiaokai Wang², Xiaohai Yang³, Haibo Mao³, Mauricio D. Sacchi⁴, and Wenchao Chen²

ABSTRACT

Recent years have witnessed many practical applications of supervised deep learning in seismic processing. However, a weak generalization behavior prevents widespread implementation on large-scale prestack data sets for coherent noise attenuation. This is particularly true when addressing strong near-surface scattered noise in land seismic data. To alleviate this problem, we have combined deep learning with an offset-vector tile (OVT) partitioning method to suppress strong scattered noise. With the OVT partitioning method, seismic data are spatially uniformly sampled, offering a favorable foundation for network learning. Specifically, the reflection probability distribution is more stationary than the noise distribution, making it easier for the network to learn

the reflections. Accordingly, we use the direct signal learning strategy rather than the commonly used residual learning strategy to train the network. To construct high-quality training labels, we adopt the 3D continuous wavelet transform (3D CWT), which can exploit the 3D spatial correlation in OVT gathers. General use of these labels can produce results similar to 3D CWT but is highly efficient. To further improve denoising performance, we propose a training sample construction approach that leverages middle-offset OVT volumes with varying azimuths in light of midoffset relatively high signal-to-noise ratio characteristics. The field data experiment demonstrates that our proposed method also has an excellent generalization ability. Despite only using six middle-offset gathers for training, the trained network is able to effectively process 1260 OVTs in a timely manner.

INTRODUCTION

Suppressing coherent noise, which often coexists with random noise, remains a technical issue in seismic data processing, particularly in prestack seismic data. In particular, the suppression of scattered noise is one of the most challenging tasks (Vermeer, 2008; Han et al., 2016; Miorali et al., 2018). Due to near-surface heterogeneities and discontinuities, as well as topography irregularities, surface waves are diffracted, resulting in secondary events, which can seriously interfere with seismic records. Note that the preceding scattering is not random: it is complex but coherent and deterministically reproducible (Strobbia et al., 2014). Many

prestack attributes rely heavily on high-quality prestack data, so scattered noise attenuation is crucial to seismic data processing.

The noise attenuation method has been described in numerous published studies. These denoising methods rely on prior knowledge of the physical representation of the signal or coherent noise and can be roughly divided into three groups. First, the filtering based upon the predictable property of the seismic signal is a general approach, such as f - x deconvolution (Gulunay, 1986) and t - x prediction filtering (Abma and Claerbout, 1995). However, large-amplitude prestack noise distorts reflection continuity, hence decreasing the denoising capability. Second, two fundamental approaches based on the low-rank prior are currently being adopted. One is low-rank factorization

First presented at the First International Meeting for Applied Geoscience & Energy, SEG, Expanded Abstracts. Manuscript received by the Editor 4 October 2021; revised manuscript received 3 May 2022; published ahead of production 30 June 2022; published online 10 August 2022.

¹Xi'an Jiaotong University, School of Information and Communications Engineering, Xi'an, China and University of Alberta, Department of Physics, Edmonton, Alberta, Canada. E-mail: 409791715@qq.com.

²Xi'an Jiaotong University, School of Information and Communications Engineering, Xi'an, China. E-mail: xkwang@xjtu.edu.cn; wencchen@xjtu.edu.cn (corresponding author).

³Xinjiang Oilfield Company Ltd., Institute of Geophysics, Urumqi, China. E-mail: yxhai@petrochina.com.cn; maohaibo@petrochina.com.cn.

⁴University of Alberta, Department of Physics, Edmonton, Alberta, Canada. E-mail: msacchi@ualberta.ca.

© 2022 Society of Exploration Geophysicists. All rights reserved.

(Trickett, 2003; Bekara and Van der Baan, 2007; Orosez and Sacchi, 2011; Gao et al., 2015) and the other is nuclear norm minimization (Gu et al., 2014; Li et al., 2017). Typically, most of them obtain successful denoising results by assuming linear events, and the denoising results can be further improved with more flattened reflections. Third, sparse priors are widely imposed on denoising methods, requiring that the reflected or noise coefficients within the transformed image be sparse. However, the energy of the prestack noise is nonuniform, which seriously distorts the reflections in some areas. This undermines the validity of this sparse assumption, thereby weakening the denoising ability. In summary, prestack data have richer information than poststack data, making it possible to achieve good denoising results in theory. However, conventional methods have difficulty in fully exploiting such information in practice due to the preceding reasons. Therefore, selecting a suitable seismic data domain is essential for achieving more effective denoising methods.

In recent years, seismic exploration targets have gradually turned to complex structures, complex strata, and lithologic trap reservoirs (Yin et al., 2018). These exploration targets have high requirements for seismic processing accuracy. The wide-azimuth seismic acquisition can obtain massive high-quality seismic data and is commonly used to improve the imaging accuracy of seismic data in complex areas. As a core wide-azimuth seismic data processing technology, offset-vector tile (OVT) processing technology has been proposed by Vermeer (1998). Each OVT gather is composed of multiple common-midpoint traces with a limited range of offsets and azimuths, thus having a good energy consistency. In addition, OVT can enhance the spatial continuity of the prestack wavefield, which is favorable to noise attenuation. As a result, noise attenuation in the OVT domain has captured considerable attention. Li et al. (2015) improve the signal-to-noise ratio (S/N) of OVT-domain seismic data with the volume τ - p transform. Vinje et al. (2015) propose an approach based on the OVT gathers to reduce acquisition footprint artifacts caused by irregular subsurface illumination. Duan et al. (2016) reveal that the footprint can be effectively suppressed from the source by integrating 5D interpolation and migration in the OVT domain. Ling and Hu (2019) propose an internal multiple attenuation method after binning data into the OVT domain and demonstrate that such OVT gathers are appropriate for this task. Sun et al. (2019) apply a 2D curvelet transform to denoise weak signals in the OVT domain. To improve conventional denoising approaches, Li et al. (2019) use a unified learning-based framework to extract residual weak signals from strong noise after sorting data into the OVT domain. In conclusion, the OVT technique delivers an effective and accurate data domain for performing seismic processing. Although the OVT technique has the potential to improve prestack denoising performance, it also faces two main issues. First, each OVT gather can be regarded as low-fold-coverage seismic data. Therefore, the S/N of OVT gathers is usually relatively low, thus requiring a more refined denoising process. Second, containing 3D spatial coordinates as well as rich azimuth and offset information, OVT gathers have a vast amount of data, which further put forward higher requirements for the processing speed of the denoising algorithm. Therefore, it is vital to propose a robust and efficient denoising algorithm for OVT gathers.

Recent years have seen considerable advancements in deep learning techniques, particularly in the field of image processing. To date, numerous deep learning algorithms have been applied to seismic signal processing and have yielded good results (Yu and Ma, 2021). Deep learning methods are mainly composed of supervised

methods and unsupervised methods. Unsupervised methods eliminate the necessity for massive and laborious labels, and thus attract considerable attention (Saad and Chen, 2021). For instance, Pham and Li (2022) train a frequency threshold classifier in a supervised manner to assist in the unsupervised separation of coherent ground roll. Nevertheless, strong scattered noise has a considerable frequency overlap with useful signals, making it challenging to add a frequency constraint for strong scattered noise attenuation. Although some other unsupervised deep learning methods have shown promise in weak scattered noise areas (Liu et al., 2020a), they need to optimize the network for each input gather and cannot use the well-trained network to test other data directly. In addition, another tricky aspect of unsupervised learning is the parameter selection. Therefore, our paper focuses on supervised deep learning because it is computationally efficient and well established.

The denoising convolutional neural network (DnCNN) (Zhang et al., 2017), U-net (Ronneberger et al., 2015), and generative adversarial network (Goodfellow et al., 2014; Kaur et al., 2019, 2020) are the most commonly used network architectures. Among them, due to the straightforward network architecture and effective residual learning strategy, DnCNN has received the most attention and has been used to attenuate random noise (Zhang et al., 2018), multiples (Yu et al., 2019), ground roll (Li et al., 2018), desert noise (Zhao et al., 2019), linear noise (Zheng et al., 2020), and blended noise (Matharu et al., 2020). Although DnCNN has succeeded in suppressing prestack coherent noise, there are still many remaining challenges compared with poststack noise, limiting its widespread application in practice. In general, the successful prediction of deep learning is dependent on the essential assumption that test and training data sets originate from the same distribution. Due to the similarity of the underlying structure, poststack data from the same work area have relatively similar data distribution characteristics, thus roughly satisfying this assumption. Moreover, owing to the strong noise suppression ability of the stacking, data usually possess a high S/N, low-noise energy, and weak amplitude variation, which provides a good data basis for network training. Therefore, applying DnCNN to poststack data does not present a significant challenge. In contrast, prestack noise has varying signatures, which are quite distinct from each other. Moreover, many types of noise have strong amplitudes and are unevenly distributed, causing trouble when dividing the training data into patches. Another key factor contributing to the unstable network training is the severe amplitude decay and considerable waveform changes of reflections. Taking near-surface scattering noise in land acquisition as an example, it often has high amplitudes and exhibits complex diffraction patterns associated with the spatial positions of point scatterers, obscuring prime reflections in t - x and f - k domains. Known as one of the most troublesome forms of coherent noise, it presents a severe challenge to the conventional attenuation strategies performed in the common-shot and common-receiver domains. Not surprisingly, supervised deep learning methods, such as DnCNN, have failed to make a breakthrough in it. On the one hand, there is a lack of high-quality training samples for network training. On the other hand, DnCNN has difficulties in capturing the desired features because the noise amplitude at near, middle, and far offsets is nonuniform. As a result, the well-trained DnCNN has a limited generalization. Unlike poststack data, even within the same work area, the network struggles to provide satisfactory denoising results for all data. Therefore, it is essential for prestack coherent noise attenuation that the network

is trained on the appropriate domain where it can easily learn the features in an accurate manner.

We present a novel method for attenuating prestack scattered noise by integrating deep learning and the OVT partitioning technique. This integration can make full use of the advantages of both sides. From the OVT aspect, deep learning is well suited for high computational efficiency thanks to access to parallel computing. Benefiting from the rapid development of graphics processing unit (GPU) technology, deep learning can substantially reduce the processing time of OVT gathers to an acceptable level. As a result, deep learning makes it no longer challenging to process large-scale OVT gathers in practice, so that OVT techniques more applicable for wide-azimuth exploration. From the network aspect, seismic data in the OVT domain have a smooth and continuous wavefield. Moreover, each OVT gather is single-fold coverage of the whole acquisition area, and it samples the complete 3D subsurface. Therefore, the spatial and temporal amplitude variation in the OVT domain is reduced and uniform regardless of near-, middle-, and far-offset volumes. Thus, these merits of OVT data provide an ideal data foundation to train the network and further improve the network generalization ability. Similar to 3D-DnCNN (Liu et al., 2020b) which was applied to poststack data, we adopt a 3D network to exploit the 3D spatial correlation of the OVT volumes. On this basis, we have made some modifications so that it is more suitable for our task. In particular, removing the residual learning strategy in 3D-DnCNN is superior for strong noise attenuation, as demonstrated by the synthetic data. To construct high-quality training samples and fully exploit the dip information provided by the field seismic data, we adopt a 3D continuous wavelet (3D CWT) to suppress scattered noise in the OVT volumes. By randomly selecting OVT volumes with varying azimuths inside the middle-offset ranges, we are able to further increase the denoising performance in efficiency and quality compared with 3D CWT. To demonstrate this, the well-trained network is tested on an actual seismic data set from Western China. It takes roughly 6 min to effectively denoise an OVT gather with a size of $200 \times 200 \times 3001$, which is approximately one-tenth of 3D CWT.

