Improving the Signal-to-Noise Ratio of Seismological Datasets by Unsupervised Machine Learning

by Yangkang Chen, Mi Zhang, Min Bai, and Wei Chen

ABSTRACT

Seismic waves that are recorded by near-surface sensors are usually disturbed by strong noise. Hence, the recorded seismic data are sometimes of poor quality; this phenomenon can be characterized as a low signal-to-noise ratio (SNR). The low SNR of the seismic data may lower the quality of many subsequent seismological analyses, such as inversion and imaging. Thus, the removal of unwanted seismic noise has significant importance. In this article, we intend to improve the SNR of many seismological datasets by developing new denoising framework that is based on an unsupervised machine-learning technique. We leverage the unsupervised learning philosophy of the autoencoding method to adaptively learn the seismic signals from the noisy observations. This could potentially enable us to better represent the true seismic-wave components. To mitigate the influence of the seismic noise on the learned features and suppress the trivial components associated with low-amplitude neurons in the hidden layer, we introduce a sparsity constraint to the autoencoder neural network. The sparse autoencoder method introduced in this article is effective in attenuating the seismic noise. More importantly, it is capable of preserving subtle features of the data, while removing the spatially incoherent random noise. We apply the proposed denoising framework to a reflection seismic image, depth-domain receiver function gather, and an earthquake stack dataset. The purpose of this study is to demonstrate the framework's potential in real-world applications.

INTRODUCTION

Seismic phases from the discontinuities in the Earth's interior contain significant constraints for high-resolution deep Earth imaging; however, they sometimes arrive as weak-amplitude waveforms (Rost and Weber, 2001; Rost and Thomas, 2002; Deuss, 2009; Saki *et al.*, 2015; Guan and Niu, 2017, 2018; Schneider *et al.*, 2017; Chai *et al.*, 2018). The detection of these weak-amplitude seismic phases is sometimes challenging because of three main reasons: (1) the amplitude of these phases is very small and can be neglected easily when seen next to the amplitudes of neighboring phases that are much larger; (2) the coherency of the weak-amplitude seismic phases is seriously degraded because of insufficient array coverage and

spatial sampling; and (3) the strong random background noise that is even stronger than the weak phases in amplitude makes the detection even harder. As an example of the challenges presented, the failure in detecting the weak reflection phases from mantle discontinuities could result in misunderstanding of the mineralogy or temperature properties of the Earth interior.

To conquer the challenges in detecting weak seismic phases, we need to develop specific processing techniques. In earthquake seismology, in order to highlight a specific weak phase, recordings in the seismic arrays are often shifted and stacked for different slowness and back-azimuth values (Rost and Thomas, 2002). Stacking serves as one of the most widely used approaches in enhancing the energy of target signals. Shearer (1991a) stacked long-period seismograms of shallow earthquakes that were recorded from the Global Digital Seismograph Network for 5 yr and obtained a gather that shows typical arrivals clearly from the deep Earth. Morozov and Dueker (2003) investigated the effectiveness of stacking in enhancing the signals of the receiver functions. They defined a signal-to-noise ratio (SNR) metric that was based on the multichannel coherency of the signals and the incoherency of the random noise, and they showed that the stacking can significantly improve the SNR of the stacked seismic trace. However, stacking methods have some drawbacks. First, they do not necessarily remove the noise present in the signal. Second, they require a large array of seismometers. Third, they require coherency of arrivals across the array, which are not always about earthquake seismology. From this point of view, a single-channel method seems to be a better substitute for improving the SNR of seismograms (Mousavi and Langston, 2016; 2017).

In the reflection seismology community, many noise attenuation methods have been proposed and implemented in field applications over the past several decades. Prediction-based methods utilize the predictive property of the seismic signal to construct a predictive filter that prevents noise. Median filters and their variants use the statistical principle to reject Gaussian white noise or impulsive noise (Mi *et al.*, 2000; Bonar and Sacchi, 2012). The dictionary-learning-based methods adaptively learn the basis from the data to sparsify the noisy seismic data, which will in turn suppress the noise (Zhang, van der Baan, *et al.*, 2018). These methods require experimenters to solve the dictionary-updating and sparse-coding methods and can be very

expensive, computationally speaking. Decomposition-based methods decompose the noisy data into constitutive components, so that one can easily select the components that primarily represent the signal and remove those associated with noise. This category includes singular value decomposition (SVD)-based methods (Bai *et al.*, 2018), empirical-mode decomposition (Chen, 2016), continuous wavelet transform (Mousavi *et al.*, 2016), morphological decomposition (Huang *et al.*, 2017), and so on. Rank-reduction-based methods assume that seismic data have a low-rank structure (Kumar *et al.*, 2015; Zhou *et al.*, 2017). If the data consist of κ complex linear events, the constructed Hankel matrix of the frequency data is a matrix of rank κ (Hua, 1992). Noise will increase the rank of the Hankel matrix of the data, which can be attenuated via rank reduction. Such methods include Cadzow filtering (Cadzow, 1988; Zu *et al.*, 2017) and SVD (Vautard *et al.*, 1992).

Most of the denoising methods are largely effective in processing reflection seismic images. The applications in more general seismological datasets are seldom reported, partially because of the fact that many seismological datasets have extremely low data quality. That is, they are characterized by low SNR and poor spatial sampling. Besides, most traditional denoising algorithms are based on carefully tuned parameters to obtain satisfactory performance. These parameters are usually data dependent and require a great deal of experiential knowledge. Thus, they are not flexible enough to use in application to many real-world problems. More research efforts have been dedicated to using machine-learning methods for seismological data processing (Chen, 2018a,b; Zhang, Wang, *et al.*, 2018; Bergen *et al.*, 2019; Lomax *et al.*, 2019; McBrearty *et al.*, 2019). Recently, supervised learning (Zhu *et al.*, 2018) has been successfully applied for denoising of the seismic signals. However, supervised methods with deep networks require very large training datasets (sometimes to an order of a billion) of clean signals and their noisy contaminated realizations. In this article, we develop a new automatic denoising framework for improving the SNR of the seismological datasets based on an unsupervised machine-learning (UML) approach; this would be the autoencoder method. We leverage the autoencoder neural network to adaptively learn the features from the raw noisy seismological datasets during the encoding process, and then we optimally represent the data using these learned features during the decoding process. To effectively suppress the random noise, we use the sparsity constraint to regularize the neurons in the hidden layer. We apply the proposed UML-based denoising framework to a group of seismological datasets, including a reflection seismic image, a receiver function gather, and an earthquake stack. We observe a very encouraging performance, which demonstrates its great potential in a wide range of applications.

METHOD

Unsupervised Autoencoder Method

We will first introduce the autoencoder neural network that we use for denoising seismological datasets. Autoencoders are specific neural networks that consist of two connected parts (decoder and encoder) that try to copy their input to the output layer. Hence, they can automatically learn the main features of the data in an unsupervised manner. In this article, the network is simply a three-layer architecture with an input layer, a hidden layer, and an output layer. The encoding process in the autoencoder neural network can be expressed as follows:

$$\mathbf{p} = \xi(\mathbf{W_1}\mathbf{x} + \mathbf{b_1}), \tag{1}$$

in which $\mathbf{x}$ is the training sample ($\mathbf{x} \in \mathbb{R}^n$), ξ is the activation function.

The decoding process can be expressed as follows:

$$\widehat{\mathbf{x}} = \xi(\mathbf{W_2}\mathbf{x} + \mathbf{b_2}). \tag{2}$$

In equations (1) and (2), $\mathbf{W_1}$ is the weighting matrix between the input layer and the hidden layer; $\mathbf{b_1}$ is the forward bias vector; $\mathbf{W_2}$ is the weighting matrix between the hidden layer and output layer; $\mathbf{b_2}$ is the backward bias vector; and ξ is the activation function. In this study, we use the softplus function as the activation function:

$$\xi(\mathbf{x}) = \log(1 + e^x). \tag{3}$$

Sparsity Regularized Autoencoder

To mitigate the influence of the seismic noise on the learned features and suppress the trivial components associated with low-amplitude neurons in the hidden layer, we apply a sparsity constraint to the hidden layer; that is, the output or last layer of the encoder. The sparsity constraint can help dropout the extracted nontrivial features that correspond to the noise and a small value in the hidden units. It can thus highlight the most dominant features in the data—the useful signals. The sparse penalty term can be written as follows:

$$\tilde{\mathbf{p}} = R(\mathbf{p}), \tag{4}$$

in which R is the penalty function:

$$R(\mathbf{p}) = \sum_{j=1}^{h} \mathrm{KL}(\mu \| p_j), \tag{5}$$

in which h is the number of neurons in the hidden layer and μ is a sparsity parameter. The sparsity parameter μ typically is a small value close to zero (e.g., 0.05). In other words, we would like the average activation of each hidden neuron to be close to 0.05. To satisfy this constraint, the hidden unit activations must mostly be near 0. p_j denotes the jth element of the vector $\mathbf{p}$. $\mathrm{KL}(\cdot)$ is the Kullback–Leibler divergence (Kullback and Leibler, 1951) function:

$$\mathrm{KL}(\mu \| p_j) = \mu \log \frac{\mu}{p_j} + (1 - \mu) \log \frac{1 - \mu}{1 - p_j}. \tag{6}$$

An important property of the KL function is that $\mathrm{KL}(\mu \| p_j) = 0$ if $\mu = p_j$, otherwise its value increases monotonically as p_j diverges from μ.

The cost function thus becomes:

$$\mathbf{J}(\mathbf{W}, \mathbf{b}) = \frac{1}{2} \|\widehat{\mathbf{x}} - \mathbf{x}\|_2^2 + \boldsymbol{\beta} R(\mathbf{p}), \qquad (7)$$

in which $\boldsymbol{\beta}$ is the weight controlling the sparsity penalty term. The cost function can be minimized using a stochastic gradient method. The gradients with respect to $\mathbf{W}$ and $\mathbf{b}$ can be derived from the backpropagation method (Vogl *et al.*, 1988).

We can extract the feature learned by the ith unit in the hidden layer and plot it in a 2D image. The learned feature of the ith unit corresponds to the part of the input image x that would maximally activate the ith hidden unit. Assume that the input x is normalized in the sense that $\|\mathbf{x}\|^2 \leq 1$, then the input part of the training data that maximally activates the ith hidden unit is given by:

$$y^j = \frac{W_1^{i:j}}{\sqrt{\sum_{j=1}^{N^2}(W_1^{i:j})^2}}, \qquad (8)$$

in which y_j denotes the jth element in the feature image corresponding to the ith hidden unit. Here, $\mathbf{y}$ denotes a vectorized 2D image with size $N \times N$. To view the feature in a 2D view, $\mathbf{y}$ needs to be rearranged into a 2D matrix and be plotted.

Patching and Unpatching

The learning process uses patch-based samples. In this article, preparing the training samples from the seismological datasets is referred to as the patching process. Conversely, reconstruction of the seismological datasets from filtered patches is referred to as the unpatching process. The patching and unpatching processes are illustrated in Figure 1. In the patching process, we slide a window of the patch size from the top to the bottom, as well as the left to the right, of the 2D seismic data. Thus, we obtain a patch in each sliding step. To avoid the discontinuity between patches when reconstructing, we arrange it so that each pair of neighbor patches shares an overlap. The size of the overlapping part is called the shift size. In this article, we define the shift size as half of the patch size. A large patch size would cause the learning process to miss small-scale features, whereas a small patch size would make the learning process incapable of learning meaningful waveform features. In this article, we define the patch size as approximately half of the dominant wavelength of data. The patches obtained from the sliding process are arranged as a 2D matrix, which is incorporated into the learning process. In the unpatching process, we reinsert each filtered patch from the 2D data matrix back into the seismological datasets. In the overlapping part of the reconstructed trace, we take the average of the two neighbor patches. The proposed UML algorithm is not limited to multichannel seismic data. It can also be used to learn
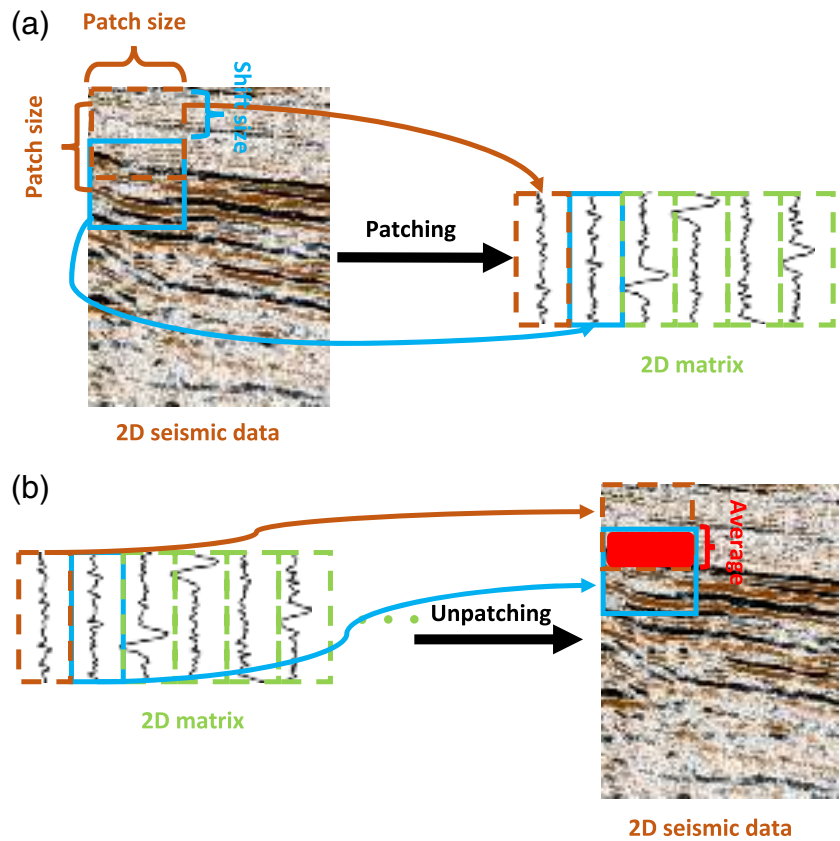

▲ **Figure 1.** Cartoons illustrating the principles of (a) patching and (b) unpatching. The color version of this figure is available only in the electronic edition.

the features from 1D seismic data, such as sparsely recorded earthquake data or microseismic data.