METHOD

OVT partitioning technique and data preprocessing

Complex geologic exploration targets require high seismic exploration accuracy. Limited by the nonuniformity of the near-, medium-, and far-offset amplitude and the difficulty of actual azimuth anisotropy processing, narrow-azimuth seismic processing technology has difficulties in further improving seismic reservoir prediction accuracy. Wide-azimuth processing technology effectively solves these problems, thus becoming one of the mainstream directions of seismic exploration technology development at this stage.

OVT is one of the most widely used wide-azimuth processing technologies in the industry. At present, OVT gathers mainly come from orthogonal observation systems. Sorting seismic data into OVT gathers data is realized in the cross-spread domain. The OVT size is decided first, and then OVT gathers are extracted. As shown in Figure 1a,

a cross spread is a collection of traces that share a common source line and a common receiver line. Then, the midpoint area of the cross spread can be divided into many tiles, as shown in the yellow square. The size of each tile is decided by the receiver line interval and source line interval, usually twice these intervals. Theoretically, for the 3D orthogonal observation system, this partitioning rule can make each tile have only one coverage at each common midpoint, obtaining the best data regularity. Each tile comprises several common midpoints within a limited source and receiver range, thereby comprising particular azimuth and offset information. Then, an OVT gather can be obtained by patching together all of the tiles with the same offset ranges in different cross spreads, as depicted in Figure 1b. The trace assignment using in the traditional common-offset domain is nonuniform. Therefore, aliased noise is prevalent, especially in the inline direction, due to insufficient spatial sampling. On the contrary, the OVT partitioning method can produce uniformly sampled data and show good data consistency.

Although the adjacent traces in the OVT gathers have a reasonable correlation, many empty traces need to be interpolated. To prepare the final output of OVT gathers, we perform a 5D seismic data reconstruction based on the antileakage Fourier transform (Xu et al., 2010). In addition, normal moveout, static correction, and abnormal energy suppression also are applied to improve seismic data quality. It is worth mentioning that we do not attenuate noise during data preprocessing because denoising is our primary objective.

Training sample construction

To achieve good results with supervised deep learning, the most challenging aspect is to have high-quality training samples. Having ground-truth reflections is a significant advantage when training networks with synthetic data. However, it is an overwhelming task to simulate realistic scattered noise. Therefore, even though this training sample construction technique can achieve good results when applied to a small-scale test, the practical application of this method at a large scale is not very reliable. Due to this, we choose the second method, which is creating training labels with a traditional method.

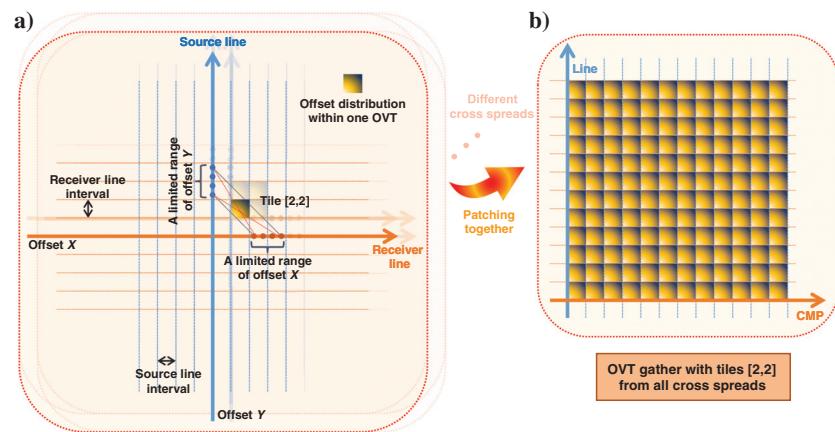


Figure 1. OVT gathers can be formed based on the cross-spread geometry. (a) A cross spread is a collection of traces that share a common source line and a common receiver line. Traces from a specified range of offset X and offset Y are grouped into an OVT. (b) Patching together all of the tiles that have the same two offset ranges from different cross spreads produces an OVT gather, which is a single-fold data set with similar offsets and azimuths that covers the whole survey region. CMP: common midpoint.

Sparsity is a widely used regularization for denoising seismic data. With sparse constraints, we can denoise the noisy data in some specific sparse domain based on the spatial continuity of the signal. Previous methods based on sparse representation fail to estimate good results when dealing with scattered noise due to the absence of a suitable sparse domain for representing useful signals. For example, in the shot domain, the spatial amplitude variation of the noise is evident, and the sparsity of effective signals is thus affected, resulting in poor denoising results. However, OVT gathers exhibit good spatial consistency and therefore provide an ideal domain for denoising. We use continuous wavelet transform (CWT) to denoise OVT gathers based on sparse regularization. To fully exploit the 3D spatial continuity of OVT volumes, 3D CWT (Wang et al., 2010; Zhao et al., 2021) is used to obtain clean data, and the definition is

$$(W_\psi f)(a, \mathbf{b}, \theta, \varphi) = \langle f, \psi_{a, \mathbf{b}, \theta, \varphi} \rangle = \frac{1}{a^3} \int_{\mathbb{R}^3} f(\mathbf{u}) \psi^* \left(\frac{1}{a} r_{-\theta, -\varphi}(\mathbf{u} - \mathbf{b}) \right) d^3 \mathbf{u}, \quad (1)$$

where $f(\mathbf{u})$ is the input 3D OVT gather. Here, $a > 0$, \mathbf{b} , and $(\theta, \varphi) \in [0, 2\pi]$ represent the scale factor, translation operation, and rotation parameter, respectively, used to generate a series of wavelets:

$$\psi_{a, \mathbf{b}, \theta, \varphi} = \frac{1}{a^3} \psi \left(\frac{1}{a} r_{-\theta, -\varphi}(\mathbf{u} - \mathbf{b}) \right), \quad \mathbf{u}, \mathbf{b} \in \mathbb{R}^3. \quad (2)$$

For our applications, we select the high-dimensional Morlet wavelet (Morlet et al., 1982; Grossmann and Morlet, 1984; Farge, 1992) as the mother wavelet $\psi(\mathbf{u})$ because it is a directional wavelet, and its formulation is similar to a Gabor transform, which is easy to implement. The high redundancy of CWT allows for greater denoising flexibility. After decomposing the OVT gathers into a 6D coefficient domain, we can filter the unwanted coefficient corresponding to the noise component with the hard threshold function. After inverse 3D CWT, we can obtain satisfactory results, which can be seen as labels. Even though these results are impressive, 3D CWT requires a substantial amount of computational time. Herein lies the motivation for our use of deep learning.

Furthermore, it is essential to consider how to choose more appropriate OVT gathers when constructing training samples because they have different azimuths and offsets. At the near offset, the amount of noise energy is too high, primarily due to the existence of coherent ground roll. At the far offset, the recording time for signals is short, and the wavefield is more complex. In addition, it is susceptible to interference from coherent noise, such as multiple refraction. Therefore, the near and far offsets are not suitable for training. In the middle-offset range, there are typically more traces than in the near-offset range and in the far-offset range, so it usually has a higher S/N, which is conducive to a better denoising result using conventional methods. Therefore, middle-offset gathers are selected to train the network. In addition, OVT gathers can conveniently provide azimuthal anisotropy. We use a randomization operator to select different azimuth gathers from middle-offset OVT gathers to use this azimuth information fully. It is inspired by the fact that simple random sampling is an unbiased approach to garner

the responses from a large set, making it easier for the network to learn the correct distribution of the signal.

Model formulation and network architecture

Network architecture can be viewed as an implicit regularization. Choosing the appropriate network architecture can make use of the preceding high-quality training samples more efficient. We model the seismic data, denoted by a vector $\mathbf{y} \in \mathcal{Y} = \mathbb{R}^m$, as a superposition of reflections and noise:

$$\mathbf{y} = \mathbf{x} + \mathbf{n}, \quad (3)$$

where $\mathbf{x} \in \mathcal{X} \subset \mathbb{R}^m$ represents useful signals and \mathbf{n} represents scattered and random noise. Under supervised deep learning, the denoising problems are viewed as regression problems. The network training process is to develop a mapping between noisy seismic data and clean reflections or noise with a large number of training sample pairs containing noisy data and clean labels,

$$(y_i, x_i) \sim (Y, X) = (X + N, X), \quad i = 1, \dots, K, \quad (4)$$

or containing noisy data and noise,

$$(y_i, n_i) \sim (Y, N) = (X + N, N), \quad i = 1, \dots, K, \quad (5)$$

where the clean labels and noise are separated by the conventional method. Here, Y , X , and N are random variables taking values in \mathcal{Y} , \mathcal{X} , and \mathcal{N} , respectively, and K is the total number of training samples. Through network training, deep learning attempts to find the regression function based on learning reflections,

$$h^* = \underset{h}{\operatorname{argmin}} \mathbb{E}_{X, N} \{ L(h(X + N), X) \}, \quad (6)$$

or learning noise,

$$h^* = \underset{h}{\operatorname{argmin}} \mathbb{E}_{X, N} \{ L(h(X + N), N) \}, \quad (7)$$

which is an expected risk minimization task. The loss function L we choose is a pixel-wise mean square error, which is the most commonly used one. Then, equation 6 can be expressed as direct signal learning,

$$L(h(X + N), X) = \|h(X + N) - X\|_2^2, \quad (8)$$

and equation 7 can be expressed as residual learning,

$$L(h(X + N), X) = \|h(X + N) - N\|_2^2. \quad (9)$$

However, either equation 8 or equation 9 is usually intractable due to the fact that the joint distribution function $P(Y, X)$ is unknown. Therefore, the empirical risk is used to estimate the expectation, which is derived from the sample mean over the training data set. The empirical risk minimization task is achieved by optimizing over the convolutional neural network (CNN) parameterized mappings $f_\theta: \mathcal{Y} \rightarrow \mathcal{X}$ or $f_\theta: \mathcal{Y} \rightarrow \mathcal{N}$ with parameters θ . As a result, the network training process equates to determining the optimal parameters

$$\theta^* = \operatorname{argmin}_{\theta} \sum_{i=1}^N \|f_{\theta}(y_i) - x_i\|_2^2 \quad (10)$$

or

$$\theta^* = \operatorname{argmin}_{\theta} \sum_{i=1}^N \|f_{\theta}(y_i) - (y_i - x_i)\|_2^2, \quad (11)$$

which minimizes the objective function by training the network with sample pairs $\{y_i, x_i\}_{i=1}^K$ or $\{y_i, n_i\}_{i=1}^K$. When the network finishes training, the network f_{θ^*} is used to recover useful signals or noise from unseen noisy data. We should mention that the distribution of training and test data should be independent and identical, which is one of the prerequisites for implementing deep learning algorithms successfully. Otherwise, the test results may be incorrect, and the data set bias leads to unreliable predictions.