RESULTS

We first apply the proposed algorithm to a reflection seismic image. The image is presented in Figure 2a. The 2D seismic image is extracted from a migrated 3D seismic image that is related to an oilfield in China. There is significant noise in the 2D seismic image, which compromises the coherency of the seismic events. There are several complicated structures in this 2D seismic image. First, the amplitude exhibits a strong variation from the left to the right. Second, there are some weak events in the 2D section, particularly in the deep part around 1.7 s. Third, the strong noise causes obvious discontinuities of the events, which makes the tracking of most seismic events difficult. The denoised data using the proposed method are shown in Figure 2d. The removed noise from the noisy data using the proposed method is plotted in Figure 2g. Upon the removal of the random noise, the seismic events become more continuous, and the weak events in the deep part become more evident. Additionally, the spatial amplitude variations in the dataset are well preserved. In the removed noise section (Fig. 2g), we do not see much coherent energy, which indicates that the removed noise is purely random noise and that we are not damaging any useful signals. In this example, we compare
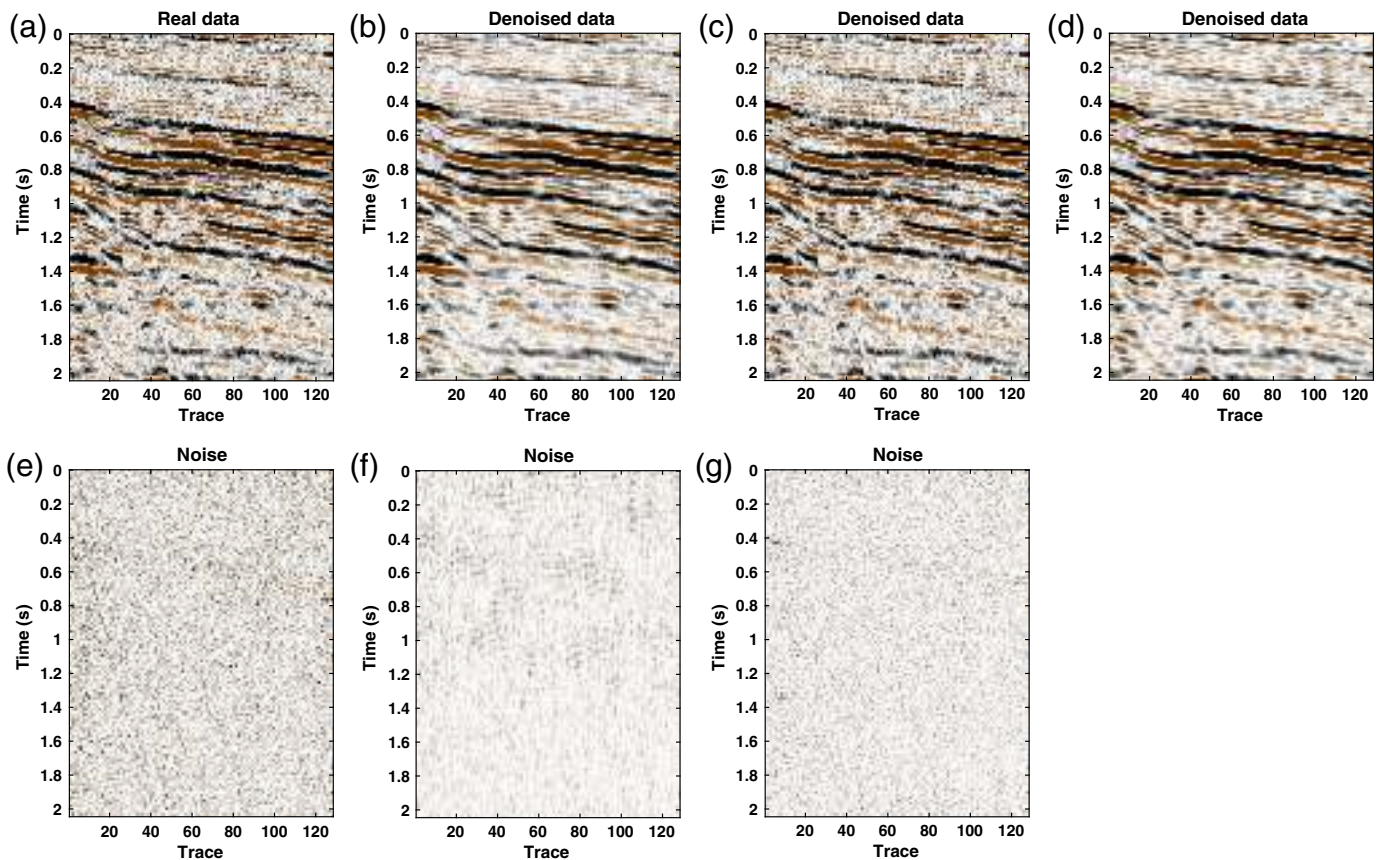
▲ **Figure 2.** Denoising performance of the reflection seismic image. (a) Reflection seismic image; (b) denoised data using the prediction-based method; (c) denoised data using the band-pass-filtering method; (d) denoised data based on the unsupervised machine learning (UML) method; (e) removed noise corresponding to (b); (f) removed noise corresponding to (c); and (g) removed noise corresponding to (d). The color version of this figure is available only in the electronic edition.

the performance of the proposed algorithm with the most widely used methods in the industry, namely the frequency-space domain prediction-based method (Canales, 1984) and the band-pass-filtering method. The result from the prediction-based method is displayed in Figure 2b, where we use a filter length that is equal to six points. The removed noise corresponding to the prediction-based method is shown in Figure 2e. However, from the denoised data shown in Figure 2b, we can observe that there is a significant amount of residual noise left in the image. The result from the band-pass-filtering method is shown in Figure 2c, where we use it to preserve the frequency contents between 0 and 25 Hz. It is difficult to compromise the signal preservation and noise removal for the band-pass-filtering method. If we use a higher cutoff frequency, then more noise will be left in the result, and the denoising performance will not be obvious. If we use a lower cutoff frequency, we will inevitably remove some signal's energy. The removed noise is shown in Figure 2f, which contains significant coherent signals.

Because there is no ground-truth solution in the real data example, we cannot use a quantitative metric (e.g., the SNR) to evaluate the denoising performance. However, we can use the local similarity metric to quantitatively measure the signal

damage. The local similarity metric is based on the assumption that the denoised signal and removed noise should be orthogonal to each other and have low similarity locally. The detailed introduction of utilizing the local similarity metric to evaluate denoising performance is given in Chen and Fomel (2015). For two competing methods, when a similar amount of noise is removed, more signal damages indicate a poorer denoising performance. We calculate the local similarity maps between the denoised data and the removed noise for the proposed method and the prediction-based method, and we show them in Figure 3. In the local similarity maps, the high local similarity anomaly shows where the denoised signal and the removed noise are very similar; it thus points out where large signal damage (or leakage) exists. From Figure 3, it is obvious that the local similarity values of the prediction-based method and the band-pass-filtering method are higher than that of the proposed method. Thus, the proposed method helps preserve useful signals more effectively than the prediction-based method. It is worth noting that the same concept was also proposed in Li *et al.* (2018), where the local similarity is defined as the signal consistency between the examined station and its nearest neighbors. In this article, the local similarity is a more general concept to evaluate the closeness of two arbitrary signals.
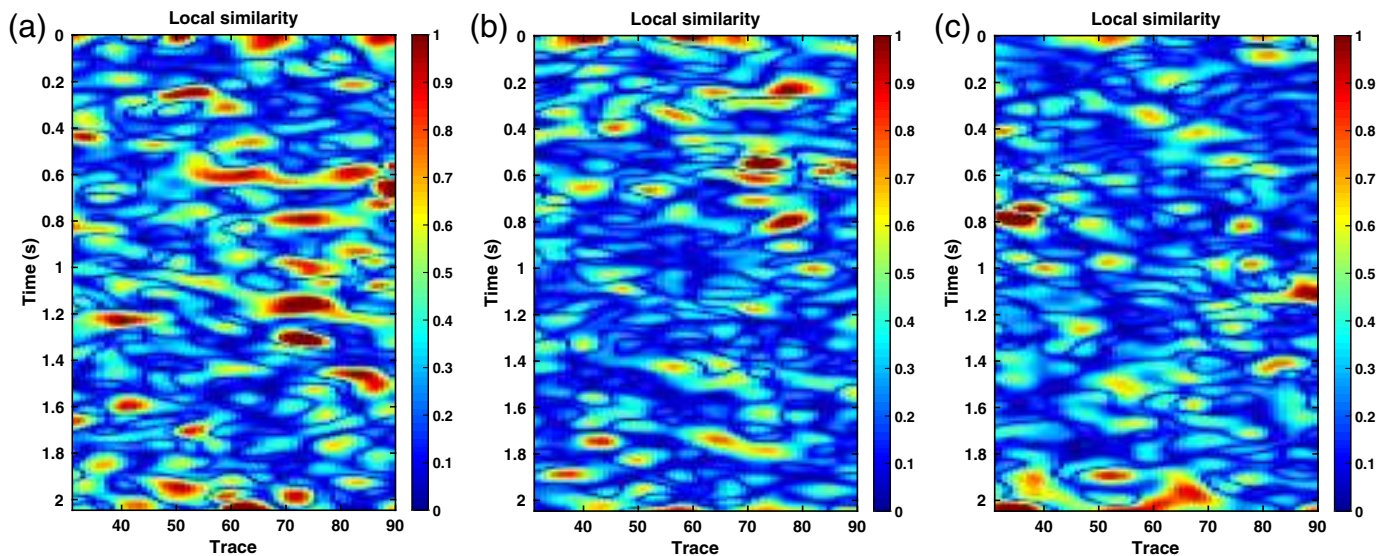
▲ **Figure 3.** Local similarity between the denoised data and the removed noise. The high similarity anomaly indicates areas with serious signal damages. (a) Local similarity corresponding to the prediction-based method. (b) Local similarity corresponding to the band-pass-filtering method. (c) Local similarity corresponding to the proposed method. Note the similarity anomalies in (a) and (b) are obviously higher than in (c). The color version of this figure is available only in the electronic edition.

Figure 4 shows the extracted 64 features using the proposed UML algorithm. Each feature is rearranged into a 40×40 2D matrix. It is clear that the extracted features correspond to different structural features of the seismic image.

We then apply the proposed denoising algorithm to a receiver function dataset. Figure 5a shows a stacked common receiver gather for the WALA station at Waterton Lake, Alberta. The WALA station belongs to the Canadian National Seismograph Network (Gu *et al.*, 2015). Each column in the matrix (Fig. 5a) corresponds to the stacked receiver function data of one specific epicentral distance corresponding to the WALA station. The two green solid lines in Figure 5a show the expected arrivals of the converted waves, P410s and P660s. To enhance the structure revealed from the receiver function data, the time-domain receiver function gather (Fig. 5a) is first transformed to the depth domain to correct the phase moveout; then, all receiver function data of different epicentral distances are stacked to output the structure, such as the 410 and 660 discontinuities, underneath the WALA station. The converted receiver function data in the depth domain are shown in Figure 5b, where the seismic phases are well aligned horizontally. However, because of the strong noise, the stacked receiver function data and the inferred Earth structure are of low fidelity and thus not reliable. We apply the proposed method to filter the strong random noise and obtain a much better receiver function gather with obviously more coherent seismic phases, which is plotted in Figure 5c. The removed noise from the noisy receiver function data (Fig. 5b) is shown in Figure 5d. From the removed noise, we can barely see that obvious signal energy and the noise are mostly spatially incoherent; this indicates a signal-preserving denoising performance of the proposed method.

To evaluate the fidelity of filtered receiver function gather, we use the local similarity metric. We calculate the local similarity between denoised data and noisy data and show it in Figure 6b. The high local similarity anomaly in Figure 6a indicates where the denoised signal is distinctly close to the noisy data and thus is of high fidelity. It is also clear that the 410 and 660 arrivals are marked with high fidelity, which ensures more reliable structures of the discontinuities within the mantle transition zone (MTZ) revealed from the receiver function gather. Figure 6b plots the local similarity between the removed noise and the noisy data. It is clear that this local similarity map is mostly zero and only contains a few areas with a high anomaly. The high anomaly indicates locations where the denoising algorithm may damage the useful signals. Because most areas are marked with low local similarity, it demonstrates that the proposed method does not cause significant damages to the useful converted-wave signals. The stacked traces from the raw depth-domain data and the denoised data are shown in Figure 5e. The red line plots the filtered data, and the blue dashed line plots the raw data. The two green dashed lines point out the expected positions of the 410 and 660 km discontinuities. From Figure 5e, we observe clearly that the waveforms corresponding to the 410 and 660 km discontinuities are of significantly higher resolutions. Because the amplitude in the denoised data is of higher fidelity due to the much reduced noise, we conclude that the proposed denoising method helps image more reliable MTZ discontinuities with a higher resolution.

Finally, we apply the proposed denoising method to an earthquake stack data. The dataset was originally used in Shearer (1991a,b). The seismic data of many earthquakes are stacked according to their epicentral distances (in degrees). To further improve the SNR of the final stack, the datasets from

Downloaded from https://pubs.geoscienceworld.org/ssa/srl/article-pdf/90/4/1552/4790732/srl-2019028.1.pdf
by Seismological Society of America, Mattie Adam

▲ **Figure 4.** Learned features from the UML method. The color version of this figure is available only in the electronic edition.

different earthquakes are also stacked. The dataset is then arranged in a 2D format, with the first axis denoting the recording time and the second axis denoting the epicentral distances. We can see a lot of seismic phases highlighted by the stack data in Figure 7a. However, there is still a lot of random noise existing in the earthquake gather. To remove the random noise, we apply the proposed UML method to the earthquake stack data. The denoised earthquake stack data are shown in Figure 7b. The seismic phases have been obviously enhanced, and the coherency of the main-wave components have become stronger; this is particularly true of the relatively weak seismic phases, which make the interpretation and further usages of these seismic phases more reliable. Figure 7c plots the removed noise from the raw stack data. Only a few obviously coherent signal components corresponding to the strongest phases are

seen in the removed noise, which indicates that the proposed method preserves most weak seismic phases well.

DISCUSSIONS

Denoising Accuracy and Reliability

To test the denoising accuracy, we create a synthetic example and conduct the denoising tests on the synthetic data. The advantage of the synthetic data test is that we have the ground-truth solution and then can evaluate the denoising performance by comparing the filtered data with the ground-truth solution, which would be the noise-free data. The synthetic example is shown in Figure 8. Figure 8a plots the clean data. We manually add some random noise into the clean data and obtain noisy data in Figure 8b. Figure 8c and 8d shows two denoised data using
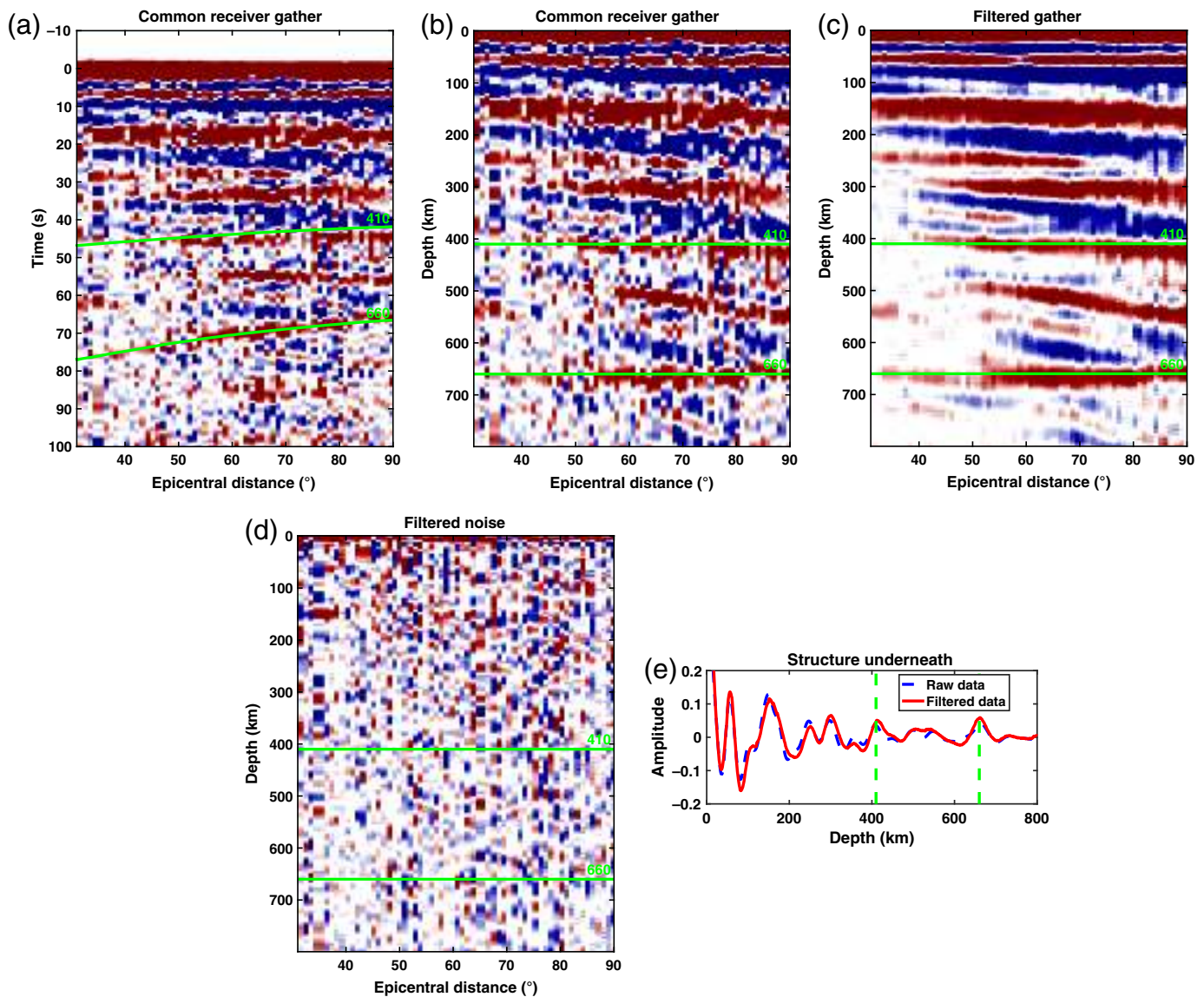
▲ **Figure 5.** Denoising performance of the receiver function data. (a) The noisy common receiver gather corresponding to the WALA station in the Canadian National Seismograph Network (CNSN) in the time domain; (b) the noisy data in the depth domain after time-to-depth conversion; (c) denoised data using the proposed method; and (d) removed noise using the proposed method. The two green solid lines highlight the expected arrivals of the converted waves, meaning the P410s and P660s. (e) Stacked RF data from common receiver gather shown in (b,c). The stacked trace depicts the discontinuity structure underneath the seismic station WALA in the CNSN. The blue dashed line shows the stacked data of the raw data. The red solid line shows the stacked data of the denoised data. Two green dashed lines denote the 410 and 660 km discontinuities, respectively. The color version of this figure is available only in the electronic edition.