When dealing with image denoising problems, it is common to introduce noise that follows the Gaussian distribution. The noise distribution is fixed, thus easier to train the network with a learning noise strategy. The widespread use of DnCNN with residual learning proves it. However, the reflection distribution and noise distribution typically vary heavily in the shot domain. In other words, learning reflections or noise in the shot domain is not an easy task. Fortunately, seismic data in the OVT domain have many sound characteristics, offering a favorable data foundation for network learning. OVTs with different offsets are single-fold coverage of the entire survey, all responding to the same underground structure. Therefore, the distribution of their reflections is similar. Moreover, the prestack wavefield of OVT gathers is consistent and continuous. There is a minor spatial and temporal amplitude variation of reflections in the OVT domain. In contrast, the noise has substantial differences at the near, middle, and far offsets. Therefore, the reflection probability distribution is more stationary than the noise distribution, making the network easier to learn the reflections.

Corresponding to the preceding analysis, we remove the residual learning structure in the 3D DnCNN. As illustrated in Figure 2, we present a network architecture that directly maps the noisy data into clean reflections through a feedforward neural network. In the input layer, there is a convolution function followed by an activation function to increase the nonlinearity of the neural network. The most commonly used activation function is the rectified linear unit (ReLU) because it accelerates the network training and, generally, produces better practice performance than any other activation function. Therefore, ReLU is set as the default activation function. The middle layers contain 15 hidden layers. To facilitate the deep-network training process, batch normalization (BN) is applied prior to each activation function, which is different from the input layer. Finally, there is a convolutional output layer. We use the same padding strategy for each convolution operation to ensure that the output is the same size as the input. It is noteworthy that all convolution operations are performed in three dimensions to maximize the use of seismic spatial structure information.

EXAMPLES

In this section, we test the proposed method by examining synthetic data with added real scattered noise and a field seismic data example. The S/N of the denoised data is calculated using the following expression:

$$S/N = 10 \log_{10} \frac{\|\mathbf{d}^{\text{true}}\|_2^2}{\|\mathbf{d}^{\text{true}} - \mathbf{d}^{\text{denoised}}\|_2^2}, \quad (12)$$

where \mathbf{d}^{true} denotes the true data prior to denoising and $\mathbf{d}^{\text{denoised}}$ denotes the recovered data after denoising.

Synthetic example

The synthetic data are used to demonstrate the rationality of not using residual learning. To construct the synthetic training data set, we follow these steps. We first construct 5D synthetic data and then sort them into the common-offset gathers. The 5D data set is parameterized by a superposition of 18 parabolas of random curvature and intercept times. The wavelet central frequency is 20 Hz with a 2 ms sampling rate. Figure 3a displays a middle-offset gather. We strive to simulate the OVT gathers by flattening most reflections. Then, we use 3D CWT to extract the noise from real OVT gathers to make the noise look realistic. Figure 3b illustrates scattered noise gathers from a middle-offset OVT gather with almost no leakage in it due to carefully selected parameters. Finally, adding them together results in noisy data in Figure 3c and the network can be trained by feeding it pair-wise clean data and noisy data.

Two different networks, i.e., with the direct signal learning in equation 8 and with the residual learning in equation 9, are trained with these training samples. Note that all of the parameters and hyperparameters are kept the same for these two networks for a fair comparison. To ensure that these two networks have learned the ability to denoise scattered noise, we apply them to denoise the noisy training data. From Figure 4, it can be seen that both networks can produce good results. Even if the results of our network with direct signal learning are slightly better, there is little difference. The same conclusion can be drawn from the S/N in Table 1.

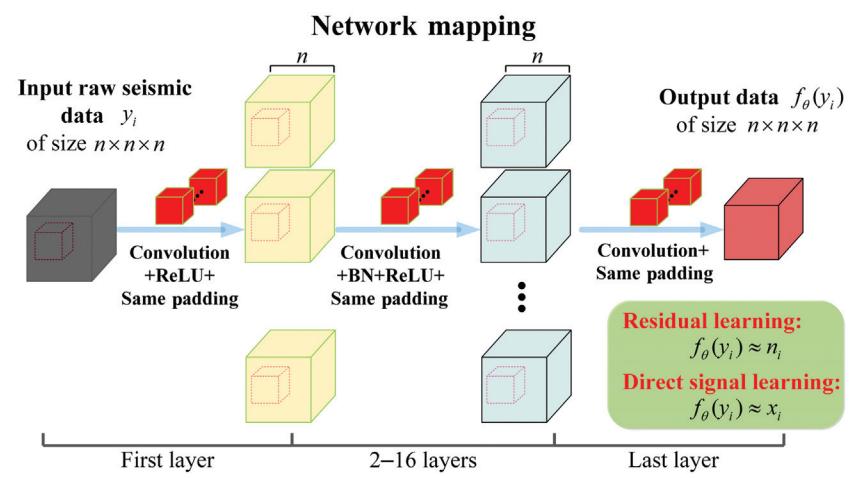


Figure 2. Network architectures with two learning strategies. Unlike the well-known residual learning strategy in 3D-DnCNN, we adopt the direct signal learning strategy to make it more adaptive to the problem of scattered noise attenuation.

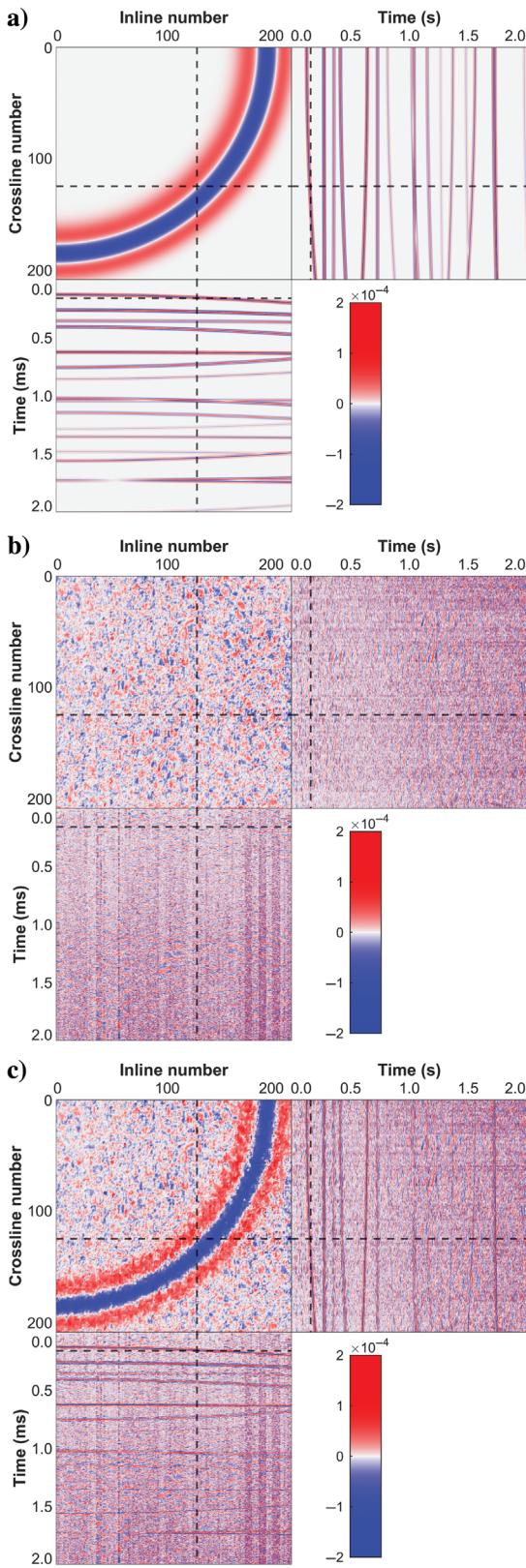


Figure 3. Synthetic common-offset training data. (a) A clean middle-offset gather from the 5D data set. (b) The scattered noise from real data. (c) The noisy data by adding scattered noise to the synthetic data.

After training, both networks can improve the S/N of noisy data from -5.7 dB to more than 12 dB. Their S/N is almost the same, indicating that the networks have converged.

The denoising performance of these two networks is evaluated using another gather from the same 5D data set. The clean data, which are treated as the ground truth, are shown in Figure 5a. Scattered noise from near-offset gather is displayed in Figure 5b. The noise has a strong energy and almost completely covers all reflections as shown in Figure 5c, which makes denoising challenging. As also can be seen in Table 1, the S/N of the noisy data is -18.8 dB, which is extremely low. Figure 6a and 6b shows the results produced by the network with residual learning. From the denoised gather, it is evident that there is some residual noise. Especially in the shallow layer, the reflections are still obscured by scattered noise. This is because we have not encountered such noise types in our training samples. Thus, it is difficult to obtain good test results with learning noise strategies. On the contrary, our network produces better results, as shown in Figure 6c and 6d, because we train the network from the perspective of signals rather than noise. Similarly, these findings also can be clearly demonstrated by the quantitative measurement results in Table 1, indicating that the S/N of our network is further improved by almost 7 dB. It is evident that a direct mapping strategy is effective when dealing with high amplitude and high variation noise. In addition, we acknowledge that the residual learning technique is an excellent tool for dealing with different sorts of noise. We also show the results with the classic f - x - y prediction filtering method (Chase, 1992) in Figures 4e, 4f, 6e, and 6f. It is worth mentioning that the scattered noise shows visible linear features, which are much more complicated to remove than random noise. In addition, the scattering has a strong energy in near-offset OVTs. Correspondingly, the conventional method fails to effectively remove it, as it only achieves a 6 dB improvement on the test data in Table 1.

Field data example

The workflow we adopt to process field seismic data is illustrated in Figure 7. As discussed in the previous section, we randomly select six medium-offset OVT gathers with varying azimuths to serve as training data. Then, an effective conventional method, 3D CWT (Zhao et al., 2021), is used to generate corresponding high-quality labels. Furthermore, we also perform sample selection and sample cleansing to double-check that the sample quality is satisfactory before feeding them for training. Once the network has been well trained, it is used to denoise other testing OVT gathers.

A 3D wide-azimuth survey of Western China has been used to demonstrate the validity of the proposed method. The offset-azimuth locations of OVT volumes are displayed in polar coordinates, as shown in Figure 8. Taking a closer look, we see that this survey contains 35 distinct offsets and 36 distinct azimuths, thus totaling 1260 OVT volumes. The overall amount of data approximately reaches 1 TB. In Figure 8, six red points indicate our training data set construction strategy that six middle-offset OVT volumes with varying azimuths are filtered by 3D CWT, which is then fed to the network for training. A baseline approach with a general use of labels is represented by six green points, which selects three near-offset OVTs and three middle-offset OVTs located in the same azimuth.

We evaluate our results on the prestack OVT gathers, stacked OVT gathers, and common-midpoint gathers sequentially. We show three different types of prestack OVT gathers: near, medium, and far

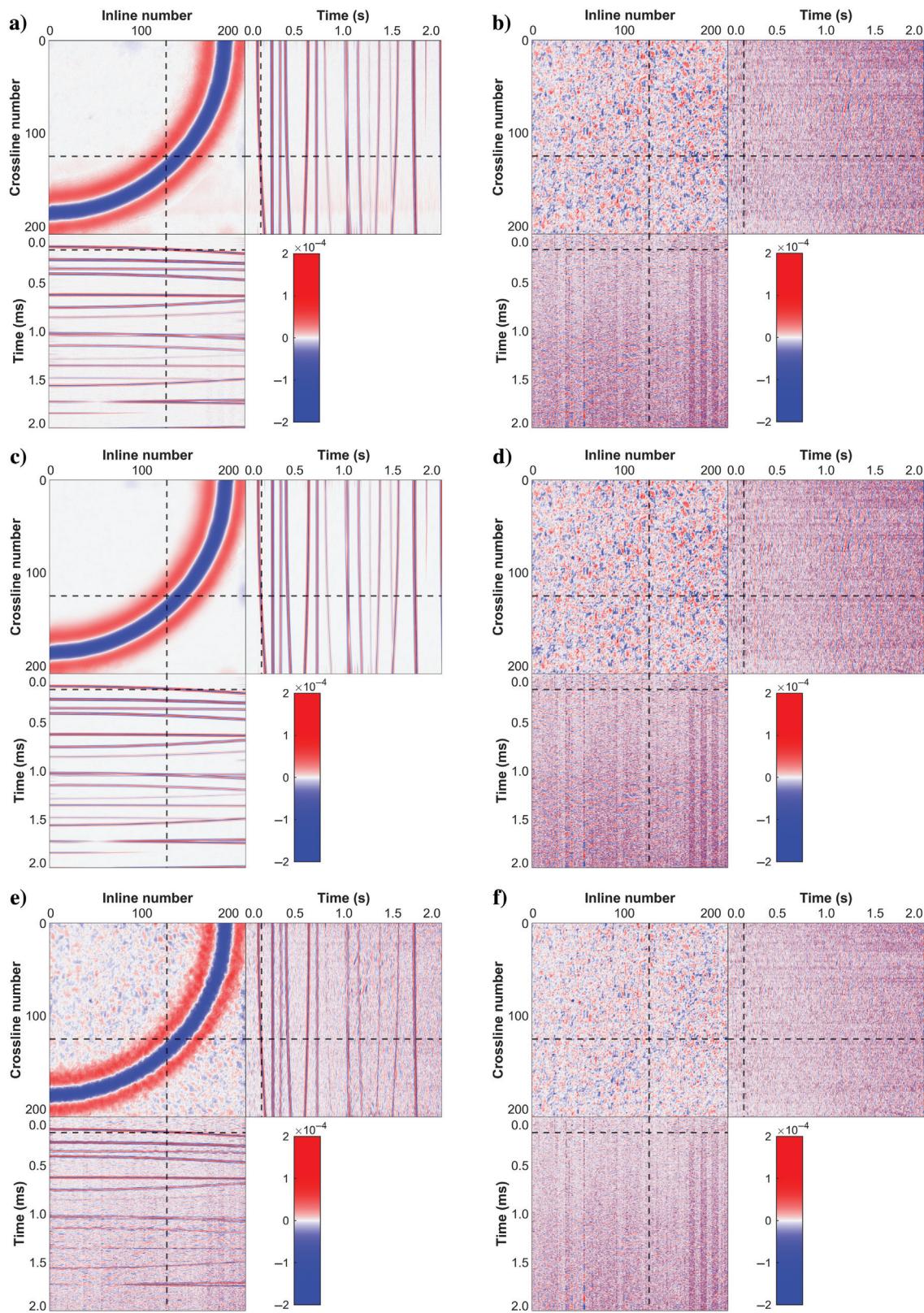


Figure 4. Results of synthetic training data. We apply the network to the training data set to check whether the network is well trained. (a) The denoised results obtained by the network with residual learning. (b) The scattered noise removed by the network with residual learning. (c) The denoised results obtained by the network with direct signal learning. (d) The scattered noise removed by the network with direct signal learning. It can be seen that their performance toward training data are good and similar, meaning that both networks are well trained. (e) The denoised results obtained by the f - x - y filtering method. (f) The scattered noise removed by the f - x - y filtering method.

offsets, whose locations are highlighted by different colored points in Figure 8. First, we compare near-offset results. Near-offset traces, which have an almost vertical incident angle, are useful for detecting azimuth variations, as well as for performing amplitude-variation-with-offset analyses. However, near-offset traces in land seismic data are severely distorted by noise, and we can hardly find the reflections. As shown in Figure 9a, near-surface weak signals are embedded in much stronger noise, which makes many conventional methods less effective at extracting those signals. Figure 9b and 9c represents the denoised results by 3D CWT. Even though the original data have a low S/N, one can still notice a significant reduction of the coherent noise in Figure 9b, as well as much clearer reflections. This proves the fact that the OVT domain is a good choice for noise reduction because there is a continuous wavefield, and 3D CWT is an effective tool for denoising the strong scattered noise in the OVT domain. In this sense, the combination of deep learning, which uses 3D CWT to build labels, and OVT techniques is promising. Moreover, the noise distribution is nonuniform, as shown in Figure 9c. Therefore, learning a mapping of noise with the network should not be the best option. Figure 9d and 9e shows the denoised data and removed noise by our proposed method. Our method uses the network to learn the reflections. The S/N after noise attenuation has been significantly improved, and no obvious damage to signals can be detected in the noise section. In addition, according to the red boxes, 3D CWT has little residual noise whereas our network can effectively remove it.

A general use of labels produces highly similar results to 3D CWT. Specific to the near-offset section shown in Figure 9, their deep-layer results are slightly different. The near-offset denoised results based on the general use of labels are shown in Figure 9f and 9g. Results of the shallow and middle layers (0–4 s) are consistent with our expectation that they perform similarly to 3D CWT. However, the deep-layer (4–6 s) results exhibit a severe signal leakage, as shown in the bottom green box in Figure 9. The possible explanation lies in the simultaneous use of near- and midoffset labels. Due to their common underground structure, near- and middle-offset deep-layer labels should be similar. However, 3D CWT performs better in middle-offset OVT volumes than in near-offset volumes. Accordingly, we sometimes train the network with two different labels for similar input noisy data, making the network ambiguous. Therefore, the network outputs approximately zero values in certain regions, and leakage of useful signals occurs. In contrast, the proposed method uses all middle-offset labels, which means that all labels have high consistency, resulting in more continuous denoised results as displayed in the top green box in Figure 9. From a closer inspection of the results obtained by residual learning in Figure 9h and 9i, we can draw the same conclusion as

Table 1. S/N comparison of different network architectures for synthetic data.

	Training data (dB)	Test data (dB)	Test data gain (dB)
Original noisy data	-5.7	-18.8	—
With residual learning	12.4	-6.0	12.8
With direct signal learning	13.2	1.5	20.3
f-x-y filtering method	2.7	-12.8	6.0

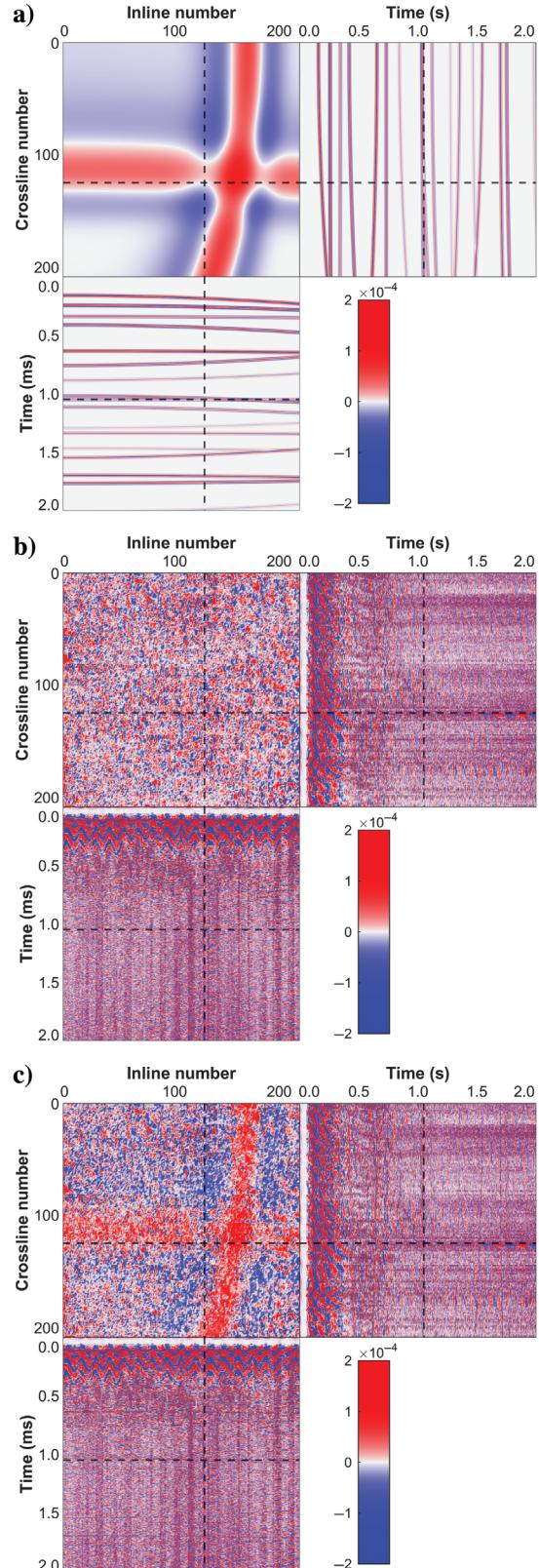


Figure 5. Synthetic common-offset test data. (a) A clean near-offset gather from the 5D data set. **(b)** The scattered noise from real data. **(c)** The noisy data by adding scattered noise to the synthetic data.