the prediction-based (or predictive) denoising method and the proposed UML method, respectively. The comparison is positive in supporting the proposed method compared with the ground-truth solution. The denoised data using the predictive method still contains significant residual noise, but the denoised data using the proposed method are much closer to clean data. It is clear that the proposed method even preserves very subtle features in the data, such as the weak energy in the right up corner of the image. Because in this example we have the clean data, we can use the following SNR metric (Liu *et al.*, 2009; Chen, 2017) to evaluate the denoising accuracy:

$$\mathrm{SNR} = 10 \log_{10} \frac{\|s\|_2^2}{\|s - \widehat{s}\|_2^2}, \tag{9}$$

in which s denotes the noise-free data and $\widehat{s}$ denotes the noisy or denoised data. The calculated SNR of the noisy data (Fig. 8b) is 1.63 dB. The predictive method increases the SNR to 6.21 dB, whereas the proposed method increases the SNR further to 9.23 dB. The much higher SNR indicates that the proposed method can obtain higher accuracy, thus the resulting data are more reliable.
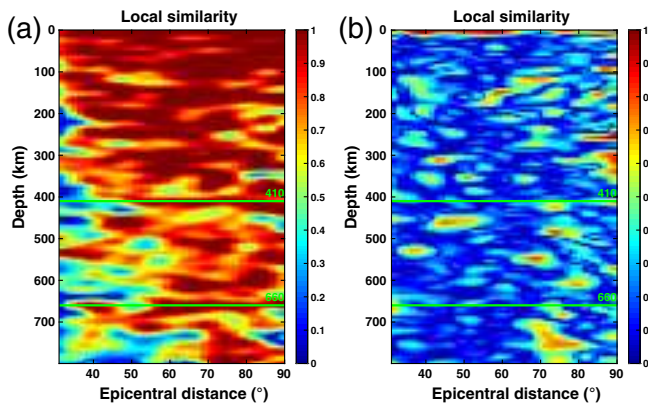
▲ **Figure 6.** (a) Local similarity between the denoised data and the noisy data. The high similarity anomaly indicates areas with high fidelity. (b) Local similarity between the removed noise and the noisy data. The high similarity anomaly in (b) indicates areas with high denoising uncertainties. The color version of this figure is available only in the electronic edition.

Effect of Noise

To investigate the effect of noise to the denoising performance of the proposed algorithm, we conduct several denoising tests in the case of different noise variances. We calculate SNRs for the noisy data, denoised data using the predictive method, and denoised data using the proposed method, when noise variance increases from 0.1 to 1. The calculated SNRs for three datasets are plotted in Figure 8e. From the diagrams, we can see that when noise level increases, the SNR of all three datasets decreases smoothly. This indicates that both the proposed denoising algorithm and the predictive method are robust to noise. Here, robust means that there will not be unstable issues when the noise level becomes very strong. However, the proposed method denoted by the blue line is always above the red line, indicating superior performance of the proposed method. Besides, the slope of the blue curve is slightly smaller than the red curve, indicating that the proposed method is slightly more insensitive to noise than the predictive method.

Boundary Effect

The boundary effect may occur when patching and unpatching the seismological datasets for training or prediction purposes. For an arbitrary size of the input data, it may require extension of the original data to create samples that cover the whole seismic section. For example, for the reflection seismic image shown in Figure 2a, the size is 512 × 128. When using a patch size of 40 × 40 with a shift size of 20 in each direction (vertical or horizontal), we need to extend the original seismic image to the size of 520 × 140, as shown in Figure 9a. We can see a narrower and a wider blank area on the right and bottom sides of the image, which are the extended areas. However, constructed patches from these blank areas have distinct features compared with patches from other areas, such as the patches shown in Figure 9d. There are obvious brown stripes in Figure 9d, indicating the patches created from the right and
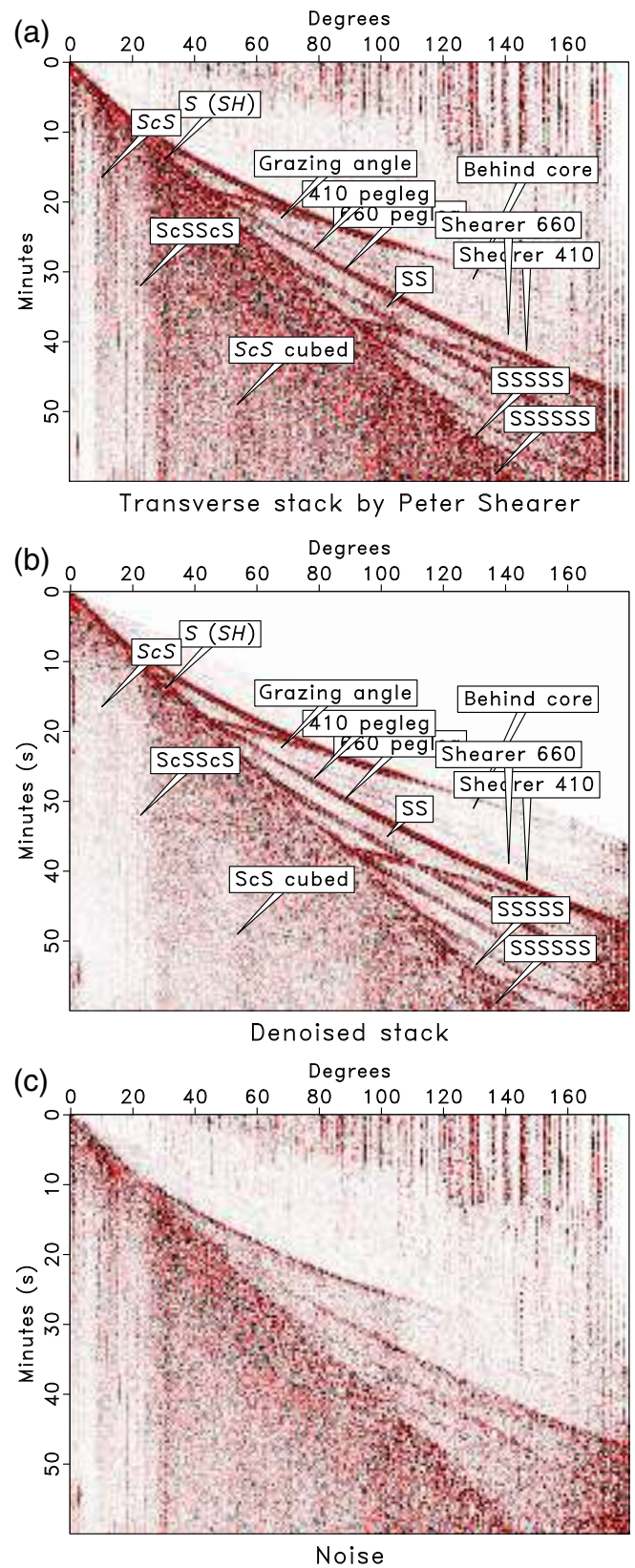
▲ **Figure 7.** Denoising performance for the earthquake stack. (a) Raw stack, (b) denoised data, and (c) removed noise. The color version of this figure is available only in the electronic edition.
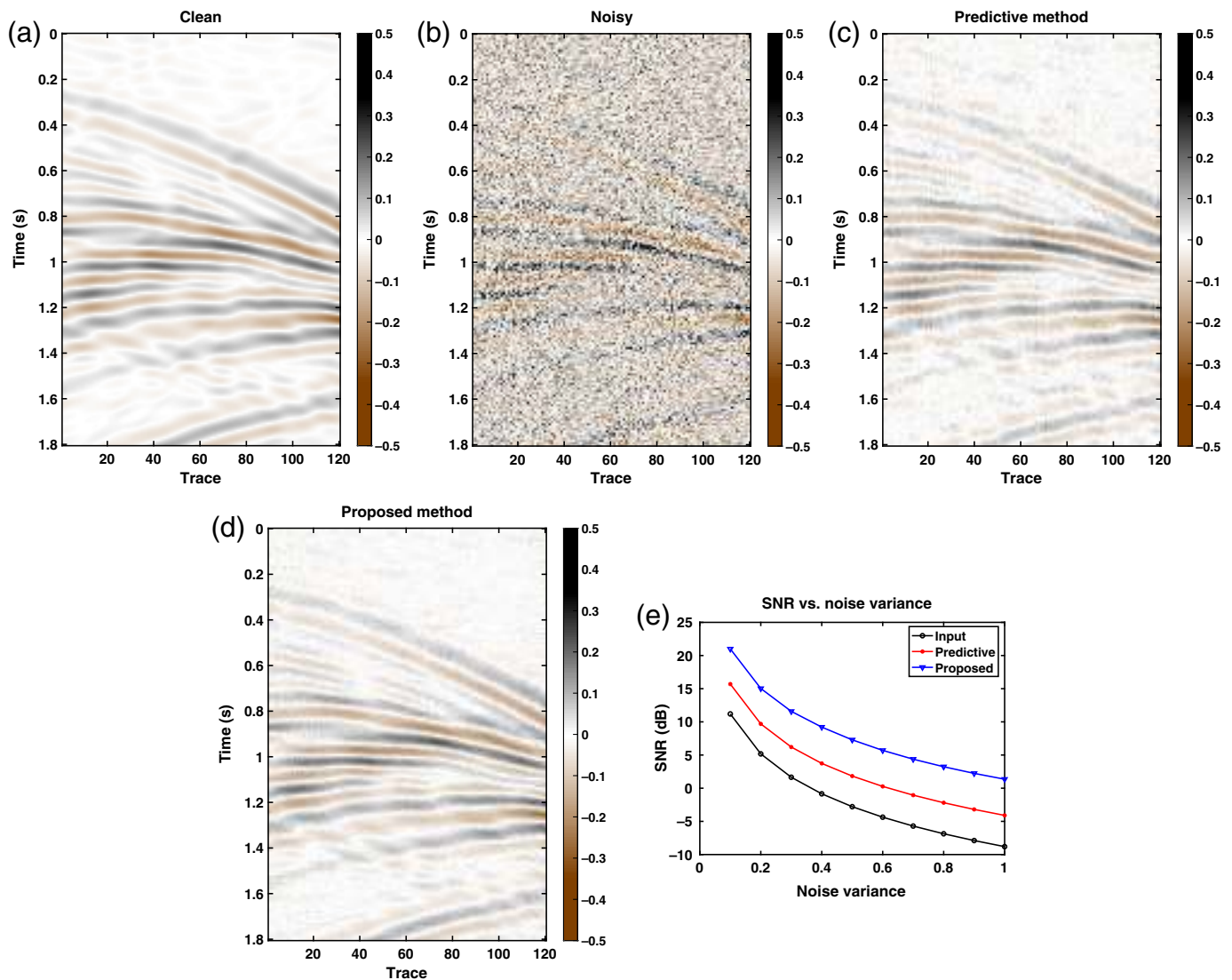
▲ **Figure 8.** Synthetic example. (a) Clean data, (b) noisy data (signal-to-noise ratio [SNR] = 1.63 dB), (c) denoised data using the prediction-based method (SNR = 6.21 dB), (d) denoised data using the proposed method (SNR = 9.23 dB), and (e) SNR diagrams in the case of different noise levels. The color version of this figure is available only in the electronic edition.