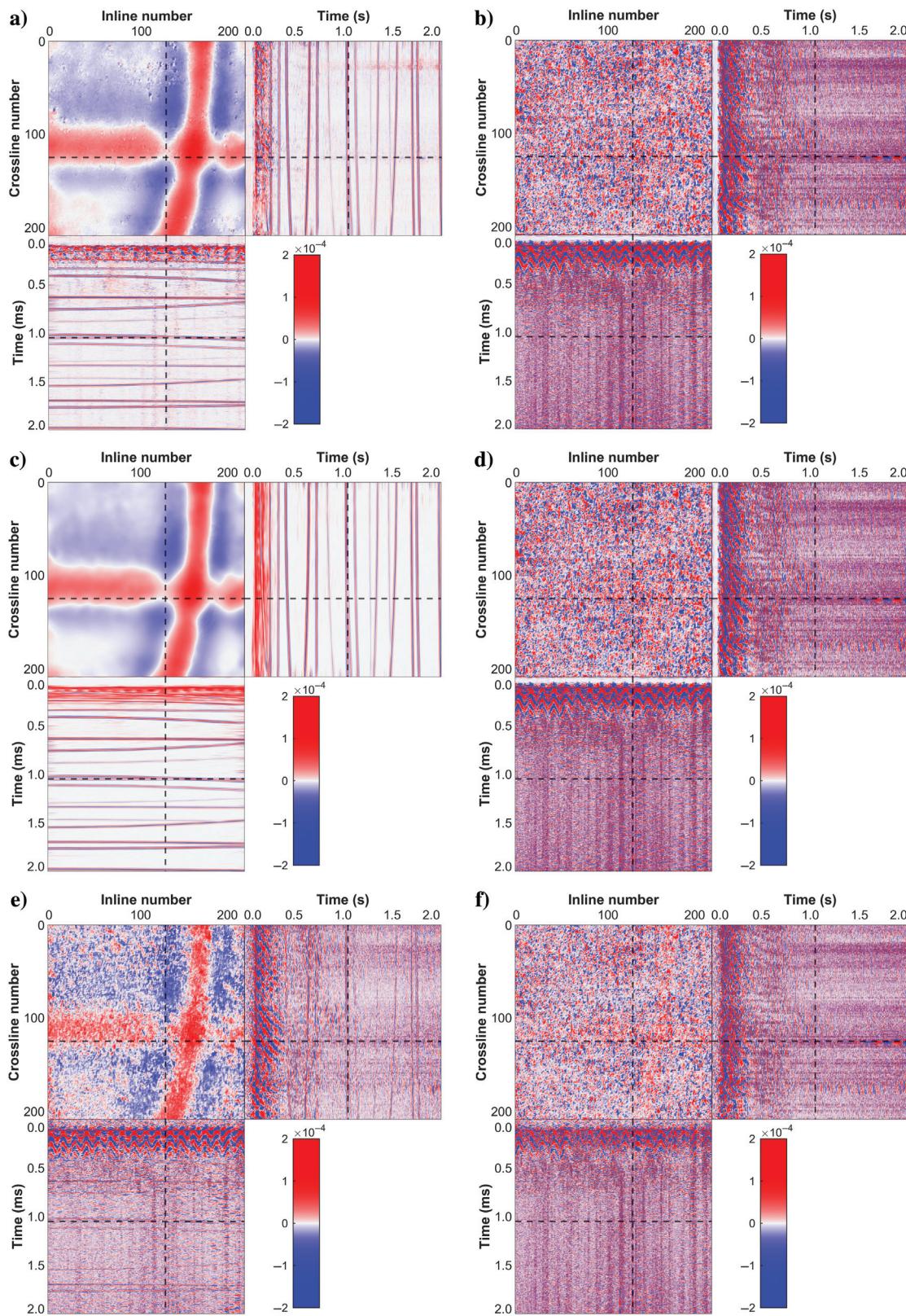


Figure 6. Results of synthetic test data. We apply the well-trained network to the test data set. (a) The scattered noise removed by the network with residual learning. (b) The denoised results obtained by the network with residual learning. (c) The scattered noise removed by the network with direct signal learning. (d) The denoised results obtained by the network with direct signal learning. It can be seen that removing residual learning can achieve better scattered noise attenuation. (e) The denoised results obtained by the f - x - y filtering method. (f) The scattered noise removed by the f - x - y filtering method. The scattering is coherent and contains visible linear features, making the conventional method ineffective.

the previous synthetic experiment that the direct signal learning strategy is superior residual learning in strong scattered noise attenuation.

Next, we compare middle-offset results. It is interesting to note that the S/N of noisy middle-offset gathers is relatively high, and 3D CWT achieves good results on it, which is the reason why we train our network on middle-offset gathers. From the time slice results in Figure 10, we can see that both methods successfully attenuate the scattered noise. As we can see in Figure 10c and 10e, there is hardly any evidence of signal leakage, illustrating the high reliability of both methods. Closer examination of the time slices in Figure 10b and 10d shows that our method performs better than 3D CWT because 3D CWT results still have some noise residues. What also is clear is that our network not only learns the denoising ability from a 3D CWT but also improves it further. Especially in the area shown in the red boxes, some structural interferences can be observed in the denoised results using 3D CWT, but our method yields a more complete suppression of scattered noise.

The S/N of far-offset OVT gathers is lower than the middle offset, as displayed in Figure 11a. Observing the denoised results in Figure 11d, our method also yields a good result, indicating that the network has mastered the ability for denoising the whole survey. Moreover, the red boxes in Figure 11 highlight that our network performs better than 3D CWT on the strong energy noise at the boundary. It results from the adoption of an effective strategy by direct learning signals, as the signal is relatively constant whereas the scattered noise fluctuates greatly and is difficult to learn. This also confirms that our method is more adaptive than 3D CWT. Despite the fact that the training samples are constructed by 3D CWT, we find that many details of the network results are superior to 3D CWT.

In the next step, we explore the stacked results. The OVT volumes are stacked according to the stack range shown in Figure 8. The partial stacking process is widely used for prestack fracture prediction because the S/N of a single OVT is too low. The crossline section results are shown in Figure 12. From the stacked results without noise attenuation in Figure 12a, it is obvious that the S/N of the stacked OVT data has improved significantly in comparison with any original single OVT volume. In spite of this, it also is evident that the resolution is reduced after stacking, and a considerable amount of noise can still be detected. The stacked results of denoised data in Figure 12b and 12d reveal that our proposed methods and 3D CWT can enhance weak signals and produce clearer

events. Furthermore, the energy consistency of our denoised results is better. Particularly near the yellow arrow, 3D CWT distorts the reflections and leaks useful signals to the noise section at a greater rate. Figure 12f and 12g presents the stacked results after denoising by a general use of labels. This also accords with our previous observations in Figure 9f and 9g, which showed that a general use of labels leads to more signal leakage, especially in the red arrow region. In addition, our method can obtain more continuous results in the time slice, as shown in Figure 13d. Compared with Figure 13b, it is clearly noticeable that there are obvious noise residues of 3D CWT, but our method attenuates them well.

To further illustrate the effectiveness of our method, we denoise all 1260 OVT gathers and sort them into common-midpoint gathers. Due to the scattered noise, the primary reflections are seriously obscured in Figure 14a. Our methods and 3D CWT are successful in attenuating strong scattered noise, as demonstrated by the results in Figure 14b and 14d, which also has been confirmed previously. Through detailed examination, it is detectable that there is a very modest degree of signal leakage as indicated by the yellow arrows in Figure 14c. However, our method has better fidelity. In addition, we enlarge the green box area for better comparison. An inspection of the red boxes in Figure 15 reveals that events have a better continuity after denoising with our method. In addition, it is a

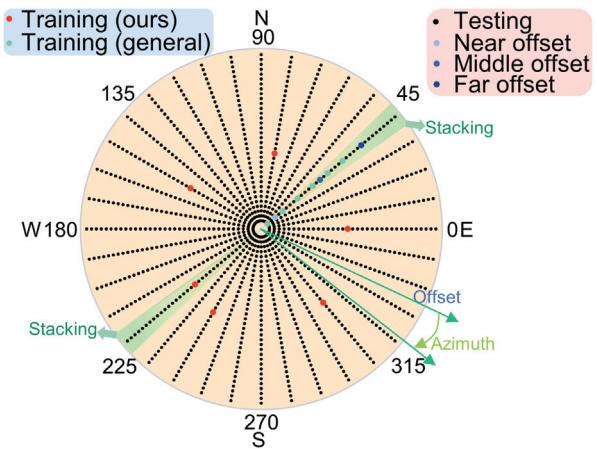


Figure 8. Maps of OVT volumes in offset-azimuth polar coordinates.

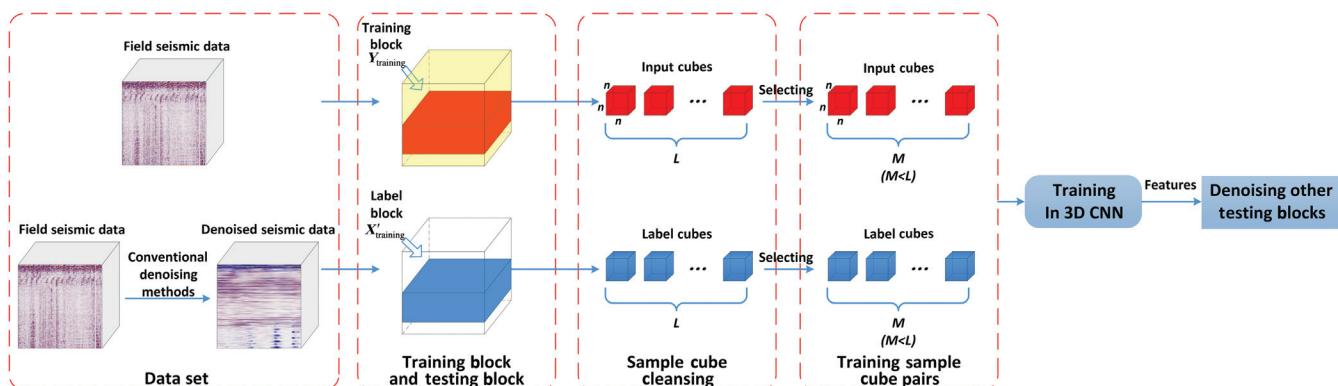


Figure 7. Workflow of the proposed method.

substantial benefit that our method can simultaneously suppress coherent and random noise, making it useful in practice.

In addition to the denoising effect, calculation time also is one of our major concerns. Table 2 presents the time consumption comparison. The advantage of 3D CWT is that it does not require any training. However, it takes 68.5 min to process one OVT gather with the Intel Xeon E5-1620 v3 CPU, so the total time to process all OVT traces is still very long, approximately 61 days. The computational time is measured based on 3D CWT with fast Fourier transform using the C program, so it can hardly dramatically reduce the computing time of 3D CWT. Our method provides a significant

reduction in computational time. Including the training time, it only takes 4.1 days, which can alleviate the excessive computation time problem in the OVT domain. In addition, we point out that we use a single GPU here, namely the Nvidia GeForce RTX 2080 Ti, and the calculation time can be further reduced with additional GPUs.

DISCUSSION

Denoising performance

OVT gathers have excellent data consistency, and 3D CWT and our method benefit from it. It is somewhat surprising that deep

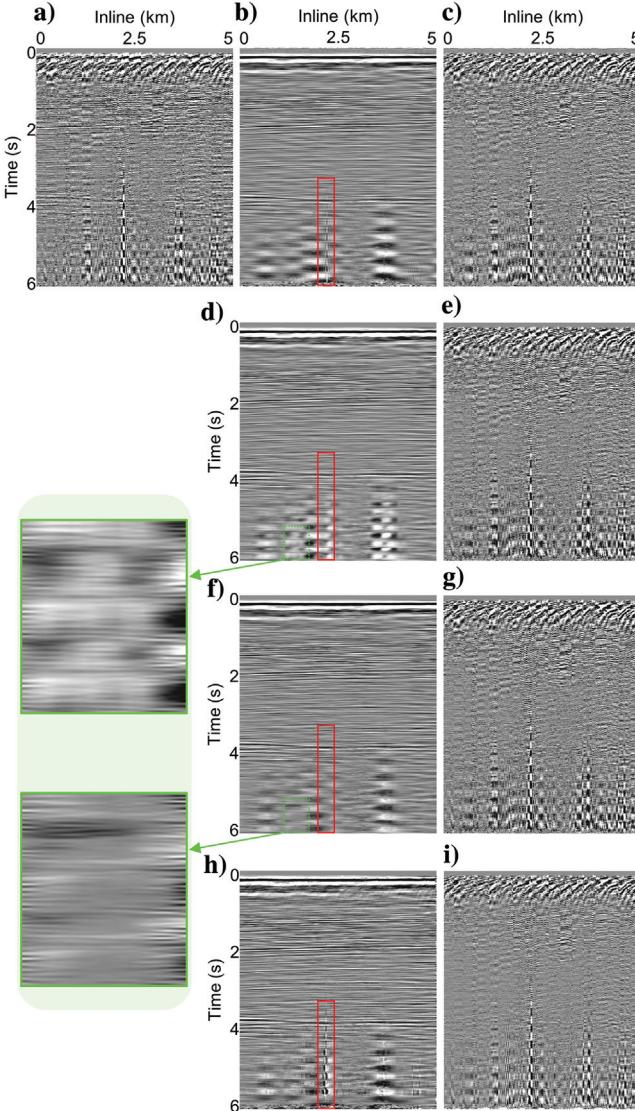


Figure 9. Crossline section results of a near-offset OVT volume at 2.5 km. (a) Noisy data, (b and c) results using 3D CWT, (d and e) our results, (f and g) a general use of labels using our proposed network architecture, and (h and i) the results based on residual learning using the same training samples as our method. As indicated by the red boxes, 3D CWT has few noise residues, whereas our network can effectively remove them. Residual learning also has an inferior performance due to evident noise residues. The green boxes reveal that a general use of labels causes more deep-layer signal leakage than our method.

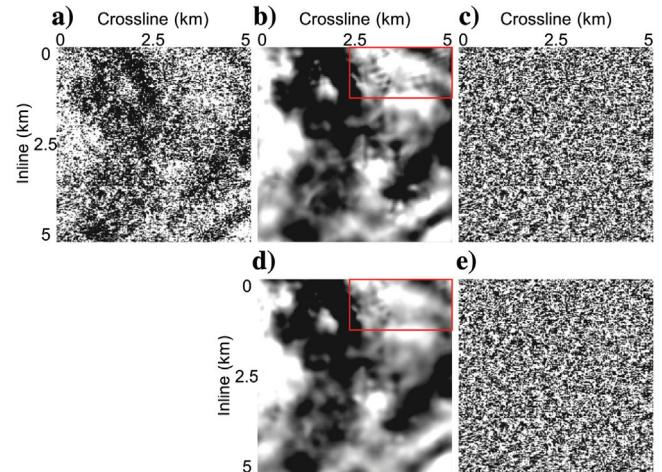


Figure 10. Time slice results of a middle-offset OVT volume at 3500 ms. (a) Noisy data, (b and c) results using 3D CWT, and (d and e) our results. As is evident in the red boxes, some structural interferences can be observed in the denoised results using 3D CWT, but our method yields a more complete suppression of scattered noise.

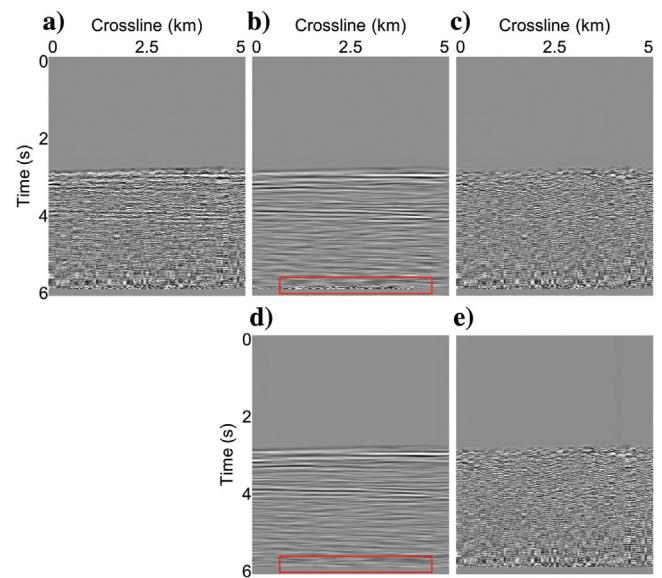


Figure 11. Inline section results of a far-offset OVT volume at 2.5 km. (a) Noisy data, (b and c) results using 3D CWT, and (d and e) our results. The red boxes indicate that our network performs better than 3D CWT on the strong energy noise.

learning outperforms the 3D CWT methods even though the training labels are constructed by 3D CWT. In the field of image denoising, it is almost impossible for supervised deep learning to surpass labels because their labels are ground truth. In addition to the appropriate network structure, the main reason for our better performance is the rational use of labels. On the one hand, middle offsets have a relatively high S/N, and thus 3D CWT can achieve better denoising results. With high-quality samples, it is easier for the network to learn the distribution of the reflections, so it produces cleaner denoising results at near- and far-offset gathers compared with 3D CWT. In other words, deep learning has learned the most beneficial aspects of 3D CWT and generalized them to the entire survey. In a contrast, parameters of 3D CWT are sensitive to noise energy and must be fine-tuned with different noise levels, which is highly challenging in practice. Therefore, fixed parameters in 3D CWT lead to a poor denoising effect in certain areas. On the other hand, OVT is a popular wide-azimuth cross-spread oriented processing technology that can analyze azimuthal anisotropy in a convenient way. To a certain extent, we have incorporated azimuthal information because we randomly chose different azimuthal OVT gathers. Nevertheless, we acknowledge that six OVTs in different

directions were chosen based on experience, which requires further research to determine at least how many OVTs are adequate.

In light of the preceding analysis, there are several possible ways to continue to improve the denoising effect of the network. First, we can construct more realistic synthetic data whose distribution is closer to the field data and jointly train the network with the field data. The network will learn a more accurate mapping toward reflections, thus improving the denoising effect. Second, the denoising performance can be further enhanced by incorporating azimuth-related information on a large scale. Combining multimodal deep learning and reciprocal OVT gathers Vermeer (2007) may be helpful. Third, another straightforward way is to find a more appropriate network architecture, such as the transformer (Tay et al., 2020), or by using the neural architecture search. However, this puts forward higher requirements for computing resources.

Generalization capability

The typical objective of a supervised deep learning system is to minimize the noncomputable expected risk by approximately minimizing the computable empirical risk (Kawaguchi et al., 2020). The generalization gap refers to the difference between empirical and expected risk, which explains the dependency of a trained model on the unseen training data set. One desirable goal of machine learning is enhancing generalization ability (minimizing the generalization gap). Because it is difficult to collect many noisy and clean data training pairs, a mismatch between training and testing data commonly exists. The latter is a significant obstacle in applying deep learning for processing seismic field data. Using a direct learning strategy, we have introduced a novel way of attenuating scattered noise in the OVT domain. In this situation, the networks effectively capture the intrinsic features of valuable signals. Although noise exhibits different features in different OVT gathers, the network still generalizes well to other OVT gathers in the same study area. The field data experiment shows that the proposed strategy requires a reasonably low label proportion of $6/1260 \approx 0.48\%$, considerably less than the usual

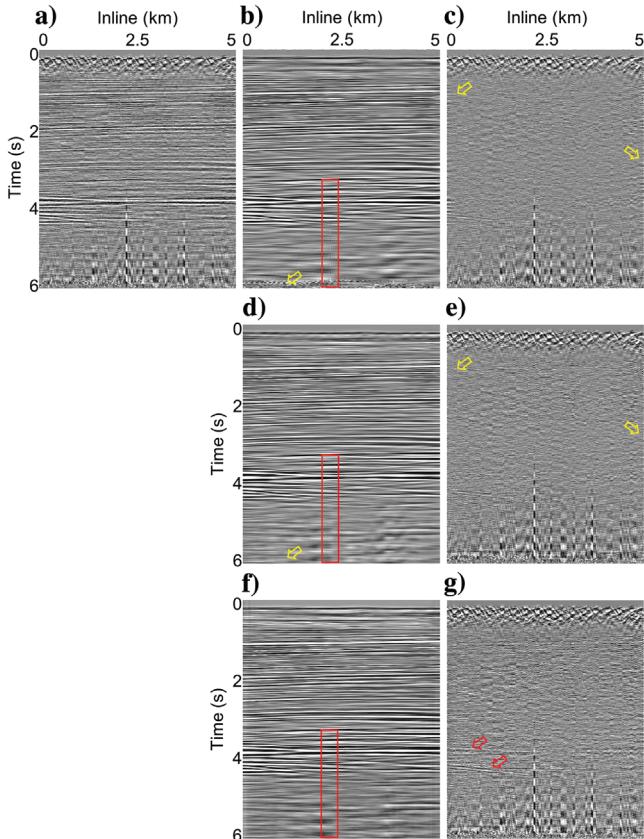


Figure 12. Crossline section results of a stacked OVT volume at 2.5 km. (a) Noisy data without noise attenuation, (b and c) results using 3D CWT, (d and e) our results, and (f and g) a general use of labels using our proposed network architecture. The red boxes illustrate that 3D CWT has few noise residues, whereas our network can remove them effectively. The yellow arrows indicate that our network is more robust than 3D CWT regarding strong energy noise at the boundary. As indicated by the red arrows, a general use of labels leads to more signal leakage.