bottom boundaries. In the proposed algorithm, we use randomly selected patches from the input seismic image as the training dataset and then use all the regularly selected patches from the input data for testing; that is, we use them for prediction and denoising. If the boundary patches are not included in the training dataset, the algorithm will not be accurate in predicting the testing dataset. Ideally, those brown stripes in Figure 9d should be preserved during the prediction process; however, due to insufficient coverage of the training datasets, the predicted datasets, as shown in Figure 9e, will be far from the corrected data. The incorrect prediction will result in denoised data with strong boundary artifacts, which is shown in Figure 9b. To avoid the boundary effect, we need to include the boundary patches in the training dataset, so that the trained machine can take the boundary extension of the original seismic image into consideration and make a correct

prediction of the input testing datasets. The predicted testing data after including the boundary patches are shown in Figure 9f, which preserves the brown stripes (the boundaries) well. The reconstructed denoised data via an unpatching step from Figure 9f is shown in Figure 9c, which no longer contains the boundary artifacts.

Effect of the Training Data Size

It is known that the training data size may affect the performance in many machine-learning applications. Here, we intend to investigate how the training data size will affect the denoising performance in the proposed algorithm. We increase the number of randomly selected patches for training from 1000 to 6000. For each training data size, we conduct the training and prediction separately. We calculate the SNRs for each case and plot the SNR diagram with respect to variable
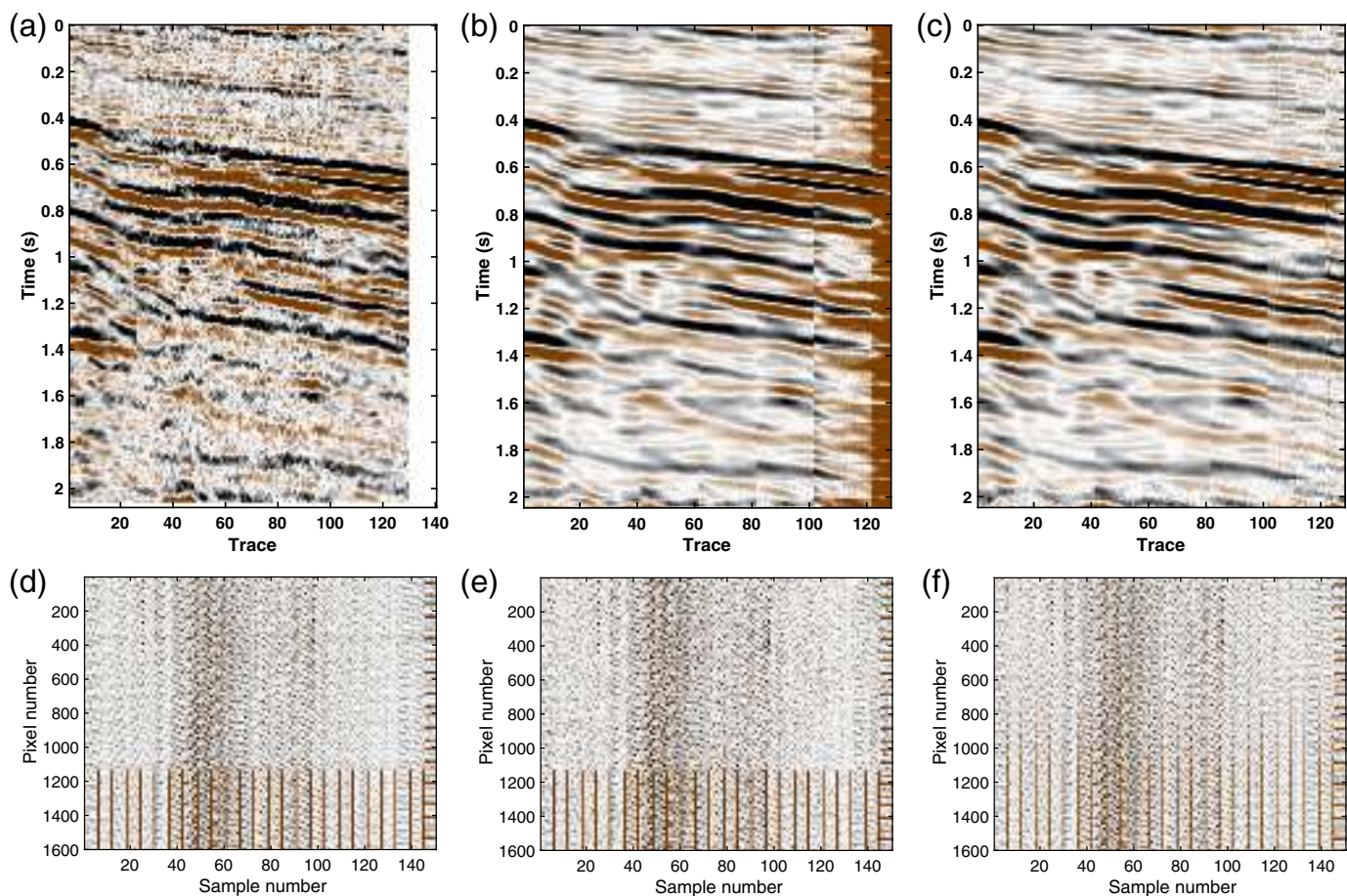
▲ **Figure 9.** (a–c) Demonstration of the edge effect. (a) Extended image for constructing the patches with size 40 × 40. (b) The denoised data when the boundary patches are not considered in the training samples. (c) The denoised data when the boundary patches are included in the training samples. (d–f) The comparison of the patches constructed from data shown in (a–c). (d) Patches constructed from the extended image. (e) Patches after applying the trained encoding and decoding network when not including the boundary patches. (f) Patches after applying the trained encoding and decoding network when considering the boundary patches. The color version of this figure is available only in the electronic edition.