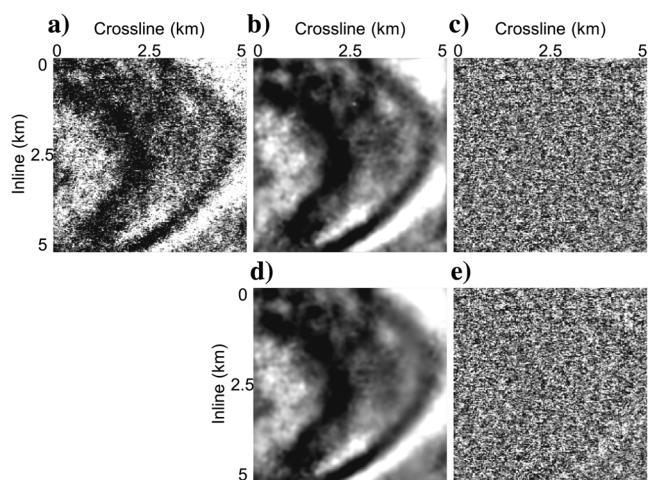


Figure 13. Time slice results of a stacked OVT volume at 3000 ms. (a) Noisy data without noise attenuation, (b and c) results using 3D CWT, and (d and e) our results. It is apparent that 3D CWT has obvious noise residues in the denoised results, but our method attenuates them well.

proportion (10%) of supervised learning currently used in similar research.

Different factors affect the generalization capability of networks for seismic data denoising, such as the type of survey used to acquire the data, subsurface geologic conditions, seismic source wavelets, noise types, S/N, etc. If one considers the scattered noise attenuation task as an example, the network may not produce satisfactory results if the input data have a different bandwidth from the training data. The preceding influencing factors may change substantially even within the same work area. For example, a land work area may span the Loess Plateau and desert area, making the network difficult to generalize. In addition, the diversities of seismic domains and noise types are two significant factors limiting the generalization of the network. For example, suppose that the network is trained to attenuate scattered noise in the OVT domains. In that case, there is a very low probability that the network can attenuate scattered noise in other seismic domains. Furthermore, seismic noise types are diverse, and each kind of noise has different characteristics, making it problematic to reach a significant level of generalization. For instance, training a network for scattered noise attenuation would not serve to remove ground roll because they have different features, violating the assumption that training and testing data arise from the same

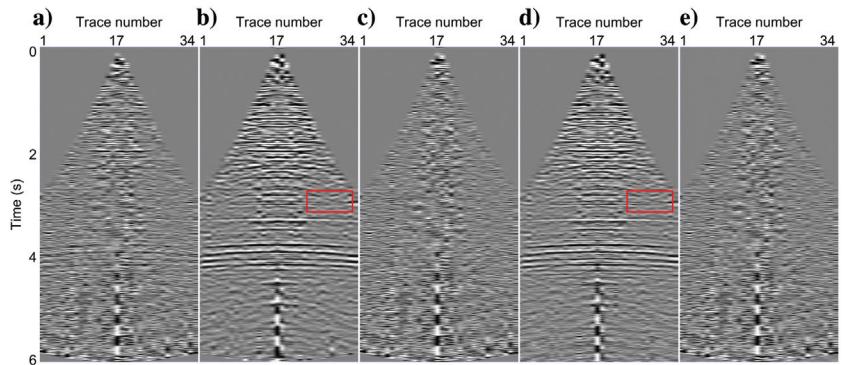


Figure 15. Magnified denoising result comparison in common-midpoint gathers. (a) Noisy data, (b and c) results using 3D CWT, and (d and e) our results. The red boxes indicate that reflections have a better continuity after denoising with our method.

Table 2. Time consumption comparison.

	Training time	Test time for one OVT (min)	Total time for 1260 OVTs (days)
3D CWT	—	68.5	61
Our method	Two days	5.9	4.1

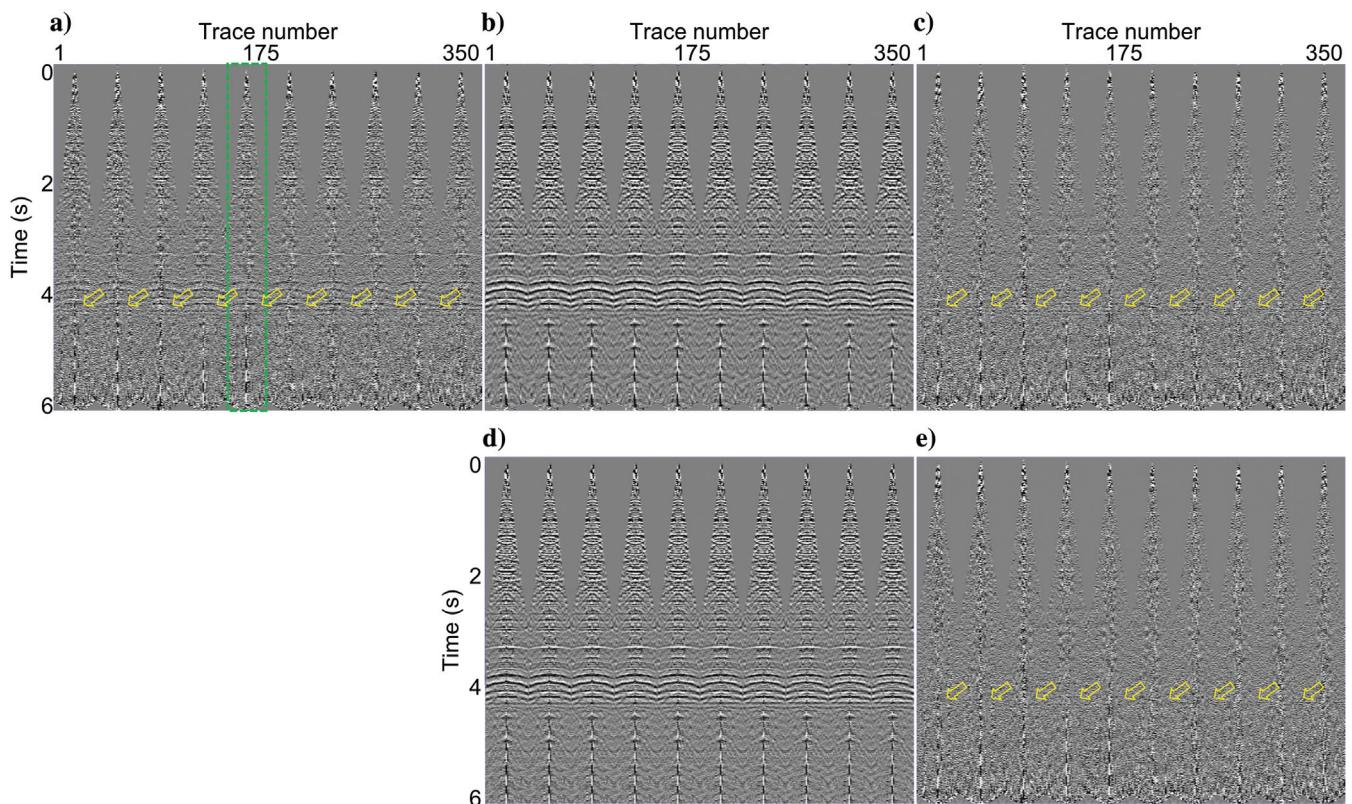


Figure 14. Denoising results of common-midpoint gathers. (a) Noisy data, (b and c) results using 3D CWT, and (d and e) our results. By comparing the yellow arrows, a lower degree of signal leakage can be observed from our results. The amplified region is marked in the green box.

distribution. As a preliminary step to generalizing the proposed method to data sets not included in the training phase, we need large-scale and diverse training data presented in different domains, noise types, frequency bandwidths, etc. The latter might be unfeasible given the size of such a data set, and hence we believe that it is often better to restrain the application of deep learning to one particular domain and for a specific type of noise as is done in this paper.

It also is important to note that overfitting in deep networks is another crucial factor that may impede the generalization. Overfitting leads to poor generalization capability, and several regularization methods have been proposed to alleviate it. Typically, there are two types of regularizations: implicit and explicit. Explicit regularizations can be adopted to avoid overfitting; they are not structural parts of the network architecture, the algorithms, or the data, so usually, explicit regularization can be added or removed easily. Typical examples are weight decay, loss function update, dropout, data augmentation, and early stopping. In contrast, the implicit regularization method incorporates the characteristics of the network architecture, the learning algorithm, or the data to regulate the effective generalization capability of a neural network. Examples are optimization methods, convolution layers, and BN. Please note that the preceding classification is not strict. For example, the dropout can sometimes be regarded as an implicit regularization method (Wei et al., 2020). Once the network is trained using regularization methods with varying data sets containing abundant noise types, it should have a strong generalization ability to process different field data sets unseen during the training phase and be suitable for other noise processing tasks.

CONCLUSION

We propose a method for denoising prestack strong scattered noise by combining the advantages of deep learning and OVT partitioning. In the OVT domain, the wavefield continuity and data consistency provide a conducive learning environment for the network. The massive amount of OVT gathers can make full use of the high computational efficiency of deep learning. We select a straightforward network architecture to learn reflections and randomly select the middle-offset OVT gathers as the training volumes. Our field data results demonstrate that only a limited number of OVT volumes is sufficient to suppress the noise of the entire survey. This approach offers great potential in practice because prestack data are enormous, and deep learning can be up to 10 times less time-consuming than the conventional methods described in this paper. In addition, weak signals in the deep layer also can be better recovered by our method. Consequently, our approach is well suited for applications in which azimuthally varying attributes are useful, such as fracture detection.

ACKNOWLEDGMENTS

This research is sponsored by the National Natural Science Foundation of China (41774135 and 41974131). The work of D. Liu was supported by the China Scholarship Council. Many thanks go to the anonymous assistant editor, the associate editor A. Stanton, and other anonymous reviewers for their insightful suggestions.

DATA AND MATERIALS AVAILABILITY

Data associated with this research are available and can be obtained by contacting the corresponding author.

REFERENCES

- Abma, R., and J. Claerbout, 1995, Lateral prediction for noise attenuation by t_x and f_x techniques: *Geophysics*, **60**, 1887–1896, doi: [10.1190/1.1443920](https://doi.org/10.1190/1.1443920).
- Bekara, M., and M. Van der Baan, 2007, Local singular value decomposition for signal enhancement of seismic data: *Geophysics*, **72**, no. 2, V59–V65, doi: [10.1190/1.2435967](https://doi.org/10.1190/1.2435967).
- Chase, M. K., 1992, Random noise reduction by FXY prediction filtering: *Exploration Geophysics*, **23**, 51–55, doi: [10.1071/EG992051](https://doi.org/10.1071/EG992051).
- Duan, W., J. Pei, and F. Li, 2016, Acquisition footprints supersession with source and receiver line interpolation in the OVT domain: *Oil Geophysical Prospecting*, **51**, 40–48, doi: [10.13810/j.cnki.issn.1000-7210.2016.01.006](https://doi.org/10.13810/j.cnki.issn.1000-7210.2016.01.006).
- Farge, M., 1992, Wavelet transforms and their applications to turbulence: *Annual Review of Fluid Mechanics*, **24**, 395–458, doi: [10.1146/annurev.fl.24.010192.002143](https://doi.org/10.1146/annurev.fl.24.010192.002143).
- 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).
- Goodfellow, I., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, 2014, Generative adversarial nets: *Advances in Neural Information Processing Systems*, 27.
- Grossmann, A., and J. Morlet, 1984, Decomposition of Hardy functions into square integrable wavelets of constant shape: *SIAM Journal on Mathematical Analysis*, **15**, 723–736, doi: [10.1137/0515056](https://doi.org/10.1137/0515056).
- Gu, S., L. Zhang, W. Zuo, and X. Feng, 2014, Weighted nuclear norm minimization with application to image denoising: *IEEE Conference on Computer Vision and Pattern Recognition*, 2862–2869.
- Gulunay, N., 1986, FXDECON and complex Wiener prediction filter: 56th Annual International Meeting, SEG, Expanded Abstracts, 279–281, doi: [10.1190/1.1893128](https://doi.org/10.1190/1.1893128).
- Han, X. L., C. Strobbia, J.-F. Dervieux, Q. Shi, C. Qi, C. E. Hinz, Z. Fan, S. Baghdadi, and J. Cui, 2016, Coherent and scattering surface wave attenuation with model-based method: A case study from tight-gas oilfield of China: *International Geophysical Conference, SPG/SEG*, 752–754, doi: [10.1190/IGCBeijing2016-229](https://doi.org/10.1190/IGCBeijing2016-229).
- Kaur, H., S. Fomel, and N. Pham, 2020, Seismic ground-roll noise attenuation using deep learning: *Geophysical Prospecting*, **68**, 2064–2077, doi: [10.1111/1365-2478.12985](https://doi.org/10.1111/1365-2478.12985).
- Kaur, H., N. Pham, and S. Fomel, 2019, Seismic data interpolation using CycleGAN: 89th Annual International Meeting, SEG, Expanded Abstracts, 2202–2206, doi: [10.1190/segam2019-3207424.1](https://doi.org/10.1190/segam2019-3207424.1).
- Kawaguchi, K., L. P. Kaelbling, and Y. Bengio, 2020, Generalization in deep learning: arXiv:1710.05468 [cs, stat].
- Li, C., Y. Zhang, and C. C. Mosher, 2019, A hybrid learning-based framework for seismic denoising: *The Leading Edge*, **38**, 542–549, doi: [10.1190/tle38070542.1](https://doi.org/10.1190/tle38070542.1).
- Li, F., W. Duan, R. Zhao, L. Luo, and Q. Dang, 2015, Seismic signal to noise ratio improvement with volume tau-p transform in OVT domain: *Oil Geophysical Prospecting*, **50**, 418–423+2.
- Li, H., W. Yang, and X. Yong, 2018, Deep learning for ground-roll noise attenuation: 88th Annual International Meeting, SEG, Expanded Abstracts, 1981–1985, doi: [10.1190/segam2018-2981295.1](https://doi.org/10.1190/segam2018-2981295.1).
- Li, J., D. Wang, S. Ji, Y. Li, and Z. Qian, 2017, Seismic noise suppression using weighted nuclear norm minimization method: *Journal of Applied Geophysics*, **146**, 214–220, doi: [10.1016/j.jappgeo.2017.09.013](https://doi.org/10.1016/j.jappgeo.2017.09.013).
- Ling, Y., and S. Hu, 2019, Internal multiple attenuation by OVT migration and its application in carbonate rock environment: 81st Annual International Conference and Exhibition, EAGE, Extended Abstracts, doi: [10.3997/2214-4609.201901021](https://doi.org/10.3997/2214-4609.201901021).
- Liu, D., W. Chen, M. D. Sacchi, and H. Wang, 2020a, Should we have labels for deep learning ground roll attenuation? 90th Annual International Meeting, SEG, Expanded Abstracts, 3239–3243, doi: [10.1190/segam2020-3425692.1](https://doi.org/10.1190/segam2020-3425692.1).
- Liu, D., W. Wang, X. Wang, C. Wang, J. Pei, and W. Chen, 2020b, Poststack seismic data denoising based on 3-D convolutional neural network: *IEEE Transactions on Geoscience and Remote Sensing*, **58**, 1598–1629, doi: [10.1109/TGRS.2019.2947149](https://doi.org/10.1109/TGRS.2019.2947149).
- Matharu, G., W. Gao, R. Lin, Y. Guo, M. Park, and M. D. Sacchi, 2020, Simultaneous source deblending using a deep residual network: *SEG 2019 Workshop: Mathematical Geophysics: Traditional vs Learning*, 13–16, doi: [10.1190/iwmg2019_04.1](https://doi.org/10.1190/iwmg2019_04.1).
- Miorali, M., K. Mills, M. Igoue, and D. McClymont, 2018, Near-surface data-driven methods for surface wave and multiple removal, onshore Kenya: 80th Annual International Conference and Exhibition, EAGE, Extended Abstracts, doi: [10.3997/2214-4609.201801415](https://doi.org/10.3997/2214-4609.201801415).
- Morlet, J., G. Arens, E. Fourgeau, and D. Giard, 1982, Wave propagation and sampling theory — Part II: Sampling theory and complex waves: *Geophysics*, **47**, 222–236, doi: [10.1190/1.1441329](https://doi.org/10.1190/1.1441329).
- Oropeza, V., and M. Sacchi, 2011, Simultaneous seismic data denoising and reconstruction via multichannel singular spectrum analysis: *Geophysics*, **76**, no. 3, V25–V32, doi: [10.1190/1.3552706](https://doi.org/10.1190/1.3552706).

- Pham, N., and W. Li, 2022, Physics-constrained deep learning for ground roll attenuation: *Geophysics*, **87**, no. 1, V15–V27, doi: [10.1190/geo2020-0691.1](https://doi.org/10.1190/geo2020-0691.1).
- Ronneberger, O., P. Fischer, and T. Brox, 2015, U-net: Convolutional networks for biomedical image segmentation: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 234–241.
- Saad, O. M., and Y. Chen, 2021, A fully unsupervised and highly generalized deep learning approach for random noise suppression: *Geophysical Prospecting*, **69**, 709–726, doi: [10.1111/1365-2478.13062](https://doi.org/10.1111/1365-2478.13062).
- Strobbia, C., A. Zarkhidze, R. May, and F. Ibrahim, 2014, Model-based attenuation for scattered dispersive waves: *Geophysical Prospecting*, **62**, 1143–1161, doi: [10.1111/1365-2478.12118](https://doi.org/10.1111/1365-2478.12118).
- Sun, M., Z. Li, Y. Qu, Z. Li, L. Li, J. Zhao, and B. Ning, 2019, A seismic denoising method based on curvelet transform with sparse constraint in OVT domain: *Geophysical Prospecting for Petroleum*, **58**, 208–218.
- Tay, Y., M. Dehghani, D. Bahri, and D. Metzler, 2020, Efficient transformers: A survey: arXiv:2009.06732 [cs].
- Trickett, S. R., 2003, F -xy eigenimage noise suppression: *Geophysics*, **68**, 751–759, doi: [10.1190/1.1567245](https://doi.org/10.1190/1.1567245).
- Vermeer, G. J. O., 1998, Creating image gathers in the absence of proper common-offset gathers: *Exploration Geophysics*, **29**, 636–642, doi: [10.1071/EG998636](https://doi.org/10.1071/EG998636).
- Vermeer, G. J. O., 2007, Reciprocal offset-vector tiles in various acquisition geometries: 77th Annual International Meeting, SEG, Expanded Abstracts, 61–65, doi: [10.1190/1.2792382](https://doi.org/10.1190/1.2792382).
- Vermeer, G. J. O., 2008, Alternative strategies for tackling scattered noise: 78th Annual International Meeting, SEG, Expanded Abstracts, 95–99, doi: [10.1190/1.3054863](https://doi.org/10.1190/1.3054863).
- Vinje, V., P. Zhao, and J. Gaiser, 2015, Offset vector tile gather extension and weighting to reduce footprint in dual-datum and converted-wave migration: 85th Annual International Meeting, SEG, Expanded Abstracts, 2103–2107, doi: [10.1190/segam2015-5899618.1](https://doi.org/10.1190/segam2015-5899618.1).
- Wang, X., J. Gao, W. Chen, and E. Zhang, 2010, On the method of detecting the discontinuity of seismic data via 3D wavelet transform: IEEE International Geoscience and Remote Sensing Symposium, 3945–3947.
- Wei, C., S. Kakade, and T. Ma, 2020, The implicit and explicit regularization effects of dropout: Proceedings of the 37th International Conference on Machine Learning, PMLR, 10181–10192.
- Xu, S., Y. Zhang, and G. Lambaré, 2010, Antileakage Fourier transform for seismic data regularization in higher dimensions: *Geophysics*, **75**, no. 6, WB113–WB120, doi: [10.1190/1.3507248](https://doi.org/10.1190/1.3507248).
- Yin, X., H. Zhang, and Z. Zong, 2018, Research status and progress of 5D seismic data interpretation in OVT domain: *Geophysical Prospecting for Petroleum*, **57**, 155–178.
- Yu, S., and J. Ma, 2021, Deep learning for geophysics: Current and future trends: *Reviews of Geophysics*, **59**, e2021RG000742, doi: [10.1029/2021RG000742](https://doi.org/10.1029/2021RG000742).
- Yu, S., J. Ma, and W. Wang, 2019, Deep learning for denoising: *Geophysics*, **84**, no. 6, V333–V350, doi: [10.1190/geo2018-0668.1](https://doi.org/10.1190/geo2018-0668.1).
- Zhang, F., D. Liu, X. Wang, W. Chen, and W. Wang, 2018, Random noise attenuation method for seismic data based on deep residual networks: International Geophysical Conference, 1774–1777, doi: [10.1190/IGC2018-435](https://doi.org/10.1190/IGC2018-435).
- Zhang, K., W. Zuo, Y. Chen, D. Meng, and L. Zhang, 2017, Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising: *IEEE Transactions on Image Processing*, **26**, 3142–3155, doi: [10.1109/TIP.2017.2662206](https://doi.org/10.1109/TIP.2017.2662206).
- Zhao, C., L. Jiang, X. Wang, D. Liu, Z. Shi, and W. Chen, 2021, Prestack seismic noise attenuation based on 3D CWT: First International Meeting for Applied Geoscience & Energy, SEG, Expanded Abstracts, 2834–2838, doi: [10.1190/segam2021-3594268.1](https://doi.org/10.1190/segam2021-3594268.1).
- Zhao, Y., Y. Li, X. Dong, and B. Yang, 2019, Low-frequency noise suppression method based on improved DnCNN in desert seismic data: *IEEE Geoscience and Remote Sensing Letters*, **16**, 811–815, doi: [10.1109/LGRS.2018.2882058](https://doi.org/10.1109/LGRS.2018.2882058).
- Zheng, Y., Y. Yuan, and X. Si, 2020, The improved DnCNN for linear noise attenuation: SEG 2019 Workshop: Mathematical Geophysics: Traditional vs Learning, SEG, Global Meeting Abstracts, 56–59, doi: [10.1190/iwmg2019_14.1](https://doi.org/10.1190/iwmg2019_14.1).

Biographies and photographs of the authors are not available.