training data size in Figure 10a. From Figure 10a, it is clear that SNR increases when the number of training patches increases. The SNR increases quickly when training data size increases from 1000 to 2000, then gradually increases from 10.46 to 12.54 dB when the training data size increases from 2000 to 5000. The SNR is nearly unchanged as the training data size changes from 5000 to 6000. This test indicates that a significantly large training data size can help obtain a better denoising performance; however, when the training data size is sufficiently large, the improvement of denoising performance is negligible.

Effect of the Patch Size

We also test the effect of the patch size on the denoising performance. We change the patch size from 20 to 60 and calculate the SNRs in different patch sizes. The SNR diagram with respect to variable patch size is shown in Figure 10b. From Figure 10b, we can observe that the SNR first increases when patch size increases from 20 to 40, and then it decreases when

the patch size changes from 40 to 60. This test tells us that an appropriate patch size needs to be adjusted to obtain the best denoising performance. This phenomenon can be explained by the fact that a large patch size would cause the learning process to miss small-scale features, whereas a small patch size would make the learning process incapable of learning meaningful waveform features. Thus, we suggest defining the patch size as approximately half of the dominant wavelength of data.

Effect of the Shift Size

Finally, we test the effect of the shift size on the denoising performance. We increase the shift size from 2 to 30 and compute the SNRs for different shift sizes. A smaller shift size corresponds to a large overlap between neighbor patches, as explained at the beginning of this article. The SNR diagram of the SNRs in different cases is shown in Figure 10c. It is evident from Figure 10c that the SNR decreases monotonically when the shift size increases from 2 to 30 points. From this test, we conclude that a large overlap between patches will help
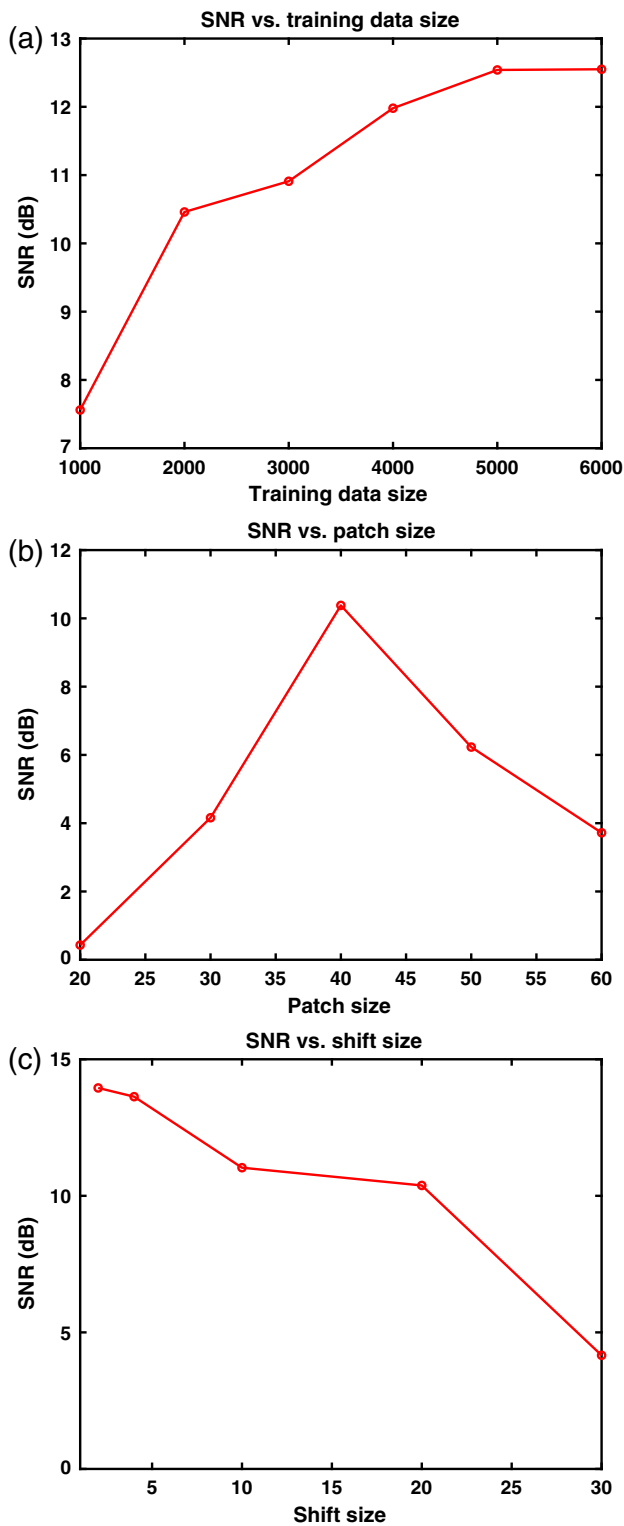
▲ Figure 10. (a) SNR diagram in the case of different training data sizes. It is clear that a larger training dataset helps obtain a better denoising performance. (b) SNR diagram in the case of different patch size. It is clear that an appropriate patch size can help obtain the best denoising performance. (c) SNR diagram in the case of different shift size. It is evident that a smaller shift size can obtain a better denoising performance. The color version of this figure is available only in the electronic edition.

obtain a better denoising performance. However, a larger overlap between patches will create a large number of redundant patches for the training process, which can be much more computationally expensive. Thus, an appropriate selection of the shift size to balance denoising performance and computational efficiency needs to be carefully designed. In this article, we simply choose half of the patch size as the shift or overlapping size.

CONCLUSIONS

Many types of seismological datasets contain strong seismic noise, which may impede the effective usage of these datasets for imaging and inversion purposes. We introduced a new denoising framework for improving the SNR of different types of seismological datasets based on an unsupervised machine-learning method. We utilize the autoencoder algorithm to adaptively learn the features from the raw noisy seismological datasets and use the sparse constraint to suppress the learned trivial features that may be associated with partial noise components. The selection of appropriate training samples is important to the learned features and also greatly affects the overall denoising performance. We use randomly selected patches that densely cover the whole dataset to obtain a satisfactory result. However, a more intelligent patch selection strategy is worth investigating in future research. Because of the nature of UML, the proposed denoising framework does not rely on carefully defined labels for the training dataset and thus can be much more flexible in practice. The applications on a multichannel reflection seismic image, a receiver function gather, and an earthquake stack data demonstrate that the proposed denoising framework can obtain better performance as opposed to the state-of-the-art competing methods. Most importantly, the proposed denoising algorithm can preserve subtle features in the seismic data while removing the spatially incoherent random noise.

DATA AND RESOURCES

ACKNOWLEDGMENTS

REFERENCES

Bai, M., J. Wu, S. Zu, and W. Chen (2018). A structural rank reduction operator for removing artifacts in least-squares reverse time migration, *Comput. Geosci.* **117,** 9–20.

Bergen, K. J., T. Chen, and Z. Li (2019). Preface to the focus section on machine learning in seismology, *Seismol. Res. Lett.* **90,** no. 2A, 477–480.

Bonar, D., and M. Sacchi (2012). Denoising seismic data using the nonlocal means algorithm, *Geophysics* **77,** no. 1, A5–A8.

Cadzow, J. A. (1988). Signal enhancement—A composite property mapping algorithm, *IEEE Trans. Acoust. Speech Signal Process.* **36,** no. 1, 49–62.

Canales, L. (1984). Random noise reduction, *54th Annual International Meeting, SEG*, Expanded Abstracts, Atlanta, Georgia, 6–7 December, 525–527.

Chai, C., C. J. Ammon, M. Maceira, and R. B. Herrmann (2018). Interactive visualization of complex seismic data and models using bokeh, *Seismol. Res. Lett.* **89,** no. 2A, 668–676.

Chen, Y. (2016). Dip-separated structural filtering using seislet thresholding and adaptive empirical mode decomposition based dip filter, *Geophys. J. Int.* **206,** no. 1, 457–469.

Chen, Y. (2017). Fast dictionary learning for noise attenuation of multidimensional seismic data, *Geophys. J. Int.* **209,** 21–31.

Chen, Y. (2018a). Automatic microseismic event picking via unsupervised machine learning, *Geophys. J. Int.* **212,** 88–102.

Chen, Y. (2018b). Fast waveform detection for microseismic imaging using unsupervised machine learning, *Geophys. J. Int.* **215,** 1185–1199.

Chen, Y., and S. Fomel (2015). Random noise attenuation using local signal-and-noise orthogonalization, *Geophysics* **80,** WD1–WD9.

Deuss, A. (2009). Global observations of mantle discontinuities using SS and PP precursors, *Surv. Geophys.* **30,** nos. 4/5, 301–326.

Gu, Y. J., Y. Zhang, M. D. Sacchi, Y. Chen, and S. Contenti (2015). Sharp mantle transition from cratons to cordillera in southwestern Canada, *J. Geophys. Res.* **120,** no. 7, 5051–5069.

Guan, Z., and F. Niu (2017). An investigation on slowness-weighted CCP stacking and its application to receiver function imaging, *Geophys. Res. Lett.* **44,** no. 12, 6030–6038.

Guan, Z., and F. Niu (2018). Using fast marching eikonal solver to compute 3-D Pds traveltime for deep receiver-function imaging, *J. Geophys. Res.* **123,** no. 10, 9049–9062.

Hua, Y. (1992). Estimating two-dimensional frequencies by matrix enhancement and matrix pencil, *IEEE Trans. Signal Process.* **40,** no. 9, 2267–2280.

Huang, W., R. Wang, S. Zu, and Y. Chen (2017). Low-frequency noise attenuation in seismic and microseismic data using mathematical morphological filtering, *Geophys. J. Int.* **211,** 1318–1340.

Kullback, S., and R. A. Leibler (1951). On information and sufficiency, *Ann. Math. Stat.* **22,** no. 1, 79–86.

Kumar, R., C. Da Silva, O. Akalin, A. Y. Aravkin, H. Mansour, B. Recht, and F. J. Herrmann (2015). Efficient matrix completion for seismic data reconstruction, *Geophysics* **80,** no. 5, V97–V114.

Li, Z., Z. Peng, D. Hollis, L. Zhu, and J. McClellan (2018). High-resolution seismic event detection using local similarity for large-N arrays, *Sci. Rep.* **8,** no. 1, Article Number 1646.

Liu, G., S. Fomel, L. Jin, and X. Chen (2009). Stacking seismic data using local correlation, *Geophysics* **74,** V43–V48.

Lomax, A., A. Michelini, and D. Jozinović (2019). An investigation of rapid earthquake characterization using single-station waveforms and a convolutional neural network, *Seismol. Res. Lett.* **90,** no. 2A, 517–529.

McBrearty, I. W., A. A. Delorey, and P. A. Johnson (2019). Pairwise association of seismic arrivals with convolutional neural networks, *Seismol. Res. Lett.* **90,** no. 2A, 503–509.

Mi, Y., X. Li, and G. F. Margrave (2000). Median filtering in Kirchhoff migration for noisy data, *2000 SEG Annual Meeting, Society of Exploration Geophysicists*, Calgary, Alberta, Canada, 6–11 August.

Morozov, I. B., and K. G. Dueker (2003). Signal-to-noise ratios of teleseismic receiver functions and effectiveness of stacking for their enhancement, *J. Geophys. Res.* **108,** no. B2, doi: 10.1029/2001JB001692.

Mousavi, S. M., and C. A. Langston (2016). Hybrid seismic denoising using higher-order statistics and improved wavelet block thresholding, *Bull. Seismol. Soc. Am.* **106,** no. 4, 1380–1393.

Mousavi, S. M., and C. A. Langston (2017). Automatic noise-removal/signal-removal based on general cross-validation thresholding in synchrosqueezed domain and its application on earthquake data, *Geophysics* **82,** no. 4, V211–V227.

Mousavi, S. M., C. A. Langston, and S. P. Horton (2016). Automatic microseismic denoising and onset detection using the synchrosqueezed continuous wavelet transform, *Geophysics* **81,** no. 4, V341–V355.

Rost, S., and C. Thomas (2002). Array seismology: Methods and applications, *Rev. Geophys.* **40,** no. 3, 2-1–2-27.

Rost, S., and M. Weber (2001). A reflector at 200 km depth beneath the northwest pacific, *Geophys. J. Int.* **147,** no. 1, 12–28.

Saki, M., C. Thomas, S. E. Nippress, and S. Lessing (2015). Topography of upper mantle seismic discontinuities beneath the North Atlantic: The Azores, Canary and Cape Verde plumes, *Earth Planet. Sci. Lett.* **409,** 193–202.

Schneider, S., C. Thomas, R. M. Dokht, Y. J. Gu, and Y. Chen (2017). Improvement of coda phase detectability and reconstruction of global seismic data using frequency–wavenumber methods, *Geophys. J. Int.* **212,** no. 2, 1288–1301.

Shearer, P. M. (1991a). Imaging global body wave phases by stacking long-period seismograms, *J. Geophys. Res.* **96,** no. B12, 20,353–20,364.

Shearer, P. M. (1991b). Constraints on upper mantle discontinuities from observations of long period reflected and converted phases, *J. Geophys. Res.* **96,** no. B11, 18,147–18,182.

Vautard, R., P. Yiou, and M. Ghil (1992). Singular-spectrum analysis: A toolkit for short, noisy chaotic signals, *Phys. Nonlinear Phenom.* **58,** no. 1, 95–126.

Vogl, T. P., J. Mangis, A. Rigler, W. Zink, and D. Alkon (1988). Accelerating the convergence of the back-propagation method, *Biol. Cybern.* **59,** nos. 4/5, 257–263.

Zhang, C., M. van der Baan, and T. Chen (2018). Unsupervised dictionary learning for signal-to-noise ratio enhancement of array data, *Seismol. Res. Lett.* **90,** no. 2A, 573–580.

Zhang, G., Z. Wang, and Y. Chen (2018). Deep learning for seismic lithology prediction, *Geophys. J. Int.* **215,** 1368–1387.

Zhou, Y., S. Li, D. Zhang, and Y. Chen (2017). Seismic noise attenuation using an online subspace tracking algorithm, *Geophys. J. Int.* **212,** no. 2, 1072–1097.

Zhu, W., S. M. Mousavi, and G. C. Beroza (2018). Seismic signal denoising and decomposition using deep neural networks, available at https://arxiv.org/abs/1811.02695 (last accessed December 2018).

Zu, S., H. Zhou, W. Mao, D. Zhang, C. Li, X. Pan, and Y. Chen (2017). Iterative deblending of simultaneous-source data using a coherency-pass shaping operator, *Geophys. J. Int.* **211,** no. 1, 541–557.

Yangkang Chen
Min Bai
School of Earth Sciences
Zhejiang University
Number 866, Yuhangtang Road, Xihu District
Hangzhou 310027, Zhejiang Province, China
yangkang.chen@zju.edu.cn

Mi Zhang
State Key Laboratory of Petroleum Resources and Prospecting
China University of Petroleum
18 Fuxue Road
Beijing 102200, China
cupmi@sina.com

Wei Chen[1,2]
Key Laboratory of Exploration Technology for Oil, and Gas
Resources of Ministry of Education
Yangtze University
Number 111, Daxue Road, Caidian District
Wuhan 430100, China
chenwei2014@yangtzeu.edu.cn

Published Online 22 May 2019

[1] Also at Hubei Cooperative Innovation Center of Unconventional Oil and Gas, Number 111, Daxue Road, Caidian District, Wuhan 430100, China.
[2] Corresponding author.